ECONOMIC POLICY

MINISTÈRE DE L'ÉCONOMIE, DES FINANCES ET DE LA RELANCE
Liberté
Égalité
Fraternité

Direction générale du Trésor

75th Economic Policy Panel Meeting

7- 8 April 2022

Immigration and Teacher Bias towards Foreign Students

Marco De Benedetto & Maria De Paola

CEPR CESifo SciencesPo.

Immigration and Teacher Bias towards Foreign Students[*,†]

Marco Alberto De Benedetto[‡1] and Maria De Paola[§2]

[1]Department of Economics, Statistics and Finance, University of Calabria (CS), Italy.

[2]National Institute of Social Security (INPS), Rome.

[2]Institute for the Study of Labor (IZA), Bonn.

Abstract

We analyze the role of changes in the geographic concentration of immigrants in shaping teachers' assessment of students' performance. By using data on Italian students attending the 5^{th} grade, we adopt an IV estimation strategy and, by controlling for student performance in blindly-scored tests of proficiency, find that an increasing presence of immigrants in the local population negatively affects the way teachers evaluate immigrant students, as opposed to their peers, in non-blindly-graded tests. Very similar results are found for 2^{nd} and 8^{th} graders, who have interacted with teachers for a shorter period than 5^{th} graders have. We also reveal that the effect is mainly driven by schools located in smaller communities and in areas with lower overall levels of educational attainment and that it is unlikely to be related to the conduct of immigrant students who live in the areas experiencing sizable increases in immigration flows. In addition, older and less qualified teachers tend to be more biased.

JEL classification: J13; J15; I21; I24.

Keywords: Immigration; Discrimination; Students' performance; Teachers' assessment; Bias in grading.

---

[*]Declaration of interest: none

[†]Acknowledgement: We would like to thank the Special Guest editor (Prof. Barbara Petrongolo) and three anonymous referees for providing us with useful comments that allowed us to enhance this paper, Guido De Blasio for having provided us with data to build a measure of the share of immigrants in the population in 1991. We are also grateful to Giorgio Brunello, Francesca Gioia, Paolo Naticchioni, Vincenzo Scoppa and participants at the 2021 EALE Conference for their suggestions.

[‡]Corresponding author: marco_alberto.debenedetto@unical.it

[§]Email: m.depaola@unical.it

1 Introduction

The economic and social consequences of migration inflows are often debated by both academics and politicians. Although immigration can bring relatively large long-term benefits to the economies of host countries (Peri, 2012; Hong and McLaren, 2015; Sequeira et al., 2017; Tabellini, 2020) and many empirical investigations have only found minimal economic consequences for native populations (Ottaviano and Peri, 2012; Manacorda et al., 2012), citizens still frequently rate it as one of the most pressing issues faced by their countries.

According to the "group threat" theory, geographic concentrations of immigrants might play a crucial role in shaping attitudes towards immigration and immigrants (Quillian, 1995): an increase in the proportion of foreigners living within a certain area induces natives to adopt more conservative attitudes as they feel threatened.[1]

Adverse attitudes toward immigrants can easily translate into discriminating behaviors which might negatively affect the socio-economic status and integration of immigrants (Johnston and Lordan, 2016). Indeed, prejudice towards minority groups might widen racial inequalities in both wages and employment opportunities. At the same time, a hostile environment is likely to have negative effects on immigrants' mental and physical health (Johnston and Lordan, 2012). These problems might be even more significant when discrimination takes place in schools and hostility comes from peers and teachers.

In this paper, we investigate the impact of the presence of immigrants in a given geographic area on the behavior of teachers in their evaluation of the performance of immigrant students compared with natives. The idea behind this research is that an increasing presence of immigrants might generate, even unconsciously, feelings of hostility that could lead to discriminatory attitudes. Due to exposure to teachers' stereotypes, young immigrants may become discouraged and make less of an effort in their study activities, with consequent negative results for their careers. Teachers might modify their behavior in class as a reaction to students' migratory status; for example, they might have stereotyped expectations of student skills which may result in a self-fulfilling prophecy (Papageorge et al. 2020).[2]

[1]On the other hand, some recent extensions of the "intergroup contact" theory (Allport et al., 1954) propose that larger immigration flows, providing greater opportunities for intergroup contact, reduce perceived threats and render natives' attitudes towards immigrants more accommodating (Wagner et al., 2006; Schlueter and Wagner, 2008). However, the evidence supporting this assumption is limited. Wagner et al. (2006) show that changes in the share of ethnic minority members across districts in Germany reduce the prejudice of the majority. Similar results are found by Schlueter and Wagner (2008) when using regionalized cross-national data from Europe. However, this evidence might be affected by endogenity issues.

[2]This is nicely documented in the well-known psychological study conducted by Rosenthal and Jacobson (1968), where it is shown that the randomly allocated student IQs, communicated to teachers at the beginning of the year,

Teachers might also demand less effort from minority students or give them less feedback, praise them less often for success and criticize them more frequently for failure, and, in turn, these students might lose motivation and self-confidence.[3]

Our analysis is based on large-scale observational data. We have access to five annual censuses of all the pupils attending the 5^{th} grade in Italian primary schools and these provide rich information on pupils' characteristics (citizenship, socio-economic background and psychological traits) and their performance in terms of both standardized test scores and marks assigned by math and language teachers. The latter information forms the basis of our empirical strategy, which relies on a combination of blind and non-blind test scores in order to investigate whether teachers in primary schools are biased in favor of native pupils in their evaluations. To analyze whether an eventual bias is related to the presence of immigrants in the area, we exploit the information offered by our data on school location and, since teachers are likely to live within the local labor market where the school is located, we consider the concentration of immigrants at the local labor market level.

We estimate a local labor market fixed effects model and handle the endogeneity and reverse causality problems, deriving from the fact that the share of immigrants in a given geographic area is not exogenous, by using an instrumental variable approach that relies on pre-existing settlement patterns. Our dependent variable is the score received by students from teachers in math and language in non-blindly-graded tests, while the independent variable of main interest is the interaction term between the dummy variable for immigrant students and the share of immigrants in the local labor market as this captures a differentiated effect of changes in the number of immigrants present in the area on the scores obtained by immigrant and native students.

From 2SLS estimates and by controlling for individual performance on standardized blindly-graded tests, we find that immigrant students obtain lower grades than natives from their teachers in non-blind tests. We also find that this grade penalty is driven by schools located in areas which are experiencing large inflows of immigrants. While teachers working in such areas tend to give lower grades to immigrant students than to natives, even though they perform in the same way in blindly-graded tests, those teachers operating in schools located in areas with low levels of immigration grade native and immigrant students similarly. We find that a one standard deviation increase in the share of immigrants in the local labor market is associated with a 0.044

affect students' performance at the end of the year.

[3]Psychologists have widely discussed how beliefs about social groups are a powerful determinant of attitudes and behavior towards members of these groups (Fiske, 1998).

(0.053) decrease in immigrants' grades in math (language) on average, corresponding to about 1/5 (1/6) of the average gap. This gap between immigrants and natives is particularly relevant when it comes to language evaluation for second generation immigrants and this, in line with what is found by Alesina et al. (2018), might reflect different teachers' expectations for immigrants who are less familiar with the Italian language.

These results are robust to the inclusion of school or class fixed effects, to alternative measures of student performance in blindly-graded tests and when, instead of considering the share of immigrants in the local labor market in which the school is located, we conduct our analysis at municipal level.

Similar findings emerge for students in the 2^{nd} and 8^{th} grade, suggesting that the amount of time for "classroom-interaction" between teachers and pupils does not play a particularly relevant role in explaining teachers' different assessment standards. Teacher bias against immigrants is weaker (and even absent) in larger and more educated municipalities and in regions where there are already large historical immigrant communities. Furthermore, younger and more educated teachers seem to be less reactive to increases in immigration flows. Our results also show that teacher bias is not related to the fact that students in areas with large migration inflows are less integrated, and as a consequence behave differently at school.

All in all, these results suggest that the penalty paid by immigrant students in non-blindly-graded tests is likely to depend on teachers working in poorer areas and in regions that are less accustomed to ethnic diversity who become more biased against immigrants as their presence in the area increases.

Our paper adds to the literature which compares "blind" and "non-blind" assessment methods in schools in order to investigate teacher bias. One strand of this literature has focused on stereotypical attitudes of teachers towards boys and girls. In a pioneering paper, using data on Israeli high schools, Lavy (2008) shows that male students face discrimination and obtain lower grades in all subjects. Similar evidence is also found by a number of studies that use data from other countries (Hinnerich et al., 2011; Hanna and Linden, 2012; Cornwell et al., 2013; Di Liberto and Casula, 2016). The same methodology has also been adopted to investigate teacher behavior towards minority groups. For instance, Burgess and Greaves (2013) exploit the English testing system of "quasi-blind" externally marked tests and "non-blind" internal assessment to investigate whether ethnic minority pupils are subject to low teacher expectations. They find that, relative to white pupils, black Caribbean and black African pupils are under-assessed, whereas

Indian, Chinese, and mixed white and Asian pupils are over-assessed. Similar results are found by Botelho et al. (2015), who show that racial discrimination in terms of biased assessment of students is prevalent within Brazilian schools, and by Alesina et al. (2018), who analyze the behavior of Italian teachers towards immigrant and native students and show that, after controlling for performance in standardized blindly-graded tests, immigrant children in middle schools receive lower teacher-assigned grades than natives do.

Our study complements this literature by providing additional evidence of teachers' biased assessment and shows that this bias is driven by teachers who work in areas with current high immigration flows. A similar analysis that focuses on teacher grading bias has been undertaken by Alesina et al. (2018) and shows that the grading bias correlates with teachers' preconceptions of immigrants as measured by the Implicit Association Test.[4]

Our paper is also connected with the literature investigating the effects of the presence of immigrants on an array of political and social outcomes. A number of studies have investigated how immigration flows within a given area influence the political views of native voters, for instance their tendency to vote for right-wing parties with anti-immigration platforms or their support for certain policies. A common finding of this strand of literature is that an increase in the proportion of immigrants in a given area leads to greater support for anti-immigration parties[5] (see, for instance Otto and Steinhardt, 2014; Barone et al., 2016; Harmon, 2018; Dustmann et al., 2019; Mayda et al., 2020; Becker and Fetzer, 2016; Halla et al., 2017; Brunner and Kuhn, 2018; Tabellini, 2020).[6]

These works address the endogeneity of immigrant location choices by using an instrumental variable strategy based on pre-existing settlement patterns. Adopting the same methodological strategy, we add to this literature by investigating whether the concentration of immigrants

[4]A different approach has been taken by a number of other papers that rely on randomized control trials. Van Ewijk (2011) runs a field experiment in which, by manipulating the names of students who had the task of writing essays, a group of teachers working in the Netherlands were induced to believe that some essays were written by Dutch students while others were written by Turkish or Moroccan students. Results show that - on average - teachers do not exhibit a grading bias, but they have lower expectations of children from ethnic minorities. By using a similar experiment in which German and Turkish names are randomly assigned to sets of essays written by fourth-graders, Sprietsma (2013) provides evidence of the presence of grade discrimination in primary education in Germany. Finally, Hanna and Linden (2012) find a negative difference between blind and non-blind test scores for members of lower castes in India (relative to upper castes), which is a clear evidence of discrimination.

[5]The negative impact of immigration flows on natives' attitudes towards immigrants might be driven by both concerns regarding the economic effects of immigration and non-economic worries, such as those derived from the idea that immigrants' cultural diversity is a threat to the values of the host communities and an impediment to social cohesion (negative externalities of immigration on compositional amenities). Card et al. (2012) find that compositional amenity concerns are 2–5 times more important in explaining attitudes towards immigrants than concerns over wages and taxes.

[6]Another strand of the literature has analyzed the effect of immigration on native attitudes relating to redistribution policies (Alesina et al., 2018; Moriconi et al., 2019). A common finding is that support for redistribution diminishes as the number of immigrants in the population increases.

within a given geographic area (influencing the social climate and attitudes towards immigrants) affects immigrants' educational outcomes. A similar attempt has been made by Bracco et al. (2022) who analyze how the change in the political climate in Italian municipalities, induced by local elections in which the anti-immigration Lega Nord party stood, affected school bullying. They find that election campaigns where Lega Nord is supported lead to a higher incidence of victimization among immigrant students.

In our work, we look at a different outcome and, instead of focusing on the impact of electoral campaigns of anti-immigration parties, we turn our attention to how teachers' behavior is affected by the local presence of immigrants. As teacher behavior based on stereotypes may have long lasting consequences on students' school choices and outcomes (Lavy and Sand, 2018; Lavy and Megalokonomou, 2017; Terrier, 2015; Carlana, 2019) and, through this channel, on their following labor market and lifetime outcomes, the results of our research are particularly relevant.

The paper is structured as follows. Section 2 describes the institutional setting and introduces the data. Section 3 focuses on the empirical strategy. In Section 4 and 5, we present our baseline results and some robustness checks. In Section 6, we try to shed light on the mechanisms underlying our findings. Section 7 concludes.

2 Institutional setting and data description

Immigration is a phenomenon of growing importance all around the world. According to ISTAT (the Italian National Institute of Statistics), the number of immigrants in Italy reached 8 percent of the total population (4,8 million) at the end of 2012, slightly lower than in other European countries such as Germany and France. Moreover, Italy has experienced immigration as a more recent event than the other large European countries. Indeed, although immigrants made up less than 1.7 percent of the population at the end of the 1990s, since 1998, Italy has experienced sizeable inflows and the proportion reached 8.8 percent at the end of 2019, while in Germany and France the share of immigrants has only changed slightly, floating around 9 and 6 percent, respectively.[7]

Along with the general trend, the presence of children from an immigrant background in Italian schools has risen rapidly over recent years, reaching 11.5 percent in primary and lower

[7]Regarding the composition of migration inflows, immigrants in Italy mainly come from countries outside the European Union and from low-income countries, with no relevant variation over the period 2012-2016. In particular, the top 15 countries of origin are Albania, Romania, Morocco, China, the Philippines, Tunisia, Serbia and Montenegro, Macedonia, Poland, India, Peru, Senegal, Egypt, Sri Lanka, and Ecuador respectively, accounting for roughly 69 percent of the total.

secondary schools and floating around 7 percent in upper secondary schools in 2019 (see Ministry of Education, University and Research - MIUR, 2020). We rely on different sources of data in order to investigate the impact of this change in the proportion of immigrants registered in the local labor market on student outcomes with regard to their citizenship. First, information on student performance comes from a dataset provided by INVALSI, a government agency which tests student attainment in both language and math every year. The assessment covers students attending 2^{nd} and 5^{th} grades (primary schools), as well as 8^{th} (lower secondary schools) and 10^{th} grades (upper secondary schools).

In our work, we focus on primary schools, as the experiences that children have in primary schools are fundamental to how they conceive of school and how they view themselves in relation to native teachers and classmates (Heckman, 2008). Any experiences of discrimination at this vulnerable age can limit the emotional benefits of early education and might result in a multitude of long-lasting negative consequences in terms of future school drop-out, academic performance and job trajectories. In particular, we restrict our analysis to students in the 5^{th} grade (although similar results hold true for 2^{nd} and 8^{th} graders).[8] We end up with a sample of 1,219,572 students in 6,861 institutes (14,847 schools) located in 5,962 municipalities and 580 local labor markets from the school-years 2012/13 to 2016/17.

From the INVALSI source of data, we gather information on standardized blindly-graded test scores and on scores assigned by teachers in non-blindly-graded tests in the two core subjects in the primary school program: the Italian Language (*Language*) and Mathematics (*Math*). The data distinguish between "written grades" and "oral grades", but, given the large percentage of missing values in written grades (69.08 and 69.01 percent for language and math, respectively), we only consider oral grades. Teacher-assigned grades are on a scale from 1 to 10.

As regards student performance in standardized tests, there are a number of alternative indicators available. The first indicator is the "Cheating-corrected" test score, which is adopted by INVALSI to avoid problems of score manipulation. Given that cheating is a common occurrence which could skew the reliability of standardized test scores (see for instance Angrist et al., 2017; Bertoni et al., 2013), INVALSI has developed a statistical solution to purge the data of this problem. This method exploits the statistical properties of the distribution of answers given in

[8]For the 10^{th} grade, the student questionnaire is available, but a large proportion of immigrant children drop out of school before this grade. While 11.3 percent of Italian natives drop out of school once they are in the 8^{th} grade, this share rises to 36.5 percent among those of immigrant background (Eurostat, 2019: https://www.openpolis.it/quanto-e-frequente-labbandono-scolastico-tra-gli-alunni-stranieri/). As this means that the remaining immigrant students are highly selected, it would be difficult to compare them with their native counterparts.

classes where the test is taken under the supervision of external examiners (randomly assigned to selected classes and schools to perform monitoring) and uses a continuous class-level probability of manipulation (similar to that estimated in Angrist et al., 2017). This probability is based on the variability of intra-class percentage of correct answers, modes of wrong answers, etc. and the resulting estimates are used to "deflate" the raw scores in the test.[9] The second indicator is the scores as computed by INVALSI, which applies the IRT Rasch model[10] to answers to the tests, in order to account for difficulties of single items (*Rasch Language score* and *Rasch Math score*). The third indicator is that of using the fraction of correct answers in language and math multiple-choice tests (*Language score* and *Math score*).[11]

It is worth noting that the INVALSI tests are identical across schools, whereas marks assigned to students in class are based on a standard that is independently set by each teacher.[12]

To differentiate between students of Italian and non-Italian extraction we use a variable, known as the "citizenship indicator", that comes from the INVALSI dataset. This variable takes three different possible values: "Italian", "First generation immigrant" and "Second generation immigrant". Using this information, we build the *Immigrant* dummy variable which takes the value 1 if the student is a first or second generation immigrant. First-generation students are students born abroad to foreign-born parents and second-generation students are children born in Italy to foreign-born parents. As shown in Table 1, where some descriptive statistics are reported, 8.46 percent of students in our dataset are immigrants, that is 6.15 percent are second generation immigrants while the remaining 2.31 percent are first generation immigrants.

[Insert Table 1]

Immigrants attending the 5^{th} grade in Italian primary schools performed worse than their peers in both non-blindly and blindly-graded tests. On average, the "cheating corrected" score on math (language) test is 22.58 (29.78) for immigrants and 25.65 (34.66) for natives, while the average grade assigned by math (language) teachers to immigrant and native pupils is 7.45 (7.27) and 7.99 (7.93), respectively. In Figure 1, to present this visually, we plot the average grade

[9]For a detailed description of the method see Campodifiori et al. (2010).

[10]The Rasch model is a logistic model which belongs to the area of the Item Response Theory (IRT) and operates a joint estimate of two types of parameters: a difficulty parameter for each test question and a skill parameter for each student. In particular, the Rasch model allows expression of the probability of choosing the correct answer in an item as a function of the difficulty of the item itself and of the student ability as measured in the entire test.

[11]Similar results are found when this alternative measure is used as a proxy for student ability in our model (not reported, but available upon request).

[12]Moreover, while INVALSI tests assess student performance on an absolute grading scale, teachers might adopt relative marking which might also be affected by class composition or by class size.

assigned by math and language teachers to immigrant and native children by quintiles of the standardized test score in math and language, respectively.

The dataset at hand also provides information on a number of students' and parents' characteristics (gender, attendance at kindergarten and pre-schooling, parents' education and employment). By exploiting this information, we build a set of dummy variables, taking the value 1 (and 0 otherwise), for: girls (*Female*); students who have attended kindergarten (*Kindergarden*), students who have attended pre-primary school (*Pre-primary*), father's level of education (elementary, middle, high-school or college), mother's level of education (elementary, middle, high-school or college) and father and mother's employment status (unemployed, homemaker, manager, entrepreneur, retailer, professional, teacher, manual worker or retired). In addition, information on the family background of students is used by INVALSI to build an indicator of socio-economic status (called ESCS-Economic and Social Cultural Status),[13] with a zero mean and unitary standard deviation.

We also have information on whether the student is younger or older than a regular student (we build a dummy variable *Regular* that takes the value 0 for grade-repeaters or early-starters and 1 otherwise). This variable could be important in reference to immigrant children in Italy, who often have different school-age enrolment patterns from native Italian students.[14]

As regards school organization, we know whether a class follows a full-day or half-day schedule and, on the basis of this information, we build a dummy variable, *Full day*, for those classes whose schedule is organized in terms of entire school days (8am-4pm usually) rather than only in the morning. We also have information on the number of students enrolled in each grade at the beginning of the school year (*Class size*) and the number of classes for the 5^{th} grade within the school (*School size*). In addition, we have computed the share of immigrant and female students in the class, i.e. *Share of immigrants in class* and *Share of female students in class* with a mean of 0.101 and 0.505, respectively.

Apart from measures of cognitive skills, the INVALSI dataset also allows us to build measures

[13]This indicator is built in accordance with the one proposed in the OECD-PISA framework and considers parents' occupation, education and the educational resources available at home (for instance, the number of books). For a detailed description see Ricci (2010), http://new.sis-statistica.org/wp-content/uploads/2013/09/RS10-SP-The-Economic-Social-and-Cultural-Background-a-continuous-index-for-the-Italian-Students-of-the-fifth-grade.pdf.

[14]In Italy, a student starts primary school in September of the calendar year (Jan-Dec), in which she turns six, e.g. children born in 2014 start primary school in September 2020. If students are slightly too young to begin school (i.e. children born in January-April 2015 in our example), parents can freely choose to let their children start primary school a year earlier. On the other hand, though, it is not uncommon for recently arrived immigrants, who are lagging in language skills or academic background, to be put into classes with pupils who are younger than them. From our data, it emerges that immigrants make up 33 percent of pupils attending a grade which is lower than their age would seem to indicate.

of individual feelings toward the standardized test, which might differ between immigrant and native students. We exploit the survey that is administered to students attending the 5^{th} grade (*Student Questionnaire*) on the same day as one of the two tests is taken and consider three questions concerning whether: 1) students were already worried before taking the test, 2) students were nervous during the test and unable to find the right answer, and 3) students had the impression they performed badly during the test. For each of these questions, students have four possible answers: *Not at all*, *A little*, *Moderately* and *Very much.* Then, for each of these questions and for each answer, we build a set of dummy variables which take the value of 1 when the student picks that specific answer and zero otherwise. Roughly 53.3 percent of the students in our sample were worried before taking the standardized test (those who replied *Moderately* and *Very much*), 54.7 percent of students were nervous during the test and 42.4 percent felt their performance was poor.

Finally, we have information on the region, province and municipality in which each school is located. This information allows us to merge the INVALSI dataset with a second source of data, the Demographic Balance and Resident Population by Sex and Citizenship (*ISTAT, Bilancio demografico e popolazione residente per sesso e nazionalità*), and, in particular, to build our main variable of interest i.e. *Share of Immigrants*, which, since 2002, has presented information on legally resident foreigners (by citizenship and sex) in each Italian municipality on 1st January of each year. Furthermore, ISTAT Territorial Statistics provides us with data on the average population in the local labor market (LLM hereafter), i.e. *Population size LLM.*

Moreover, in order to build up our instrumental variable, $Z_{(m,t)}$, we gather 1991 information from ISTAT on immigrants both by municipality of residence and area of origin in the world and details on residence permits by province and country of origin from the Italian Ministry of the Interior. In particular, following Barone et al. (2016), data by municipality and nationality are obtained by imputing for each municipality within a given area of the world the nationality share observed at the provincial level. Then, both the share of immigrants and the instrument are aggregated at the local labor market level.

3 Empirical methodology

In order to recover the impact that the share of immigrants in the local labor market exerts on the behavior of teachers in evaluating immigrant students' performance, as compared with that of natives, we estimate a LLM fixed effects regression model. We decided to focus on LLM rather than municipalities for two main reasons. First, since we do not have information on where teachers actually live, a change in the immigration flows in one municipality (a teacher's place of residence) could also affect the way that teacher evaluates immigrant students in non-blindly-graded tests in an adjacent municipality (where the school is located). For example, if teacher j lives in municipality l, but works in nearby city n, we are unable to assess whether his/her behavior is affected by the immigration flows registered in l or n. Second, the INVALSI database only provides us with information on the municipality where institutes (and, in turn, their headquarters) are based, but the municipality might include more than one school.

We estimate the following model:

$$\begin{aligned} y_{ismt} = \beta_0 + \beta_1 Immigrant_{ismt} + \beta_2 Share\ of\ Immigrants_{mt} + \beta_3 Immigrant \\ * (Share\ of\ Immigrants)_{ismt} + \beta_4 g(Invalsi\ Test\ Score)_{ismt} + \beta_5 X_{ismt} + \\ + \beta_6 W_{smt} + \beta_7 Non - cognitive\ Skills_{ismt} + \beta_8 Pop_{mt} + \gamma_m + \mu_t + \varepsilon_{ismt} \end{aligned} \quad (1)$$

where the outcome variable is measured by the scores a teacher awards student i in school s located in the LLM m at time t in non-blindly-graded tests in language and math, respectively. The interaction term $Immigrant * (Share\ of\ Immigrants)_{ismt}$ is our main variable of interest and measures the gap in non-blindly-graded tests between immigrant and native students due to a change in the share of immigrants registered in the local labor market m at time t. Following Alesina et al. (2018), we further control for student ability, i.e. $g(InvalsiTestScore)_{ismt}$, which is a flexible polynomial function (from linear to cubic) of the score obtained by students from INVALSI in a blindly graded-test in language or math.[15]

We also include among the regressors X_{ismt}, which is a vector of students' characteristics (*Female*, *Kindergarten*, *Pre-primary School*, *Regular*, dummy variables for mother and father's educational level and occupational status). W_{smt} is a vector of school characteristics and class composition features (*Full-time*, *School size*, *Class size*, *Share of immigrants in class*, *Share of female students in class*), whereas Pop_{mt} is the average population in the LLM. γ_m and μ_t are fixed effects at local labor market level and year dummies, respectively. In particular, γ_m considers unobservable time-invariant LLM characteristics affecting both immigrants' decision to move to a specific area and teachers' evaluation of immigrant and native students in non-blindly-graded tests; ε_{ismt} is the error term of the model.

[15] Since data on teacher-assigned grades are collected at the end of the first semester (January), while the INVALSI test is administered at the end of the second semester (May), this information might not be ideal for calculating bias in grading. However, it is not clear what kind of bias we should expect from this as it depends on the reaction of both immigrant and native students to teacher behavior. It could be that immigrant students who see that they are given low grades by their teachers are induced to put less (more) effort into the INVALSI test, for instance because they get discouraged (try to catch up) and, therefore, any potential bias in grading may be underestimated (overestimated). On the other hand, native students are also likely to change their behavior in response to the better grades they obtain at the end of the first semester. Overall, the direction of the bias is difficult to assess.

Nonetheless, the inclusion of LLM fixed effects in equation (1) does not allow us to interpret the OLS estimates in a causal manner. First, there could be an omitted variable that correlates with both the share of immigrants in the local labor market and the way students are evaluated by teachers in non-blindly-graded tests, leading to a potential downward/upward bias. For example, the presence of a positive shock in the economy can increase demand for immigrants and, as a consequence, this could negatively (positively) influence teachers' assessment of immigrant (native) students. In addition, immigrants are not randomly distributed across local labor markets, but rather they self-select on the basis of certain factors, for instance they may decide not to live in areas where there is hostility towards multiculturalism and/or in places where the schools are reputed to discriminate against children from immigrant backgrounds.

Finally, measurement error in the main variable of interest could be at play, leading to an overall downward bias in our estimates. Indeed, it is extremely difficult to keep a perfect track of immigrants in a specific area, not only because of the presence of unrecorded illegal immigrants, but also because some immigrants move on without informing the local authorities of their departure.

We solve the aforementioned endogeneity issues by using an instrumental variable approach. We select the instrument following Card (2001)[16] and define $Z_{(m,t)}$ in the local labor market m and at time t as:

$$Z_{(m,t)} = \frac{\sum_{c=1}^{N}(\lambda_{mc1991} * Immigrants_{ct})}{Pop_{m1991}}, \tag{2}$$

where λ_{mc1991} is the share of immigrants from country c in the local labor market m in the year 1991, well before when our sample starts (in 2012); $Immigrants_{ct}$ is the total national number of immigrants from country c in year t, $Pop_{m,1991}$ is the total resident population in the local labor market m in the year 1991,[17] whereas N stands for the most common foreign nationalities in the host country over the 2012-2016 period.[18]

Instrument $Z_{(m,t)}$ exploits the fact that immigrants tend to move to areas where a group of immigrants of the same ethnicity has already settled (enclave effect). The expected inflow rate

[16]The same strategy has been adopted by, among others, Barone et al. (2016) and Brunello et al. (2020).

[17]We follow Brunello et al. (2020) and use, as the denominator of our instrument, the population registered in 1991 rather than the current population since the latter is potentially endogenous.

[18]The most common foreign nationalities (we set N equal to 15) in Italy were the following: Albania, Romania, Morocco, China, the Philippines, Tunisia, Serbia and Montenegro, Macedonia, Poland, India, Peru, Senegal, Egypt, Sri Lanka, Ecuador. In 1991, the share of immigrants who originated from the above countries was roughly 60 percent of total immigration.

$Z_{(m,t)}$ is, therefore, a weighted average of the national inflow rates of each of the most common foreign nationalities by countries of origin (the shift) with the weights depending on the 1991 distribution of immigrants in the local labor market.

The exclusion restriction in the IV approach relies on the assumption, conditional on the full set of controls as added in equation (1), that local labor market shocks that attracted immigrants in the past (20 years before our sample period starts) do not correlate with current shocks to and characteristics of the local labor market which affect differences in the behavior of teachers toward students of Italian and immigrant backgrounds. We believe the exclusion restriction holds not only because we include local labor market fixed effects in equation (1) which capture time-invariant local employment opportunities and productivity (see Barone and Mocetti, 2011), but also because 1991 was the year before the signing of the Maastricht Treaty, which reinforced freedom of movement and residence in Europe for European citizens. In particular, the year 1991 also predates the eastwards expansion of 2004 and 2007, when a number of former communist countries, including Bulgaria, Poland and Romania, joined the EU. These important historical changes indicate that past and present local shocks are unlikely to be correlated and this supports the validity of the exclusion restriction (Brunello et al., 2020).

In Table A1 (Panel A) in the Appendix of the paper, we provide direct support for the exclusion restriction by applying the method recently developed by Oster (2019) to the first stage estimates. Since the identification strategy might fail when our instrument is correlated with un-observables, the test establishes bounds to the true value of the first stage parameters under two opposing cases. In the first case, un-observables are not taken into account and the first stage is correctly specified (column 3). In the second case, there are un-observables, but observables and un-observables are equally related to the treatment ($\delta = 1$ in column 1). If a value equal to 0 can be excluded from the bounding set, then accounting for un-observables does not change the direction of our estimates. In our case, zero is excluded from the bounding set, but the inclusion of un-observables would increase the first stage estimates and, therefore, decrease the IV estimates in absolute value.

Moreover, Panel B of Table A1 shows how we follow the test proposed by Conley et al. (2012) and, via the union-of-confidence-intervals approach, build the lower and upper bounds of the 90 percent confidence intervals of the parameter of the interaction term between the dummy variable for immigrant students and the share of immigrants in the LLM. We never find that confidence intervals include zero. This suggests that the direction of our estimate is robust even when we take into account the fact that the exclusion restriction might not hold precisely.

Finally, if some shocks, specific to local labor markets, persisted over time and affected the stock of immigrants in 1991 through long-lasting consequences on both migration patterns and educational outcomes, the exclusion restriction would be violated. To reduce our concerns about this type of violation of the exclusion restriction, we perform a test in the spirit of Mitaritonna et al. (2017), who propose that the trend in the outcome variables before the sample period be regressed on the trend in the instrument during the sample period. While we do not have information on students' evaluation by their teachers in non-blindly-graded tests before 2012, we perform, as in Mitaritonna et al. (2017) and Brunello et al. (2020), an alternative exercise: we test the validity of our instrument, which is only constructed for the second half of the sample period (2014–2016), in relation to trends in the outcome variables in the first part of the sample period (2012–2013). Results reported in Panel C of Table A1 show that there is no statistically significant correlation between the pre-trend in the outcomes and the post-trend in the instrument (see column 1).

A second threat to identification is that local shocks hit LLMs while simultaneously attracting immigrants from countries that had already sent most migrants to those same LLMs before 1991 (Borusyak et al., 2022). To verify whether this is the case, instead of using trends in the outcome variables, we check whether the instrument built in 2012 (the first available year in our sample) correlates with some pre-existing labor market characteristics, e.g. the rate of employment in the local labor market in 1981 (data come from Census data provided by ISTAT). Results show that the instrument is not correlated with any pre-existing trend of the employment rate in the local labor market (see column 2 of Panel C in Table A1).

4 Main results

Before moving to our main estimates, we first present, in Table 2, evidence of the gap between immigrant and native children in teacher assessments in both math and language tests. To be consistent with the main analysis, we estimate a LLM fixed effects model. In columns (1) and (4), in which we do not control for student ability as represented by the student's grade obtained in INVALSI standardized tests, we find that, on average, immigrant children receive a lower score than their peers by 0.286 and 0.421 in math and language, respectively.

After controlling for a linear polynomial of student performance in standardized math or language tests (see columns 2 and 5), the gap decreases by 0.089 points for math and 0.131

points for language, corresponding, in terms of standard deviation, to a downward shift in the gap of 0.024 and 0.035 percentage points for math and language, respectively. Very similar results are also obtained when we include individual performance in both math and language among regressors in order to mitigate the influence of test measurement error (columns 3 and 6). Nothing of note changes when we measure individual ability by using a quadratic or a cubic polynomial of the standardized test score (results available upon request).

[Insert Table 2]

In Table 3, we include the *Share of Immigrants* in the LLM and the interaction term between this variable and the dummy for immigrant students among regressors. In Panel (a) and (b), we report the 2SLS and First stage estimates, respectively. In all specifications, we control for student and school characteristics and for time-variant LLM features and we include interaction terms between our set of control variables and *Immigrant*. This is done in order to ensure that our main explanatory variable of interest does not include the impact of other student and school characteristics correlated with the share of immigrants in the LLM, particularly true for the share of immigrants in the class. We further allow standard errors to be clustered at LLM level and robust to heteroskedasticity.

In columns (1) and (4), for grades assigned by math and language teachers, we include a linear polynomial of student performance in both standardized tests and find that a one standard deviation increase in the share of immigrants in the LLM is associated with a 0.044 (0.053) decrease in immigrant grades in math (language) overall. This effect corresponds to about 1/5 (1/6) for math (language) of the average gap as displayed in columns (1) and (4) of Table 2, between immigrants and natives with identical performance in standardized tests. For grades assigned by language teachers, we find similar findings when we control for a quadratic and cubic polynomial of both math and language standardized test scores, respectively (see column 5 and 6). Conversely, when the outcome is measured by the teacher assigned-grades in math and we control for a quadratic or cubic polynomial of the standardized test scores (columns 2 and 3), the coefficient of our main variable of interest is still negative, but not significant at any conventional level.

In addition, Panel (b) shows a strong "enclave" effect, since the shift-share instrument and its interaction with the *Immigrant* variable correlate positively with both the share of immigrants recorded in the LLM where the school is located and *the Share of Immigrants*Immigrant* variable. Moreover, the *F-statistic* is well above 10, meaning that our estimates do not suffer from the issue of weak instruments.

Instead, in Panel (c) of Table 3, we present OLS estimation results when including LLM fixed effects. When taking into account unobservable time-invariant local heterogeneity, without handling endogeneity and measurement error issues, the interaction term between the share of immigrants in the LLM and the *Immigrant* dummy is still negative for both math and language (the magnitude of the effect is, though, smaller), but not statistically significant, implying that OLS estimates are downward biased.

Among the control variables included in our model, the share of immigrants in the class deserves particular attention. A number of papers have investigated the impact of the ethnic composition of a class on the performance of both native and immigrant students,[19] but little is known about how the concentration of immigrant students in the class affects the evaluation that teachers give to students who get the same results in standardized tests. Our estimates highlight a positive coefficient which suggests that teachers tend to give higher grades to native students when the share of immigrants in the class is large. Conversely, the interaction term *Share of Immigrants in class*Immigrant* is negative and statistically significant, suggesting that when the share of immigrants in the class increases, immigrant students obtain worse grades than native students of similar ability. However, this evidence should be taken with caution because of endogeneity issues.

[Insert Table 3]

In Table 4, we investigate whether the gap experienced by students from immigrant backgrounds is similar for first and second generation immigrants. This difference might come from the fact that second generation immigrants (being born in the host country) have a relative advantage compared to their first generation colleagues as they do not have to adapt to a new culture and learn a new language. On the other hand, teachers might take into account the additional difficulties faced by first generation immigrants and apply lower standards. Indeed, we find that the gap between immigrants and natives is mainly driven by second generation immigrants. As shown in column (2), language teachers working in areas that have been experiencing large flows of immigrants award almost the same grades to first generation immigrants and natives (the

[19] For instance, Jensen and Rasmussen, (2011) find negative effects of ethnic concentration on both native and immigrant students while Ohinata and Van Ours (2013), Contini (2013) and Schneeweis (2015) find no sizeable effect on native students and negative effects on immigrant students. As regards Italy, Tonello (2016) finds a weak negative effect on native students test scores, while, by exploiting rules of class formation, Ballatore et al. (2018) find substantial adverse effects.

coefficient of *Share of Immigrants* Immigrant (I generation)* is far from being statistically significant), while they give relatively lower grades to second generation immigrants (column 4). As regards math, when we distinguish between first and second generation students, we lose power and effects turn out not to be statistically significant.[20] These results are in line with those found by Alesina et al. (2018) who show that language teachers only tend to be biased towards second generation pupils, probably because they expect less from non-native speakers.

[Insert Table 4]

We have also investigated whether the effect is heterogeneous for male and female immigrant students and whether it is related to their socio-economic background. As shown in Table A2 in the Appendix, we find that female immigrant students tend to be penalized more by their teachers in both language and math when the share of immigrants in the population increases. As regards socio-economic background (measured by using the ESCS index proposed by INVALSI), we split the sample according to the median value of the ESCS index and show that the impact of interest is higher for immigrant students whose families are relatively better off (see Table A3 in the Appendix of the paper). This is in line with results found for first and second generation immigrants and could depend on the fact that teachers have lower expectations of students from poorer families.

In general, our findings show that larger flows of immigrants into the local labor market within which schools are located systematically shape the way immigrant students are evaluated by their teachers when compared to natives.

5 Robustness checks

As a first check for the robustness of our empirical exercise, we add in estimates reported in Table 5 school (columns 1 and 2) and class (columns 3 and 4) fixed effects and find that the coefficient of our variable of main interest is negative and statistically significant at the 1 percent level for both math and language, with a larger effect for language. Furthermore, in columns (5) and (6), rather than considering the share of immigrants recorded in the LLM, we build this variable at municipal level (the municipality in which the school is located) and include municipality fixed effects in our model. The coefficient of our main variable of interest, i.e. the interaction between

[20] *Immigrant (I generation)* - in columns (1) and (2) - and *Immigrant (II generation)* - in columns (3) and (4) - are dummy variables taking the value 1 for first or second generation immigrants, respectively, and 0 for natives. In columns (1)-(2) [(3)-(4)] observations for second [first] generation immigrants are replaced with missing values.

Immigrant and the share of immigrants registered in the municipality where the school is located is again negative and statistically significant at the 1 percent level, although the magnitude of the impact seems to be larger than that found when performing our analysis at LLM level. In the remaining columns (7 and 8), we control for province fixed effects instead and find that the results are in line with those discussed in Table 3.

[Insert Table 5]

As a second check, we replicate in Table A4 the specifications reported in Table 3 and, instead of controlling for the "Cheating-corrected" answers in the standardized tests, we use the Rasch scores as an alternative measure of performance in those tests. In columns (1) and (4), we control for a linear polynomial of the standardized test scores for math and language respectively, whereas, in columns (2) and (5), we add a quadratic and in (3) and (6) a cubic polynomial of the blindly graded-test scores in both subjects. Results remain qualitatively very similar.

Qualitatively, much the same results are also obtained when we replace the set of dummies regarding the educational attainment and occupational status of both students' parents with the indicator of socio-economic status directly computed by INVALSI (see Table A5).

We have also tried to understand whether immigrant students living in areas experiencing sizable increases in immigration flows end up being integrated and behaving differently from their peers in class, so inducing their teachers to give them worse grades. With this aim, we use two waves of the INVALSI data (2013/14 and 2014/15) which provide us with some useful information on student integration. To be more precise, we exploit two sections in the questionnaire that ask a number of questions on bullying and on socialization.[21] In the section on bullying, four questions are asked to ascertain whether the respondents had been victims of bullying or whether they had themselves taken part in bullying behavior. These questions refer to verbal bullying, physical bullying and bullying in terms of isolating individuals.[22] For each of these questions students have to choose between the following answers: 1 (never), 2 (now and then), 3 (weekly), and 4 (daily). We build a dummy variable for each of these answers which takes the value 1 for students who answered either weekly or daily. Then, we use these to generate dummy variables, *Victimization* and *Bullying*, if a student has been bullied or has bullied others weekly or daily

[21]The questionnaire has changed over time and questions on social interactions and bullying were not proposed in the other waves.

[22]This is the list of questions for bullying (victimization). This school year, how often have you: Bullied/hassled (by) other students at school by making fun of them? Bullied/hassled (by) other students at school by insulting them? Bullied/hassled (by) other students at school by isolating them? Bullied/hassled (by) other students at school by beating them?

in at least one way.[23] We proceed in a similar way when considering questions relating to social relationships with classmates.[24] The possible answers to these questions were "none", "few", "some", "many" or "all". We build a dummy variable which takes a value of 1 for the first two answers and 0 otherwise.

All these variables are then used alternatively as outcome variables in the regression model specified in equation (1). As highlighted in Table A6, we do not find evidence that students from an immigrant background have a higher probability of experiencing victimization or bullying in areas with larger immigration flows (the interaction term *Share of Immigrants*Immigrant* is never statistically significant). Similar results are found when looking at social relationships with classmates. As an additional check, we have included these dummy variables among the regressors in the model described in equation (1): results are qualitatively very similar to those obtained without these controls (see Table A7).

All in all, these results suggest that we do not have evidence to support the idea that teacher grading is related to the behavior of immigrant students living in areas which experience sizable changes in immigration flows.

Finally, we implement a placebo test to assess whether there is a teacher bias towards other groups of students, e.g. low achieving natives, that varies with the share of immigrants in the LLM. With this aim, we first build a dummy variable *Low-achieving* which takes a value of 1 if a student belongs to the first quartile (below the 25^{th} percentile) of both math and language standardized test score distribution, and 0 otherwise, and then perform the same analysis as before on the subsample of native students. The results reported in Table A8 show that teacher bias as a consequence of the proportion of immigrants in the LLM does not change towards this population subgroup, i.e. the coefficient of *Share of Immigrants*Low-achieving* is negative, but not statistically significant at any conventional level.

6 Mechanisms

In this section, we try to understand the mechanisms underlying our results. In order to support the idea that some teachers react negatively to a rise in the numbers of immigrants in their area and that this is reflected in the way they grade immigrant students, we explore variations

[23] We obtain the same results when we consider separately each of the dummy variables that refer to the different forms of victimization and bullying as outcome variable.

[24] We use these questions: How many of your classmates do you get on well with?; How many of your classmates do you consider as your friends? See also https://invalsireaprove.cineca.it/docs/attach/05˙Questionario˙STAMPA.pdf

in both the municipal setting, as this is potentially associated with a greater hostility towards immigrants, and teachers' characteristics that might, in turn, affect their tendency to holding negative stereotypes.

As regards, the former, we start by considering the size of the cities in which schools are located. According to a number of papers analyzing the impact of immigration on political attitudes and voting (Dustmann et al., 2019; Barone et al., 2016), immigration flows especially lead to an increase in the votes for far-right parties in small towns, while in large cities there is little impact. This may be for a number of reasons, such as a higher standard of education of natives and a longer history of immigration and diversity in large cities. In Table 6, we run separate regressions for cities in the LLM which have an average population below or above the median of the distribution (19,909 inhabitants).[25] We focus on math in columns (1) and (2) and on language in columns (3) and (4) and find a negative and significant effect in smaller cities, whereas there is essentially no effect in large urban areas.

[Insert Table 6]

When we look at first and second generation immigrant students separately (Table A9), again we do not find any impact in larger towns, while in smaller towns we uncover a grade penalty, especially in language, for second generation immigrants. These results are in line with those found by Alesina et al. (2018) who show that while math teachers tend to be biased against both first and second generation immigrants, language teachers behave in such a way only with respect to second generation pupils, probably because they expect less from non-native speakers.

As one of the reasons which might explain the behavior of teachers living in larger towns is a higher level of education and awareness, we show estimation results in Table 7 for when our sample is split according to the percentiles of the average educational attainment of the local population in the LLM. Our findings support our conjecture since the effect of interest is stronger for both math and language in the first quartile (below the 25^{th} percentile) of the educational attainment distribution.

[Insert Table 7]

[25] Looking at smaller municipalities can also help limit measurement errors in perceived changes in the share of immigrants in the geographic area. Indeed, much literature (see among others, Cutler et al., 1999; Boeri et al., 2012) has found that, in big cities, segmented neighborhoods are more common. As immigrant density often shows great variability within the same city, teachers might have different perceptions according to whether they work or live in areas with a high or low presence of immigrants.

In Table 8, we report separate estimates for regions in the Center-North and in the South of Italy. It is well known that immigration into Italy started first in the Center-North. Indeed, in 1999, the first year of data available from official sources (ISTAT), immigrants made up about 1.8 percent of the population (considerably lower than that found on average in the 2012-2016 period) and this was higher in the center-northern regions (2.3 percent) than in the South (0.8 percent). Hence, as people living in the Center-North, had been exposed to the presence of immigrants for a longer period of time, they might have acquired more accommodating attitudes towards foreigners (Barone et al., 2016). Indeed, we find that an increase in the share of immigrants in the local population penalizes immigrant students living in center-northern regions to a lesser extent than it does those attending schools in the South. Our findings are in line with those highlighted by Bursztyn et al. (2021), who show that greater long-term exposure to any given immigrant group (including Arab-Muslims) decreases both explicit and implicit prejudice by natives against foreigners.

[Insert Table 8]

We provide further evidence of the potential role played by teacher characteristics, as we expect older and less educated teachers to show different attitudes towards immigrants. Even though Italian data do not allow us to match teachers' characteristics to the classes they teach, we are able to obtain information at school level on the average age and qualification of teachers from the INVALSI *School Questionnaire*.[26] In particular, in Table 9, we rank schools according to the average age of their teachers and split our sample above and below the median. Results suggest that the negative impact of an increase in the share of immigrants in the local population on grades obtained by immigrant students is particularly marked in LLMs where teachers are, on average, older. This could be related to generational differences in tolerance towards immigrants (with older cohorts being less tolerant) or to the fact that people become less tolerant towards outsiders as they get older (Card et al., 2012).

[Insert Table 9]

Previous literature has shown that more educated individuals have more liberal and positive attitudes towards immigrants. In order to investigate this aspect, we consider the average number of years of education of the teachers in each school. Teachers in Italian primary schools are

[26] The questionnaire is completed by teachers who are randomly sampled by INVALSI. Hence, we build indicators of both the average age and educational attainment of teachers at LLM level. However, this questionnaire is only available for the last three waves, i.e. the 2014-2015, 2015-2016 and 2016-2017 school-years.

required to obtain a university degree in the field of "Education" or a lower qualification if first employed before the 2001-2002 school year. The average number of years of education of math (language) teachers in our sample is 14.37 (14.57) with a standard deviation of 1.22 (1.27). Even though teachers' level of education negatively correlates with their age, the correlation is not very strong.[27] As shown in Table 10, the effect of an increase in the share of immigrants in the LLM on scores obtained in non-blindly-graded tests by immigrant children is larger in schools employing less qualified teachers.

[Insert Table 10]

To better understand the driving forces behind our results we have also tried to investigate whether teacher response to an increase in the share of immigrants in the local population is affected by the age of their students and/or by statistical discrimination. Teachers might rely on prior information about the ability of immigrant students and these expectations might also be affected by the presence of immigrants in the local area.[28] Nonetheless, the relevance of these aspects should decrease with the amount of "classroom-interaction" time the teacher and a given student experience. Since teachers in Italian primary schools typically follow students from the 1^{st} to the 5^{th} grade, statistical discrimination would mean a larger gap for 2^{nd} graders as the time students have spent with their teachers is less than that with 5^{th} graders. Using data on 2^{nd} graders from the school-years 2012/13 to 2016/17,[29] we find that, on average, immigrant children score 0.311 and 0.475 less than natives in math and language, respectively (which is similar to that observed for 5^{th} graders).[30]

As shown in Table 11, where we replicate the specifications of our model reported in Table 3 on the sample of 2^{nd} graders, an increase of a one standard deviation in the share of immigrants relates to a 0.078 (0.104) decrease in immigrant grades in math (language) on average. This effect is very similar to that found for 5^{th} graders, suggesting that teachers' behavior does not depend on the age of their students and/or on the amount of "classroom-interaction" time. However, we acknowledge that teachers in the 2^{nd} grade might have already accumulated sufficient knowledge

[27] The correlation between age and educational attainment of math teachers is -0.247, whereas we find a correlation of -0.249 for language teachers.

[28] The effect might be due to statistical discrimination if the expected ability of immigrant students negatively correlates with the presence of immigrants in the area. This might be the case if an increase in the share of immigrants in the LLM is driven by an upward change in migration inflows from poorer countries.

[29] These data provide the same information as that used for 5^{th} graders with the exception of variables regarding individual feelings towards the standardized test. Indeed, due to their age, 2^{nd} graders are not required to complete the *Student Questionnaire*, i.e. the survey administered to 5^{th} graders on the same day as one of the two tests is taken, which covers these aspects.

[30] We estimate the gap between immigrants and natives attending the 2^{nd} grade by replicating specification (1) and (4) of Table 2. Results available upon request.

of their students since they have interacted with these students for two academic years on a daily basis before the INVALSI test is taken, and, in turn, our evidence might be insufficient to prove that the lower grades obtained by immigrant children from their teachers do not depend on statistical discrimination.[31]

[Insert Table 11]

7 Concluding remarks

From a theoretical point of view, a change in the density of the immigrant population in a given area might positively/negatively affect the attitudes the host community has towards immigrants. On the one hand, "intergroup contact" theory predicts a positive reaction due to closer contact between immigrants and natives in areas with a higher share of immigrants. On the other hand, according to the "group threat" hypothesis, immigration may be perceived as a threat in local contexts characterized by a high number of immigrants. While a significant amount of empirical literature considers the impact of a concentration of immigrants on the political views of native voters and, more generally, on attitudes towards immigrants, less is known about whether these negative feelings translate into discriminatory behavior. We have provided some evidence of such behavior at school level, a particularly crucial environment where exposure to teachers' negative stereotypes can have disastrous consequences on the education and professional careers of immigrants.

Using data from several cohorts of students attending primary schools in Italy, we investigated whether an increase in the share of immigrants in the local population induces teachers to assist native students in their evaluations. We estimated a local labor market fixed effects model and, since the share of immigrants in a given geographical area is not exogenous, we handled endogeneity problems by using an instrumental variable approach that relies on pre-existing settlement patterns.

In line with previous evidence, after controlling for performances in standardized blindly-graded tests, we found that immigrant children receive lower teacher-assigned grades than natives. We also found that the grade penalty suffered by children from immigrant backgrounds is particularly driven by the behavior of teachers working in areas with large recent inflows of

[31] As shown in Table A10, we also find a similar impact for 8^{th} graders, who, given the Italian school system, have interacted with their teachers for three years. In particular, a one standard deviation increase in the share of immigrants is associated with a 0.039 (0.049) decrease in immigrant grades in math (language) overall. This effect corresponds to about 1/4 and 1/5 of the average gap between immigrants and natives for math (0.179) and language (0.232), respectively.

migrants. Results are robust to the inclusion of class fixed effects and of alternative measures of performance in blindly-graded tests among regressors and when we conduct our analysis at municipal level instead of considering the density of immigration in the local labor market where the school is located.

In order to shed light on the mechanisms that might drive this effect, we investigated whether the impact of interest is heterogeneous due to differences in both teacher characteristics and the areas where teachers live and work. We found that older and less qualified teachers tend to react more to increases in the share of immigrants in the local population, giving lower grades to immigrants than to their classmates with the same level of ability. In addition, the effect is greater in schools located in smaller communities with lower overall levels of education, while it is less pronounced in areas where residents have interacted with immigrants over a longer period of time.

Our results confirm previous findings which show that teachers' evaluations are often biased against minority students. While the penalization, in terms of negative evaluations, experienced by immigrant students could reflect facets of knowledge not captured by standardized tests, we also show that teacher bias is related to exogenous changes in the number of immigrants in the local population, which are not linked to differences between native and immigrant children in terms of behavior or skills. We also find that this bias is more pronounced in southern areas of Italy and among certain groups of teachers (i.e. older and less qualified).

Since teachers' negative assessments of immigrant children could produce significant effects on these students' future educational and employment prospects (see Lavy and Sand, 2018), policies aimed at addressing this bias might have a decidedly beneficial effect on both individual and social welfare. According to Ministry of Education directives, in order to boost integration at school, principals need to introduce: (1) specific policies and measures to support children from immigrant backgrounds, such as language and peer support through a mentor (peer mentoring); (2) measures to support newly arrived immigrant students, including assessment of prior knowledge or the creation of closer liaison between parents and schools; and (3) policies and measures aimed at promoting education and intercultural dialogue, as well as those aimed at improving the skills of teachers in all the aforementioned areas. Our findings may help raise awareness of unconscious behavior that can harm immigrant students and highlight the importance of policies aimed at informing teachers of the misplaced biases they might hold towards immigrant groups.

References

Alesina, A., Carlana, M., Ferrara, E. L., and Pinotti, P. (2018a). Revealing stereotypes: Evidence from immigrants in schools. Technical report, National Bureau of Economic Research.

Alesina, A., Miano, A., and Stantcheva, S. (2018b). Immigration and redistribution. Technical report, National Bureau of Economic Research.

Allport, G. W., Clark, K., and Pettigrew, T. (1954). The nature of prejudice. Addison-Wesley Reading, MA.

Angrist, J. D., Battistin, E., and Vuri, D. (2017). In a small moment: Class size and moral hazard in the Italian Mezzogiorno. *American Economic Journal: Applied Economics*, 9(4):216–49.

Ballatore, R. M., Fort, M., and Ichino, A. (2018). Tower of babel in the classroom: Immigrants and natives in italian schools. *Journal of labor economics*, 36(4):885–921.

Barone, G., D'Ignazio, A., de Blasio, G., and Naticchioni, P. (2016). Mr. rossi, mr. hu and politics. The role of immigration in shaping natives' voting behavior. *Journal of Public Economics*, 136:1–13.

Barone, G. and Mocetti, S. (2011). With a little help from abroad: The effect of low-skilled immigration on the female labour supply. *Labour Economics*, 18(5):664–675.

Becker, S. O., Fetzer, T., et al. (2016). Does migration cause extreme voting? *Center for Competitive Advantage in the Global Economy and The Economic & Social Research Council*, pages 1–54.

Bertoni, M., Brunello, G., and Rocco, L. (2013). When the cat is near, the mice won't play: The effect of external examiners in Italian schools. *Journal of Public Economics*, 104:65–77.

Boeri, T., De Philippis, M., Patacchini, E., and Pellizzari, M. (2012). Moving to segregation: Evidence from 8 Italian cities. *IZA Discussion Paper.*

Borusyak, K., Hull, P., and Jaravel, X. (2022). Quasi-experimental shift-share research designs. *The Review of Economic Studies*, 89(1):181–213.

Botelho, F., Madeira, R. A., and Rangel, M. A. (2015). Racial discrimination in grading: Evidence from Brazil. *American Economic Journal: Applied Economics*, 7(4):37–52.

Bracco, E., De Paola, M., Green, C. P., and Scoppa, V. (2022). The spillover of anti-immigration politics to the schoolyard. *Labour Economics. https://doi.org/10.1016/j.labeco.2022.102141.*

Brunello, G., Lodigiani, E., and Rocco, L. (2020). Does low skilled immigration increase the education of natives? Evidence from Italian provinces. *Labour Economics*, 63:101794.

Brunner, B. and Kuhn, A. (2018). Immigration, cultural distance and natives' attitudes towards immigrants: Evidence from Swiss voting results. *Kyklos*, 71(1):28–58.

Burgess, S. and Greaves, E. (2013). Test scores, subjective assessment, and stereotyping of ethnic minorities. *Journal of Labor Economics*, 31(3):535–576.

Bursztyn, L., Chaney, T., Hassan, T. A., and Rao, A. (2021). The immigrant next door: Exposure, prejudice, and altruism. *National Bureau of Economic Research Working Paper.*

Campodifiori, E., Figura, E., Papini, M., and Ricci, R. (2010). *Un indicatore di status socio-economico-culturale degli allievi della quinta primaria in Italia.* Invalsi.

Card, D. (2001). Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. *Journal of Labor Economics*, 19(1):22–64.

Card, D., Dustmann, C., and Preston, I. (2012). Immigration, wages, and compositional amenities. *Journal of the European Economic Association*, 10(1):78–119.

Carlana, M. (2019). Implicit stereotypes: Evidence from teachers' gender bias. *The Quarterly Journal of Economics*, 134(3):1163–1224.

Conley, T. G., Hansen, C. B., and Rossi, P. E. (2012). Plausibly exogenous. *Review of Economics and Statistics*, 94(1):260–272.

Contini, D. (2013). Immigrant background peer effects in Italian schools. *Social science research*, 42(4):1122–1142.

Cornwell, C., Mustard, D. B., and Van Parys, J. (2013). Noncognitive skills and the gender disparities in test scores and teacher assessments: Evidence from primary school. *Journal of Human resources*, 48(1):236–264.

Cutler, D. M., Glaeser, E. L., and Vigdor, J. L. (1999). The rise and decline of the American ghetto. *Journal of Political Economy*, 107(3):455–506.

Di Liberto, A. and Casula, L. (2016). Teacher assessments versus standardized tests: Is acting'girly'an advantage? *IZA Discussion Paper*.

Dustmann, C., Vasiljeva, K., and Piil Damm, A. (2019). Refugee migration and electoral outcomes. *The Review of Economic Studies*, 86(5):2035–2091.

Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. *The handbook of social psychology*, 2(4):357–411.

Halla, M., Wagner, A. F., and Zweimüller, J. (2017). Immigration and voting for the far right. *Journal of the European Economic Association*, 15(6):1341–1385.

Hanna, R. N. and Linden, L. L. (2012). Discrimination in grading. *American Economic Journal: Economic Policy*, 4(4):146–68.

Harmon, N. A. (2018). Immigration, ethnic diversity, and political outcomes: Evidence from Denmark. *The Scandinavian Journal of Economics*, 120(4):1043–1074.

Heckman, J. J. (2008). The case for investing in disadvantaged young children. *CESifo DICE Report*, 6(2):3–8.

Hinnerich, B. T., Höglin, E., and Johannesson, M. (2011). Are boys discriminated in Swedish high schools? *Economics of Education review*, 30(4):682–690.

Hong, G. and McLaren, J. (2015). Are immigrants a shot in the arm for the local economy? Technical report, National Bureau of Economic Research.

Jensen, P. and Rasmussen, A. W. (2011). The effect of immigrant concentration in schools on native and immigrant children's reading and math skills. *Economics of Education Review*, 30(6):1503–1515.

Johnston, D. W. and Lordan, G. (2012). Discrimination makes me sick! An examination of the discrimination–health relationship. *Journal of Health Economics*, 31(1):99–111.

Johnston, D. W. and Lordan, G. (2016). Racial prejudice and labour market penalties during economic downturns. *European Economic Review*, 84:57–75.

Lavy, V. (2008). Do gender stereotypes reduce girls' or boys' human capital outcomes? Evidence from a natural experiment. *Journal of Public Economics*, 92(10-11):2083–2105.

Lavy, V. and Megalokonomou, R. (2017). Persistency in teachers' grading biases and effect on longer term outcomes: University admission exams and choice of field of study. *Unpublished Manuscript*.

Lavy, V. and Sand, E. (2018). On the origins of gender gaps in human capital: Short-and long-term consequences of teachers' biases. *Journal of Public Economics*, 167:263–279.

Manacorda, M., Manning, A., and Wadsworth, J. (2012). The impact of immigration on the structure of wages: Theory and evidence from Britain. *Journal of the European economic association*, 10(1):120–151.

Mayda, A. M., Peri, G., and Steingress, W. (2020). The political impact of immigration: Evidence from the United States. *American Economic Journal: Applied Economics (forthcoming)*.

Mitaritonna, C., Orefice, G., and Peri, G. (2017). Immigrants and firms' outcomes: Evidence from France. *European Economic Review*, 96:62–82.

MIUR (2020). Gli alunni con cittadinanza non italiana a.s. 2018/2019. Technical report, Gestione Patrimonio Informativo e Statistica.

Moriconi, S., Peri, G., and Turati, R. (2019). Immigration and voting for redistribution: Evidence from European elections. *Labour Economics*, 61:101765.

Ohinata, A. and Van Ours, J. C. (2013). How immigrant children affect the academic achievement of native dutch children. *The Economic Journal*, 123(570):F308–F331.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 37(2):187–204.

Ottaviano, G. I. and Peri, G. (2012). Rethinking the effect of immigration on wages. *Journal of the European Economic Association*, 10(1):152–197.

Otto, A. H. and Steinhardt, M. F. (2014). Immigration and election outcomes — evidence from city districts in Hamburg. *Regional Science and Urban Economics*, 45:67–79.

Papageorge, N. W., Gershenson, S., and Kang, K. M. (2020). Teacher expectations matter. *Review of Economics and Statistics*, 102(2):234–251.

Peri, G. (2012). The effect of immigration on productivity: Evidence from US states. *Review of Economics and Statistics*, 94(1):348–358.

Quillian, L. (1995). Prejudice as a response to perceived group threat: Population composition and anti-immigrant and racial prejudice in Europe. *American sociological review*, pages 586–611.

Ricci, R. (2010). The economic, social, and cultural background: A continuous index for the Italian students of the fifth grade. *Atti Convegno SIS-Società Italiana di Statistica*.

Rosenthal, R. and Jacobson, L. (1968). Pygmalion in the classroom. *The Urban Review*, 3(1):16–20.

Schlueter, E. and Wagner, U. (2008). Regional differences matter: Examining the dual influence of the regional size of the immigrant population on derogation of immigrants in Europe. *International Journal of Comparative Sociology*, 49(2-3):153–173.

Schneeweis, N. (2015). Immigrant concentration in schools: Consequences for native and migrant students. *Labour Economics*, 35:63–76.

Sequeira, S., Nunn, N., and Qian, N. (2017). Migrants and the making of America: The short-and long-run effects of immigration during the age of mass migration. Technical report, National Bureau of Economic Research.

Sprietsma, M. (2013). Discrimination in grading: Experimental evidence from primary school teachers. *Empirical economics*, 45(1):523–538.

Tabellini, M. (2020). Gifts of the immigrants, woes of the natives: Lessons from the age of mass migration. *The Review of Economic Studies*, 87(1):454–486.

Terrier, C. (2015). Giving a little help to girls? Evidence on grade discrimination and its effect on students' achievement. cep discussion paper no. 1341. *Centre for Economic Performance*.

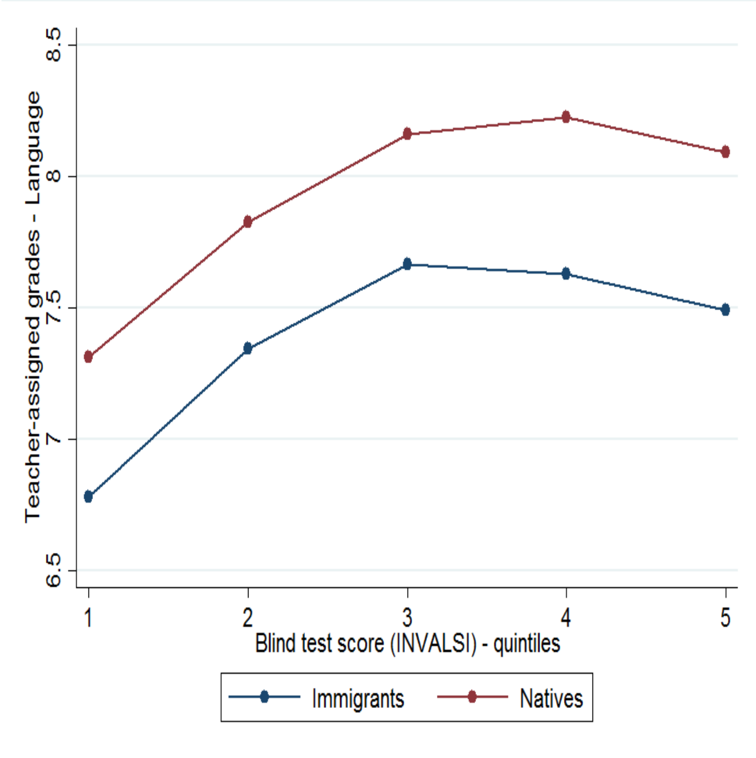
Tonello, M. (2016). Peer effects of non-native students on natives' educational outcomes: Mechanisms and evidence. *Empirical Economics*, 51(1):383–414.

Van Ewijk, R. (2011). Same work, lower grade? Student ethnicity and teachers' subjective assessments. *Economics of Education Review*, 30(5):1045–1058.

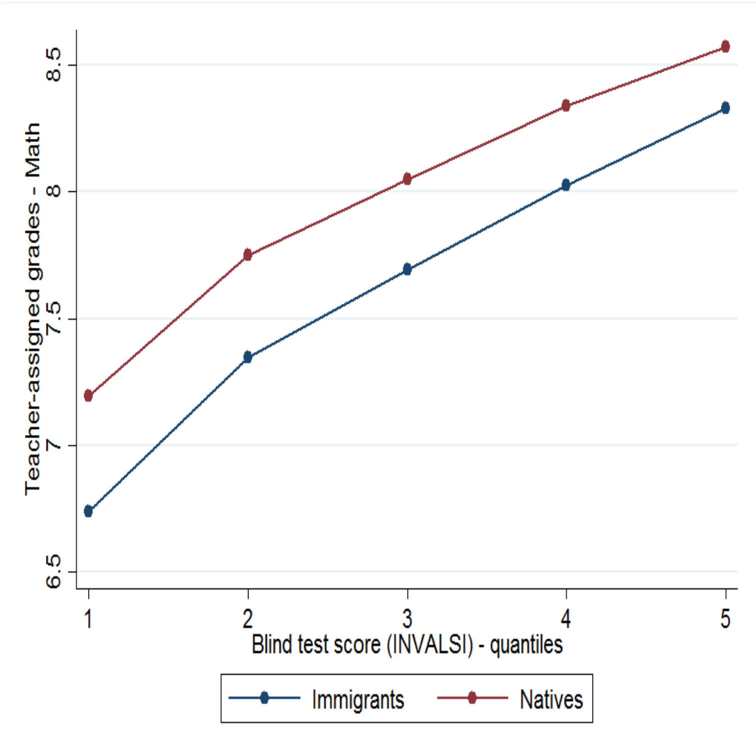
Wagner, U., Christ, O., Pettigrew, T. F., Stellmacher, J., and Wolf, C. (2006). Prejudice and minority proportion: Contact instead of threat effects. *Social psychology quarterly*, 69(4):380–390.

List of figures and tables

Figure 1: Teacher-assigned grades vs standardized test scores



(a) Language



(b) Math

Table 1: Descriptive statistics - 5^{th} graders

Variables	Obs	Mean	St. dev.	Min	Max
Outcomes					
Language score	1,219,572	7.873	1.048	5	10
Math score	1,217,165	7.943	1.097	5	10
Cheating-corrected Language score	1,219,572	34.263	16.181	0.352	82
Cheating-corrected Math score	1,218,620	25.404	8.721	0.312	50
Rasch Language score	1,219,572	0.272	1.026	-5.427	4.428
Rasch Math score	1,218,649	0.361	1.156	-5.495	4.924
Students' characteristics					
Immigrant	1,219,572	0.084	0.274	0	1
Immigrant (I generation)	1,146,188	0.023	0.151	0	1
Immigrant (II generation)	1,193,152	0.061	0.241	0	1
Female	1,219,572	0.501	0.5	0	1
Regular	1,219,572	0.971	0.168	0	1
Kindergarten	1,219,572	0.338	0.473	0	1
Pre-primary	1,219,572	0.959	0.199	0	1
Full day	1,219,572	0.107	0.309	0	1
Class size	1,219,572	20.402	4.333	1	37
School size	1,219,572	5.251	2.016	1	13
Share of female students in class	1,219,572	0.505	0.117	0	1
Share of Immigrants in class	1,219,572	0.101	0.12	0	1
Father Primary school diploma	1,219,572	0.028	0.164	0	1
Father Middle school diploma	1,219,572	0.347	0.476	0	1
Father High school diploma	1,219,572	0.479	0.5	0	1
Father Bachelor degree	1,219,572	0.146	0.354	0	1
Mother Primary school diploma	1,219,572	0.023	0.151	0	1
Mother Middle school diploma	1,219,572	0.27	0.444	0	1
Mother High school diploma	1,219,572	0.525	0.499	0	1
Mother Bachelor degree	1,219,572	0.182	0.386	0	1
Mother Unemployed	1,219,572	0.053	0.224	0	1
Mother Housewife	1,219,572	0.337	0.473	0	1
Mother Manager	1,219,572	0.012	0.11	0	1
Mother Entrepreneur	1,219,572	0.018	0.134	0	1
Mother Professional	1,219,572	0.092	0.288	0	1
Mother Retailer	1,219,572	0.078	0.268	0	1
Mother Teacher	1,219,572	0.274	0.446	0	1
Mother Workman	1,219,572	0.135	0.342	0	1
Mother Retired	1,219,572	0.001	0.032	0	1
Father Unemployed	1,219,572	0.054	0.227	0	1
Father Homemaker	1,219,572	0.003	0.056	0	1
Father Manager	1,219,572	0.035	0.184	0	1
Father Entrepreneur	1,219,572	0.059	0.235	0	1
Father Professional	1,219,572	0.139	0.345	0	1
Father Retailer	1,219,572	0.207	0.405	0	1
Father Teacher	1,219,572	0.198	0.399	0	1
Father Workman	1,219,572	0.297	0.457	0	1
Father Retired	1,219,572	0.008	0.087	0	1
ESCS	1,218,619	0.171	0.979	-3.261	2.6
Invalsi students' questionnaire					
Worried (Answer: Not at all)	1,219,572	0.158	0.365	0	1
Worried (Answer: A bit)	1,219,572	0.308	0.462	0	1
Worried (Answer: Moderately)	1,219,572	0.304	0.459	0	1
Worried (Answer: Very much)	1,219,572	0.229	0.421	0	1
Nervous during the test (Answer: Not at all)	1,219,572	0.181	0.384	0	1
Nervous during the test (Answer: A bit)	1,219,572	0.273	0.445	0	1
Nervous during the test (Answer: Moderately)	1,219,572	0.338	0.473	0	1
Nervous during the test (Answer: Very much)	1,219,572	0.209	0.407	0 1	
Low Performance Expectation (Answer: Not at all)	1,219,572	0.229	0.419	0	1
Low Performance Expectation (Answer: A bit)	1,219,572	0.347	0.476	0	1
Low Performance Expectation (Answer: Moderately)	1,219,572	0.246	0.431	0	1
Low Performance Expectation (Answer: Very much)	1,219,572	0.178	0.382	0	1
Teachers' characteristics (LLM Indicators)					
Language Teachers' Age	587,026	51.302	4.703	29	65
Math Teachers' Age	584,920	50.392	4.871	32	67
Language Teachers' Education	587,026	14.567	1.271	13	18
Math Teachers' Education	584,920	14.374	1.224	13	18
Local Labor Market characteristics					
Share of Immigrants	1,219,572	0.08	0.04	0.003	0.189
Average Population in LLM	1,219,572	186,781.45	426,399.94	783	2,208,631

Source: Data to build the outcome variables and all the controls about students and parents' characteristics are taken from INVALSI files (waves 2012/13-2016/17). *Share of Immigrants* is taken from ISTAT. The time-variant covariates come from the ISTAT Territorial Statistics.

Table 2: Gap between immigrant and native students. LLM-FE estimates – 5^{th} graders

	(1) Math score	(2) Math score	(3) Math score	(4) Language score	(5) Language score	(6) Language score
Immigrant	-0.286***	-0.197***	-0.154***	-0.421***	-0.290***	-0.291***
	(0.018)	(0.013)	(0.013)	(0.014)	(0.011)	(0.009)
Cheating-corrected Math score		0.064***	0.047***			0.029***
		(0.001)	(0.001)			(0.0001)
Cheating-corrected Language score			0.025***		0.049***	0.033***
			(0.001)		(0.001)	(0.001)
Constant	7.528***	6.317***	6.151***	7.520***	6.567***	6.325***
	(0.037)	(0.044)	(0.047)	(0.033)	(0.039)	(0.039)
Students' characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Family background	Yes	Yes	Yes	Yes	Yes	Yes
Non-cognitive skills	Yes	Yes	Yes	Yes	Yes	Yes
Cheating-corrected Math score	No	Linear	Linear	No	No	Linear
Cheating-corrected Language score	No	No	Linear	No	Linear	Linear
LLM FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,217,165	1,216,215	1,216,215	1,219,572	1,219,572	1,218,620
R-squared	0.168	0.356	0.379	0.204	0.356	0.382
No. LLM	580	580	580	580	580	580

Note: LLM-FE estimates. The dependent variable is measured by the teacher-assigned grades in Math (columns 1-3) and Language (columns 4-6). We control for LLM and year fixed effects and we focus on the waves 2012/13-2016/17. Standard Errors are robust to heteroskedasticity and are clustered at the local labor market level (shown in brackets). Significance at the 10% level is represented by *, at the 5% level by **, and at the 1% level by ***.

Table 3: FE-IV estimates. Math and language score – 5^{th} graders

	(1) FE-IV Math score	(2) FE-IV Math score	(3) FE-IV Math score	(4) FE-IV Language score	(5) FE-IV Language score	(6) FE-IV Language score
Panel (a): 2SLS						
Share of Immigrants*Immigrant	-1.101*	-1.006	-0.989	-1.335***	-1.231**	-1.209**
	(0.667)	(0.687)	(0.692)	(0.514)	(0.536)	(0.541)
Immigrant	-0.193*	-0.196*	-0.196**	-0.336***	-0.342***	-0.342***
	(0.091)	(0.091)	(0.089)	(0.087)	(0.086)	(0.086)
Cheating-corrected Math score	0.047***	0.055***	0.053***	0.028***	0.031***	0.035***
	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
Cheating-corrected Language score	0.025***	0.035***	0.047***	0.033***	0.046***	0.061***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Share of Immigrants	0.938	0.794	0.777	1.091	0.894	0.869
	(0.841)	(0.772)	(0.768)	(0.862)	(0.773)	(0.766)
Share of Immigrants in class	0.243***	0.254***	0.256***	0.261***	0.272***	0.275***
	(0.043)	(0.043)	(0.043)	(0.022)	(0.044)	(0.044)
Share of Immigrants in class*Immigrant	-0.247***	-0.253***	-0.255***	-0.179***	-0.185***	-0.188***
	(0.049)	(0.049)	(0.049)	(0.047)	(0.047)	(0.047)
Panel (b): First stage						
Instrument	0.407***	0.407***	0.407***	0.407***	0.407***	0.407****
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)
F-stat	104.02	104.24	104.19	104.20	104.41	104.36
p-value	0.000	0.000	0.000	0.000	0.000	0.000
Instrument*Immigrant	0.314***	0.314***	0.314***	0.314***	0.313***	0.314***
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)
F-stat	90.68	88.66	88.51	91.27	89.22	89.06
p-value	0.000	0.000	0.000	0.000	0.000	0.000
Panel (c): LLM-FE estimates						
Share of Immigrants*Immigrant	-0.054	-0.047	-0.061	-0.172	-0.063	-0.045
	(0.276)	(0.283)	(0.283)	(0.223)	(0.231)	(0.231)
Cheating-corrected Math score	Linear	Quadratic	Cubic	Linear	Quadratic	Cubic
Cheating-corrected Language score	Linear	Quadratic	Cubic	Linear	Quadratic	Cubic
Students' characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Family background	Yes	Yes	Yes	Yes	Yes	Yes
Non-cognitive skills	Yes	Yes	Yes	Yes	Yes	Yes
Immigrant*controls	Yes	Yes	Yes	Yes	Yes	Yes
LLM FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,216,212	1,216,212	1,216,212	1,218,617	1,218,617	1,218,617
R-squared	0.381	0.381	0.382	0.383	0.384	0.385
No. LLM	577	577	577	577	577	577

Note: FE-IV estimates. The dependent variable is measured by the teacher-assigned grades in Math (columns 1-3) and Language (columns 4-6). We control for LLM and year fixed effects and we focus on the waves 2012/13-2016/17. Standard Errors are robust to heteroskedasticity and are clustered at local labor market level (shown in brackets). Significance at the 10% level is represented by *, at the 5% level by **, and at the 1% level by ***.

Table 4: FE-IV estimates. Math and language score for I and II generation of immigrants – 5^{th} graders

	(1) FE-IV Math score	(2) FE-IV Language score	(3) FE-IV Math score	(4) FE-IV Language score
Panel (a): 2SLS				
Share of Immigrants*Immigrant (I generation)	-0.752	-0.653		
	(0.586)	(0.502)		
Share of Immigrants*Immigrant (II generation)			-0.949	-1.275***
			(0.764)	(0.548)
Immigrant (I generation)	-0.191***	-0.343***		
	(0.051)	(0.043)		
Immigrant (II generation)			-0.087	-0.172***
			(0.072)	(0.045)
Cheating-corrected Math score	0.053***	0.035***	0.053***	0.035***
	(0.003)	(0.003)	(0.003)	(0.003)
Cheating-corrected Language score	0.048***	0.061***	0.047***	0.061***
	(0.002)	(0.001)	(0.002)	(0.001)
Share of Immigrants	0.696	0.848	0.787	0.838
	(0.769)	(0.767)	(0.764)	(0.762)
Share of Immigrants in class	0.212***	0.254***	0.214***	0.232***
	(0.045)	(0.043)	(0.037)	(0.036)
Share of Immigrants in class*Immigrant	-0.273***	-0.201***	-0.232***	-0.173***
	(0.053)	(0.049)	(0.051)	(0.045)
Panel (b): First stage				
Instrument	0.407***	0.407***	0.407***	0.407***
	(0.028)	(0.028)	(0.028)	(0.028)
F-stat	99.69	99.87	103.27	103.44
p-value	0.000	0.000	0.000	0.000
Instrument*Immigrant	0.361***	0.361***	0.311***	0.316***
	(0.031)	(0.031)	(0.028)	(0.028)
F-stat	73.55	73.86	79.72	80.01
p-value	0.000	0.000	0.000	0.000
Cheating-corrected Math score	Cubic	Cubic	Cubic	Cubic
Cheating-corrected Language score	Cubic	Cubic	Cubic	Cubic
Students' characteristics	Yes	Yes	Yes	Yes
Family background	Yes	Yes	Yes	Yes
Non-cognitive skills	Yes	Yes	Yes	Yes
Immigrant*controls	Yes	Yes	Yes	Yes
LLM FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	1,143,019	1,145,261	1,189,766	1,192,105
R-squared	0.373	0.372	0.376	0.377
No. LLM	577	577	577	577

Note: FE-IV estimates. The dependent variable is measured by the teacher-assigned grades in Math (columns 1 and 3) and Language (columns 2 and 4). We control for LLM and year fixed effects and we focus on the waves 2012/13-2016/17. Standard Errors are robust to heteroskedasticity and are clustered at local labor market level (shown in brackets). Significance at the 10% level is represented by *, at the 5% level by **, and at the 1% level by ***.

Table 5: School, class, municipality, province FE-IV estimates. Math and language score – 5^{th} graders

	(1) FE-IV Math score	(2) FE-IV Language score	(3) FE-IV Math score	(4) FE-IV Language score	(5) FE-IV Math score	(6) FE-IV Language score	(7) FE-IV Math score	(8) FE-IV Language score
Panel (a): 2SLS								
Share of Immigrants*Immigrant	-0.743***	-1.051***	-0.682***	-0.974***	-1.466***	-1.896***	-0.813	-1.118**
	(0.264)	(0.248)	(0.206)	(0.188)	(0.438)	(0.356)	(0.695)	(0.538)
Immigrant	-0.119	-0.265***	-0.141	-0.289***	-0.287***	-0.316***	-0.212**	-0.352***
	(0.097)	(0.085)	(0.092)	(0.081)	(0.097)	(0.093)	(0.101)	(0.092)
Cheated-corrected Math score	0.052***	0.033***	0.055***	0.035***	0.054***	0.035***	0.054***	0.036***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Cheated-corrected Language score	0.048***	0.062***	0.047***	0.061***	0.048***	0.062***	0.048***	0.062***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Share of Immigrants	2.901***	2.711***	1.583*	1.733*	1.194*	0.905	0.411	0.271
	(1.051)	(0.926)	(0.826)	(0.767)	(0.646)	(0.611)	(0.653)	(0.619)
Share of Immigrants in class	0.241***	0.221***	0.241***	0.217***	0.347***	0.401***	0.241***	0.261***
	(0.034)	(0.032)	(0.034)	(0.032)	(0.061)	(0.058)	(0.034)	(0.043)
Share of Immigrants in class*Immigrant	-0.188***	-0.114***	-0.131***	-0.065***	-0.345***	-0.324***	-0.245***	-0.174***
	(0.034)	(0.031)	(0.029)	(0.028)	(0.067)	(0.061)	(0.051)	(0.052)
Panel (b): First stage								
Instrument	0.238***	0.252***	0.373***	0.374***	0.249***	0.314***	0.189***	0.189***
	(0.019)	(0.019)	(0.009)	(0.009)	(0.036)	(0.036)	(0.033)	(0.033)
F-stat	121.63	143.02	831.34	831.58	42.05	42.14	21.60	21.59
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Instrument*Immigrant	0.343***	0.311***	0.312***	0.312***	0.209***	0.209***	0.312***	0.312***
	(0.013)	(0.011)	(0.006)	(0.006)	(0.014)	(0.014)	(0.039)	(0.039)
F-stat	329.05	363.41	1542.28	1544.49	130.40	130.23	37.43	37.39
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Cheating-corrected Math score	Cubic	Cubic	Cubic	Cubic	Cubic	Cubic	Cubic	Cubic
Cheating-corrected Language score	Cubic	Cubic	Cubic	Cubic	Cubic	Cubic	Cubic	Cubic
Students' characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family background	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Non-cognitive skills	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Immigrant*controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	No	No	No	No	No	No
Class FE	No	No	Yes	Yes	No	No	No	No
Municipality FE	No	No	No	No	Yes	Yes	No	No
Province FE	No	No	No	No	No	No	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,215,376	1,218,500	1,215,560	1,217,951	1,215,376	1,217,779	1,216,215	1,218,620
R-squared	0.397	0.392	0.415	0.422	0.391	0.393	0.379	0.383
No. Schools/Classes/Munic./Prov.	6,861	6,861	30,692	30,723	5,959	5,962	107	107

Note: FE-IV estimates. The dependent variable is measured by the teacher-assigned grades in Math (columns 1, 3, 5 and 7) and Language (columns 2, 4, 6 and 8). We control for School FE in columns (1)-(2), for Class FE in columns (3)-(4), for Municipality FE in columns (5)-(6) and for Province FE in columns (7)-(8) and in each specification for year fixed effects. We focus on the waves 2012/13-2016/17. Standard Errors are robust to heteroskedasticity and are clustered at municipal/province level (shown in brackets). Significance at the 10% level is represented by *, at the 5% level by **, and at the 1% level by ***.

Table 6: FE-IV estimates. Heterogeneity by population size. Math and language score – 5^{th} graders

	(1) FE-IV Math score <median	(2) FE-IV Math score >median	(3) FE-IV Language score <median	(4) FE-IV Language score >median
Panel (a): 2SLS				
Share of Immigrants*Immigrant	-1.468***	0.695	-1.509***	-0.045
	(0.466)	(0.984)	(0.419)	(0.616)
Immigrant	-0.245**	-0.094	-0.348***	-0.539***
	(0.126)	(0.144)	(0.115)	(0.152)
Cheating-corrected Math score	0.052***	0.054***	0.033***	0.037***
	(0.003)	(0.004)	(0.003)	(0.004)
Cheating-corrected Language score	0.049***	0.045***	0.063***	0.058***
	(0.002)	(0.002)	(0.002)	(0.002)
Share of Immigrants	1.331	0.557	1.109	0.737
	(1.246)	(1.519)	(1.214)	(1.297)
Share of Immigrants in class	0.174**	0.201***	0.156**	0.265***
	(0.068)	(0.061)	(0.064)	(0.058)
Share of Immigrants in class*Immigrant	-0.091	-0.221***	-0.023	-0.177***
	(0.063)	(0.066)	(0.055)	(0.061)
Panel (b): First stage				
Instrument	0.385***	0.406***	0.386***	0.406***
	(0.031)	(0.036)	(0.031)	(0.036)
F-stat	77.94	87.09	77.97	86.88
p-value	0.000	0.000	0.000	0.000
Instrument*Immigrant	0.289***	0.344***	0.289***	0.344***
	(0.028)	(0.041)	(0.028)	(0.041)
F-stat	66.13	11.70	66.45	11.75
p-value	0.000	0.000	0.000	0.000
Cheating-corrected Math score	Cubic	Cubic	Cubic	Cubic
Cheating-corrected Language score	Cubic	Cubic	Cubic	Cubic
Students' characteristics	Yes	Yes	Yes	Yes
Family background	Yes	Yes	Yes	Yes
Non-cognitive skills	Yes	Yes	Yes	Yes
Immigrant*controls	Yes	Yes	Yes	Yes
LLM FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	607,866	608,342	608,997	609,611
R-squared	0.395	0.369	0.395	0.375
No. LLM	568	294	568	294

Note: FE-IV estimates. The dependent variable is measured by the teacher-assigned grades in Math (columns 1-2) and Language (columns 3-4). We control for LLM and year fixed effects and we focus on the waves 2012/13-2016/17. Standard Errors are robust to heteroskedasticity and are clustered at local labor market level (shown in brackets). Significance at the 10% level is represented by *, at the 5% level by **, and at the 1% level by ***.

Table 7: FE-IV estimates. Heterogeneity by *Education of Population*. Math and language score – 5^{th} graders

	(1) FE-IV Math score	(2) FE-IV Math score	(3) FE-IV Math score	(4) FE-IV Language score	(5) FE-IV Language score	(6) FE-IV Language score
	<25th	>25th and <75th	>75th	<25th	>25th and <75th	>75th
Panel (a): 2SLS						
Share of Immigrants*Immigrant	-1.551***	-0.316	0.913	-1.666**	-0.556	0.192
	(0.507)	(0.928)	(0.754)	(0.578)	(0.553)	(0.552)
Immigrant	-0.504**	-0.373***	-0.193	-0.664***	-0.576***	-0.447***
	(0.211)	(0.144)	(0.164)	(0.186)	(0.115)	(0.147)
Cheating-corrected Math score	0.051***	0.054***	0.062***	0.040***	0.031***	0.041***
	(0.006)	(0.004)	(0.004)	(0.006)	(0.003)	(0.005)
Cheating-corrected Language score	0.054***	0.049***	0.041***	0.066***	0.063***	0.057***
	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)
Share of Immigrants	0.852	-0.831	2.923**	0.216	-0.119	3.018**
	(1.734)	(1.225)	(1.212)	(1.559)	(1.141)	(1.191)
Share of Immigrants in class	0.399***	0.164***	0.181**	0.341***	0.185***	0.225***
	(0.104)	(0.055)	(0.078)	(0.102)	(0.054)	(0.084)
Share of Immigrants in class*Immigrant	-0.399***	-0.117**	-0.228***	-0.315***	-0.067	-0.154*
	(0.116)	(0.052)	(0.086)	(0.111)	(0.049)	(0.079)
Panel (b): First stage						
Instrument	0.311***	0.435***	0.494***	0.311***	0.435***	0.494***
	(0.036)	(0.039)	(0.036)	(0.034)	(0.037)	(0.036)
F-stat	38.58	69.12	93.11	38.68	69.19	93.57
p-value	0.000	0.000	0.000	0.000	0.000	0.000
Instrument*Immigrant	0.304***	0.294***	0.439***	0.305***	0.294***	0.439***
	(0.036)	(0.034)	(0.049)	(0.036)	(0.031)	(0.049)
F-stat	47.11	66.36	53.55	47.17	66.50	53.70
p-value	0.000	0.000	0.000	0.000	0.000	0.000
Cheating-corrected Math score	Cubic	Cubic	Cubic	Cubic	Cubic	Cubic
Cheating-corrected Language score	Cubic	Cubic	Cubic	Cubic	Cubic	Cubic
Students' characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Family background	Yes	Yes	Yes	Yes	Yes	Yes
Non-cognitive skills	Yes	Yes	Yes	Yes	Yes	Yes
Immigrant*controls	Yes	Yes	Yes	Yes	Yes	Yes
LLM FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	303,963	608,022	304,224	304,535	609,344	304,735
R-squared	0.361	0.393	0.397	0.371	0.393	0.383
No. LLM	510	445	176	510	445	176

Note: FE-IV estimates. The dependent variable is measured by the teacher-assigned grades in Math (columns 1-3) and Language (columns 4-6). We control for LLM and year fixed effects and we focus on the waves 2012/13-2016/17. Standard Errors are robust to heteroskedasticity and are clustered at local labor market level (shown in brackets). Significance at the 10% level is represented by *, at the 5% level by **, and at the 1% level by ***.

Table 8: FE-IV estimates. Heterogeneity by geographic area. Math and language score – 5^{th} graders

	(1) FE-IV Math score South	(2) FE-IV Math score Center-North	(3) FE-IV Language score South	(4) FE-IV Language score Center-North
Panel (a): 2SLS				
Share of Immigrants*Immigrant	-4.099***	-1.124	-4.174***	-1.492***
	(1.077)	(0.897)	(1.056)	(0.677)
Immigrant	-0.625*	-0.105	-0.617*	-0.327***
	(0.361)	(0.109)	(0.331)	(0.092)
Cheating-corrected Math score	0.047***	0.062***	0.041***	0.034***
	(0.005)	(0.003)	(0.005)	(0.003)
Cheating-corrected Language score	0.057***	0.039***	0.064***	0.058***
	(0.005)	(0.002)	(0.002)	(0.003)
Share of Immigrants	2.557	-0.564	3.048	-0.569
	(2.349)	(0.745)	(2.645)	(0.686)
Share of Immigrants in class	0.266	0.228***	0.254	0.239***
	(0.201)	(0.048)	(0.198)	(0.042)
Share of Immigrants in class*Immigrant	-0.447***	-0.198***	-0.395**	-0.122***
	(0.201)	(0.053)	(0.198)	(0.045)
Panel (b): First stage				
Instrument	0.275***	0.453***	0.275***	0.453***
	(0.034)	(0.027)	(0.034)	(0.027)
F-stat	33.05	133.37	33.16	137.47
p-value	0.000	0.000	0.000	0.000
Instrument*Immigrant	0.299***	0.241***	0.301***	0.241***
	(0.049)	(0.025)	(0.049)	(0.025)
F-stat	45.15	102.83	46.77	103.06
p-value	0.000	0.000	0.000	0.000
Cheating-corrected Math score	Cubic	Cubic	Cubic	Cubic
Cheating-corrected Language score	Cubic	Cubic	Cubic	Cubic
Students' characteristics	Yes	Yes	Yes	Yes
Family background	Yes	Yes	Yes	Yes
Non-cognitive skills	Yes	Yes	Yes	Yes
Immigrant*controls	Yes	Yes	Yes	Yes
LLM FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	404,795	811,417	405,831	812,786
R-squared	0.333	0.419	0.351	0.414
No. LLM	282	299	282	299

Note: FE-IV estimates. The dependent variable is measured by the teacher-assigned grades in Math (columns 1-2) and Language (columns 3-4). We control for LLM and year fixed effects and we focus on the waves 2012/13-2016/17. Standard Errors are robust to heteroskedasticity and are clustered at local labor market level (shown in brackets). Significance at the 10% level is represented by *, at the 5% level by **, and at the 1% level by ***.

Table 9: FE-IV estimates. Heterogeneity by the average age of teachers. Math and language score – 5^{th} graders

	(1) FE-IV Math score	(2) FE-IV Math score	(3) FE-IV Language score	(4) FE-IV Language score
	<median	>median	<median	>median
Panel (a): 2SLS				
Share of Immigrants*Immigrant	0.059	-1.417***	-0.555	-1.057*
	(1.208)	(0.525)	(0.717)	(0.644)
Immigrant	-0.172	-0.172	-0.617***	-0.034
	(0.161)	(0.211)	(0.144)	(0.142)
Cheating-corrected Math score	0.052***	0.071***	0.071***	0.051***
	(0.008)	(0.005)	(0.005)	(0.001)
Cheating-corrected Language score	-0.003	0.021***	0.031	0.011
	(0.007)	(0.006)	(0.005)	(0.007)
Share of Immigrants	-0.742	0.224	0.121	3.838
	(5.853)	(1.765)	(2.413)	(8.524)
Share of Immigrants in class	0.304***	0.148**	0.231***	0.264***
	(0.065)	(0.066)	(0.069)	(0.059)
Share of Immigrants in class*Immigrant	-0.268***	-0.161**	-0.124*	-0.178**
	(0.079)	(0.067)	(0.074)	(0.074)
Panel (b): First stage				
Instrument*Immigrant	0.261***	0.301***	0.288***	0.297***
	(0.034)	(0.029)	(0.027)	(0.036)
F-stat	31.18	58.88	62.49	34.67
p-value	0.000	0.000	0.000	0.000
Cheating-corrected Math score	Cubic	Cubic	Cubic	Cubic
Cheating-corrected Language score	Cubic	Cubic	Cubic	Cubic
Students' characteristics	Yes	Yes	Yes	Yes
Family background	Yes	Yes	Yes	Yes
Non-cognitive skills	Yes	Yes	Yes	Yes
Immigrant*controls	Yes	Yes	Yes	Yes
LLM FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	291,993	462,231	291,288	464,501
R-squared	0.403	0.385	0.401	0.391
No. LLM	225	559	229	563

Note: FE-IV estimates. The dependent variable is measured by the teacher-assigned grades in Math (columns 1-2) and Language (columns 3-4). We control for LLM and year fixed effects and we focus on the waves 2014/15-2016/17. Standard Errors are robust to heteroskedasticity and are clustered at local labor market level (shown in brackets). Significance at the 10% level is represented by *, at the 5% level by **, and at the 1% level by ***.

Table 10: FE-IV estimates. Heterogeneity by the average education of teachers. Math and language score – 5^{th} graders

	(1) FE-IV Math score <median	(2) FE-IV Math score >median	(3) FE-IV Language score <median	(4) FE-IV Language score >median
Panel (a): 2SLS				
Share of Immigrants*Immigrant	-1.247**	0.821	-1.194**	-0.161
	(0.587)	(1.249)	(0.558)	(0.873)
Immigrant	-0.131	-0.109	-0.739***	-0.274*
	(0.175)	(0.177)	(0.137)	(0.169)
Cheating-corrected Math score	0.067***	0.064***	0.065***	0.062***
	(0.006)	(0.006)	(0.005)	(0.009)
Cheating-corrected Language score	0.009**	0.015**	0.031***	0.008
	(0.005)	(0.008)	(0.005)	(0.007)
Share of Immigrants	-1.544	-1.275	-1.868	3.046
	(2.061)	(6.964)	(2.444)	(6.285)
Share of Immigrants in class	0.203***	0.244***	0.265***	0.227***
	(0.065)	(0.081)	(0.072)	(0.051)
Share of Immigrants in class*Immigrant	-0.183***	-0.237***	-0.197**	-0.092
	(0.065)	(0.079)	(0.078)	(0.059)
Panel (b): First stage				
Instrument*Immigrant	0.285***	0.306***	0.278***	0.318***
	(0.026)	(0.045)	(0.029)	(0.036)
F-stat	62.02	25.43	51.69	40.33
p-value	0.000	0.000	0.000	0.000
Cheating-corrected Math score	Cubic	Cubic	Cubic	Cubic
Cheating-corrected Language score	Cubic	Cubic	Cubic	Cubic
Students' characteristics	Yes	Yes	Yes	Yes
Family background	Yes	Yes	Yes	Yes
Non-cognitive skills	Yes	Yes	Yes	Yes
Immigrant*controls	Yes	Yes	Yes	Yes
LLM FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	461,637	292,593	461,679	294,155
R-squared	0.392	0.392	0.396	0.392
No. LLM	543	296	554	304

Note: FE-IV estimates. The dependent variable is measured by the teacher-assigned grades in Math (columns 1-2) and Language (columns 3-4). We control for LLM and year fixed effects and we focus on the waves 2014/15-2016/17. Standard Errors are robust to heteroskedasticity and are clustered at local labor market level (shown in brackets). Significance at the 10% level is represented by *, at the 5% level by **, and at the 1% level by ***.

Table 11: FE-IV estimates. Math and language score - 2^{nd} graders

	(1) FE-IV Math score	(2) FE-IV Math score	(3) FE-IV Math score	(4) FE-IV Language score	(5) FE-IV Language score	(6) FE-IV Language score
Panel (a): 2SLS						
Share of Immigrants*Immigrant	-2.067***	-2.004***	-1.973***	-2.738***	-2.663***	-2.625***
	(0.501)	(0.501)	(0.505)	(0.454)	(0.457)	(0.461)
Immigrant	-0.169*	-0.181*	-0.186*	-0.317***	-0.324***	-0.331***
	(0.104)	(0.104)	(0.103)	(0.091)	(0.092)	(0.091)
Cheating-corrected Math score	0.063***	0.082***	0.109***	0.052***	0.061***	0.091***
	(0.001)	(0.004)	(0.006)	(0.001)	(0.006)	(0.006)
Cheating-corrected Language score	0.021***	0.024***	0.049***	0.029***	0.038***	0.067***
	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)
Share of Immigrants	2.734***	2.664**	2.641**	2.349**	2.255**	2.226**
	(1.174)	(1.195)	(1.189)	(1.095)	(1.116)	(1.109)
Share of Immigrants in class	0.057	0.064	0.066	0.105*	0.111**	0.113*
	(0.044)	(0.045)	(0.044)	(0.045)	(0.045)	(0.045)
Share of Immigrants in class*Immigrant	-0.081**	-0.084**	-0.086**	-0.039	-0.042	-0.044
	(0.039)	(0.039)	(0.039)	(0.039)	(0.039)	(0.039)
Panel (b): First stage						
Instrument	0.374***	0.374***	0.379***	0.374***	0.379***	0.374***
	(0.029)	(0.029)	(0.029)	(0.029)	(0.029)	(0.029)
F-stat	85.63	85.62	85.62	85.77	85.76	85.76
p-value	0.000	0.000	0.000	0.000	0.000	0.000
Instrument*Immigrant	0.292***	0.291***	0.292***	0.292***	0.292***	0.292***
	(0.032)	(0.032)	(0.031)	(0.032)	(0.032)	(0.032)
F-stat	53.68	53.79	53.77	53.74	53.85	53.84
p-value	0.000	0.000	0.000	0.000	0.000	0.000
Cheating-corrected Math score	Linear	Quadratic	Cubic	Linear	Quadratic	Cubic
Cheating-corrected Language score	Linear	Quadratic	Cubic	Linear	Quadratic	Cubic
Students' characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Family background	Yes	Yes	Yes	Yes	Yes	Yes
Immigrant*controls	Yes	Yes	Yes	Yes	Yes	Yes
LLM FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,147,965	1,147,965	1,147,965	1,149,756	1,149,756	1,149,756
R-squared	0.302	0.303	0.303	0.316	0.317	0.318
No. LLM	560	560	560	561	561	561

Note: FE-IV estimates. The dependent variable is measured by the teacher-assigned grades in Math (columns 1-3) and Language (columns 4-6). We control for LLM and year fixed effects and we focus on the waves 2012/13-2016/17. Standard Errors are robust to heteroskedasticity and are clustered at local labor market level (shown in brackets). Significance at the 10% level is represented by *, at the 5% level by **, and at the 1% level by ***.

Appendix

Table A1: Validity tests

Panel A: Oster Test-First stage	(1) With un-observables	(2) R-squared	(3) Without un-observables	(4) R-squared	(5) Identified set
Instrument (Rmax=1; $\delta = 1$)	0.659	0.557	0.407	0.833	[0.407; 0.659]
Instrument*Immigrant (Rmax=1; $\delta = 1$)	0.972	0.784	0.314	0.948	[0.314; 0.972]

Panel B: Conley et al. (2012) test	(1) Math Score	(2) Language Score
Lower and Upper Bound		
Share of Immigrants*Immigrant	[-2.3327; -0.0223]	[-2.2745; -0.5724]
Cheating-corrected Math score	Linear	Linear
Cheating-corrected Language score	Linear	Linear
Students' characteristics	Yes	Yes
Family background	Yes	Yes
LLM FE	Yes	Yes
Year FE	Yes	Yes
Observations	1,216,215	1,218,620

Panel C: Mitaritonna et al. (2017) test	(1) Instrument 2014-2016	(2) Instrument 2012
Math score 2012-2013	0.004	
	(0.011)	
Language score 2012-2013	-0.003	
	(0.009)	
Employment rate 1981		0.093
		(0.066)
Observations	578	578

Note. Panel A: in columns (1) and (3) we report the betas with and without un-observables and in columns (2) and (4) the corresponding R-squared. Panel B: in columns (1) and (2) we report the 90 percent confidence interval of the parameter of *Share of Immigrants*Immigrant* after the implementation of Conley et al. (2012) test for math and language score. Panel C: variables averaged by LLM and period. Robust standard errors in parenthesis.

Table A2: FE-IV estimates. Heterogeneity by gender. Math and language score – 5^{th} graders

	(1) FE-IV Math score Female	(2) FE-IV Math score Male	(3) FE-IV Language score Female	(4) FE-IV Language score Male
Panel (a): 2SLS				
Share of Immigrants*Immigrant	-1.078*	-0.862	-1.521***	-0.887*
	(0.701)	(0.722)	(0.583)	(0.528)
Immigrant	-0.298*	-0.338***	-0.211**	-0.527***
	(0.119)	(0.127)	(0.103)	(0.126)
Cheating-corrected Math score	0.041***	0.066***	0.019***	0.051***
	(0.003)	(0.003)	(0.003)	(0.003)
Cheating-corrected Language score	0.048***	-0.045***	0.061***	0.062***
	(0.002)	(0.002)	(0.002)	(0.002)
Share of Immigrants	0.867	0.718	1.375*	0.397
	(0.737)	(0.931)	(0.808)	(0.859)
Share of Immigrants in class	0.249***	0.262***	0.297***	0.252***
	(0.049)	(0.049)	(0.047)	(0.048)
Share of Immigrants in class*Immigrant	-0.254***	-0.257***	-0.204***	-0.171***
	(0.056)	(0.052)	(0.051)	(0.052)
Panel (b): First stage				
Instrument	0.408***	0.406***	0.408***	0.406***
	(0.028)	(0.028)	(0.028)	(0.028)
F-stat	107.74	100.78	107.91	100.98
p-value	0.000	0.000	0.000	0.000
Instrument*Immigrant	0.314***	0.313***	0.314***	0.313***
	(0.028)	(0.029)	(0.028)	(0.029)
F-stat	87.26	82.18	87.85	82.45
p-value	0.000	0.000	0.000	0.000
Cheating-corrected Math score	Cubic	Cubic	Cubic	Cubic
Cheating-corrected Language score	Cubic	Cubic	Cubic	Cubic
Students' characteristics	Yes	Yes	Yes	Yes
Family background	Yes	Yes	Yes	Yes
Non-cognitive skills	Yes	Yes	Yes	Yes
Immigrant*controls	Yes	Yes	Yes	Yes
LLM FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	609,842	606,370	611,062	607,555
R-squared	0.392	0.371	0.381	0.367
No. LLM	577	577	577	577

Note: FE-IV estimates. The dependent variable is measured by the teacher-assigned grades in Math (columns 1-2) and Language (columns 3-4). We control for LLM and year fixed effects and we focus on the waves 2012/13-2016/17. Standard Errors are robust to heteroskedasticity and are clustered at local labor market level (shown in brackets). Significance at the 10% level is represented by *, at the 5% level by **, and at the 1% level by ***.

Table A3: FE-IV Estimates. Heterogeneity by socio-economic background. Math and language score – 5^{th} graders

	(1) FE-IV Math score <median	(2) FE-IV Math score >median	(3) FE-IV Language score <median	(4) FE-IV Language score >median
Panel (a): 2SLS				
Share of Immigrants*Immigrant	-1.015	-1.102*	-1.025***	-1.264**
	(0.723)	(0.581)	(0.537)	(0.592)
Immigrant	-0.018	-0.721***	-0.301***	-0.701***
	(0.044)	(0.182)	(0.103)	(0.157)
Cheating-corrected Math score	0.045***	0.061***	0.031***	0.039***
	(0.003)	(0.004)	(0.001)	(0.004)
Cheating-corrected Language score	0.049***	0.046***	0.064***	0.059
	(0.002)	(0.002)	(0.002)	(0.002)
Share of Immigrants	0.594	0.721	0.758	0.878
	(0.945)	(0.753)	(0.924)	(0.783)
Share of Immigrants in class	0.269***	0.191***	0.271***	0.257***
	(0.047)	(0.056)	(0.045)	(0.063)
Share of Immigrants in class*Immigrant	-0.277***	-0.179***	-0.201***	-0.159**
	(0.052)	(0.061)	(0.047)	(0.069)
Panel (b): First stage				
Instrument	0.395***	0.419***	0.395***	0.419***
	(0.029)	(0.027)	(0.029)	(0.027)
F-stat	89.40	125.47	89.54	125.69
p-value	0.000	0.000	0.000	0.000
Instrument*Immigrant	0.308***	0.339***	0.308***	0.341***
	(0.028)	(0.033)	(0.028)	(0.033)
F-stat	83.49	82.76	83.61	83.52
p-value	0.000	0.000	0.000	0.000
Cheating-corrected Math score	Cubic	Cubic	Cubic	Cubic
Cheating-corrected Language score	Cubic	Cubic	Cubic	Cubic
Students' characteristics	Yes	Yes	Yes	Yes
Family background	Yes	Yes	Yes	Yes
Non-cognitive skills	Yes	Yes	Yes	Yes
Immigrant*controls	Yes	Yes	Yes	Yes
LLM FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	605,807	610,405	607,035	611,582
R-squared	0.355	0.321	0.355	0.315
No. LLM	577	577	577	577

Note: FE-IV estimates. The dependent variable is measured by the teacher-assigned grades in Math (columns 1-2) and Language (columns 3-4). We control for LLM and year fixed effects and we focus on the waves 2012/13-2016/17. Standard Errors are robust to heteroskedasticity and are clustered at local labor market level (shown in brackets). Significance at the 10% level is represented by *, at the 5% level by **, and at the 1% level by ***.

Table A4: FE-IV estimates with different polynomials of Rasch math and language score – 5^{th} graders

	(1) FE-IV Math score	(2) FE-IV Math score	(3) FE-IV Math score	(4) FE-IV Language score	(5) FE-IV Language score	(6) FE-IV Language score
Panel (a): 2SLS						
Share of Immigrants*Immigrant	-1.045	-0.867	-0.873	-1.249**	-1.098**	-1.107**
	(0.723)	(0.727)	(0.712)	(0.566)	(0.573)	(0.558)
Immigrant	-0.153*	-0.142*	-0.155*	-0.307***	-0.298***	-0.308***
	(0.091)	(0.091)	(0.089)	(0.085)	(0.086)	(0.085)
Rasch Math score	0.274***	0.337***	0.354***	0.158***	0.196***	0.201***
	(0.011)	(0.011)	(0.011)	(0.007)	(0.005)	(0.006)
Rasch Language score	0.249***	0.261***	0.292***	0.321***	0.337***	0.378***
	(0.003)	(0.003)	(0.003)	(0.005)	(0.005)	(0.005)
Share of Immigrants	-0.736	-0.553	-0.511	-0.559	-0.427	-0.382
	(0.878)	(0.821)	(0.819)	(0.829)	(0.789)	(0.791)
Share of Immigrants in class	0.198***	0.234***	0.227***	0.226***	0.253***	0.247***
	(0.045)	(0.045)	(0.044)	(0.045)	(0.041)	(0.044)
Share of Immigrants in class*Immigrant	-0.194***	-0.213***	-0.207***	-0.133***	-0.148***	-0.142***
	(0.051)	(0.051)	(0.048)	(0.048)	(0.047)	(0.047)
Panel (b): First stage						
Instrument	0.407***	0.407***	0.407***	0.407***	0.407***	0.407***
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)
F-stat	104.05	104.08	104.09	104.22	104.25	104.26
p-value	0.000	0.000	0.000	0.000	0.000	0.000
Instrument*Immigrant	0.314***	0.314***	0.314***	0.314***	0.314***	0.314***
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)
F-stat	90.71	90.31	90.31	91.30	90.89	90.89
p-value	0.000	0.000	0.000	0.000	0.000	0.000
Rasch Math score	Linear	Quadratic	Cubic	Linear	Quadratic	Cubic
Rasch Language score	Linear	Quadratic	Cubic	Linear	Quadratic	Cubic
Students' characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Family background	Yes	Yes	Yes	Yes	Yes	Yes
Non-cognitive skills	Yes	Yes	Yes	Yes	Yes	Yes
Immigrant*controls	Yes	Yes	Yes	Yes	Yes	Yes
LLM FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,216,241	1,216,241	1,216,241	1,218,646	1,218,646	1,218,646
R-squared	0.369	0.382	0.382	0.381	0.389	0.391
No. LLM	577	577	577	577	577	577

Note: FE-IV estimates. The dependent variable is measured by the teacher-assigned grades in Math (columns 1-3) and Language (columns 4-6). We control for LLM and year fixed effects and we focus on the waves 2012/13-2016/17. Standard Errors are robust to heteroskedasticity and are clustered at local labor market level (shown in brackets). Significance at the 10% level is represented by *, at the 5% level by **, and at the 1% level by ***.

Table A5: FE-IV estimates. Math and language score with ESCS index as control – 5^{th} graders

	(1) FE-IV Math score	(2) FE-IV Math score	(3) FE-IV Math score	(4) FE-IV Language score	(5) FE-IV Language score	(6) FE-IV Language score
Panel (a): 2SLS						
Share of Immigrants*Immigrant	-1.117*	-1.011	-0.992	-1.309***	-1.199***	-1.176**
	(0.669)	(0.689)	(0.695)	(0.476)	(0.499)	(0.505)
Immigrant	-0.164**	-0.169***	-0.169***	-0.328***	-0.336***	-0.334***
	(0.051)	(0.051)	(0.051)	(0.043)	(0.043)	(0.043)
Cheating-corrected Math score	0.049***	0.055***	0.055***	0.029***	0.032***	0.036***
	(0.001)	(0.003)	(0.003)	(0.001)	(0.002)	(0.002)
Cheating-corrected Language score	0.026***	0.036***	0.049***	0.034***	0.047***	0.062***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Share of Immigrants	1.026	0.871	0.852	1.201	0.993	0.966
	(0.842)	(0.767)	(0.764)	(0.873)	(0.776)	(0.768)
Share of Immigrants in class	0.235***	0.246***	0.248***	0.251***	0.262***	0.265***
	(0.043)	(0.043)	(0.049)	(0.041)	(0.042)	(0.042)
Share of Immigrants in class*Immigrant	-0.253***	-0.259***	-0.261***	-0.182***	-0.187***	-0.189***
	(0.051)	(0.051)	(0.051)	(0.047)	(0.047)	(0.047)
Panel (b): First stage						
Instrument	0.407***	0.407***	0.407***	0.407***	0.407***	0.407***
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)
F-stat	103.82	104.02	103.98	103.99	104.19	104.15
p-value	0.000	0.000	0.000	0.000	0.000	0.000
Instrument*Immigrant	0.315***	0.315***	0.315***	0.315***	0.315***	0.315***
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)
F-stat	89.55	87.66	87.52	89.88	88.06	87.92
p-value	0.000	0.000	0.000	0.000	0.000	0.000
Cheating-corrected Math score	Linear	Quadratic	Cubic	Linear	Quadratic	Cubic
Cheating-corrected Language score	Linear	Quadratic	Cubic	Linear	Quadratic	Cubic
Students' characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Family background	Yes	Yes	Yes	Yes	Yes	Yes
Non-cognitive skills	Yes	Yes	Yes	Yes	Yes	Yes
Immigrant*controls	Yes	Yes	Yes	Yes	Yes	Yes
LLM FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,215,274	1,215,274	1,215,274	1,217,673	1,217,673	1,217,673
R-squared	0.368	0.369	0.369	0.369	0.371	0.372
No. LLM	577	577	577	577	577	577

Note: FE-IV estimates. The dependent variable is measured by the teacher-assigned grades in Math (columns 1-3) and Language (columns 4-6). We control for LLM and year fixed effects and we focus on the waves 2012/13-2016/17. Standard Errors are robust to heteroskedasticity and are clustered at local labor market level (shown in brackets). Significance at the 10% level is represented by *, at the 5% level by **, and at the 1% level by ***.

Table A6: FE-IV estimates. Bullying, victimization and socialization of students – 5^{th} graders

	(1) FE-IV Victimization	(2) FE-IV Bullying	(3) FE-IV Feeling good with classmates	(4) FE-IV Friends with classmates
Panel (a): 2SLS				
Share of Immigrants*Immigrant	-0.004	0.229	0.163	0.078
	(0.105)	(0.159)	(0.123)	(0.114)
Immigrant	0.006	-0.015	-0.045	0.054
	(0.038)	(0.011)	(0.057)	(0.058)
Cheating-corrected Math score	-0.004***	-0.007***	-0.005***	-0.005***
	(0.001)	(0.001)	(0.001)	(0.001)
Cheating-corrected Language score	-0.014***	-0.008***	-0.004***	0.004
	(0.001)	(0.001)	(0.001)	(0.001)
Share of Immigrants	0.711	1.271	0.793**	0.449
	(0.452)	(0.797)	(0.341)	(0.341)
Share of Immigrants in class	0.036***	-0.041**	-0.066***	-0.043***
	(0.011)	(0.017)	(0.013)	(0.011)
Share of Immigrants in class*Immigrant	-0.006	0.088***	0.076***	0.052***
	(0.013)	(0.021)	(0.015)	(0.012)
Panel (b): First stage				
Instrument	0.391***	0.391***	0.391***	0.391***
	(0.042)	(0.042)	(0.042)	(0.042)
F-stat	53.04	53.04	53.22	53.32
p-value	0.000	0.000	0.000	0.000
Instrument*Immigrant	0.314***	0.314***	0.314***	0.314***
	(0.031)	(0.031)	(0.031)	(0.031)
F-stat	56.01	56.01	55.43	55.76
p-value	0.000	0.000	0.000	0.000
Cheating-corrected Math score	Cubic	Cubic	Cubic	Cubic
Cheating-corrected Language score	Cubic	Cubic	Cubic	Cubic
Students' characteristics	Yes	Yes	Yes	Yes
Family background	Yes	Yes	Yes	Yes
Immigrant*controls	Yes	Yes	Yes	Yes
LLM FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	449,973	449,973	447,846	448,315
R-squared	0.035	0.035	0.022	0.012
No. LLM	556	556	556	556

Note: FE-IV estimates. The dependent variable is on top of each column. We control for LLM and year fixed effects and we focus on the waves 2013/14-2014/15. Standard Errors are robust to heteroskedasticity and are clustered at local labor market level (shown in brackets). Significance at the 10% level is represented by *, at the 5% level by **, and at the 1% level by ***.

Table A7: FE-IV estimates with bullying, victimization and socialization of students as controls – 5^{th} graders

	(1) FE-IV Math score	(2) FE-IV Language score
Panel (a): 2SLS		
Share of Immigrants*Immigrant	-1.854***	-1.931***
	(0.644)	(0.499)
Immigrant	-0.003	-0.159
	(0.126)	(0.164)
Cheating-corrected Math score	0.052***	0.043***
	(0.003)	(0.004)
Cheating-corrected Language score	0.011***	0.018***
	(0.004)	(0.004)
Share of Immigrants	-2.393	-1.087
	(2.739)	(2.329)
Share of Immigrants in class	0.325***	0.314***
	(0.064)	(0.069)
Share of Immigrants in class*Immigrant	-0.309***	-0.186***
	(0.059)	(0.067)
Panel (b): First stage		
Instrument	0.391***	0.391***
	(0.042)	(0.042)
F-stat	53.65	53.47
p-value	0.000	0.000
Instrument*Immigrant	0.313***	0.314***
	(0.031)	(0.031)
F-stat	55.24	55.34
p-value	0.000	0.000
Cheating-corrected Math score	Cubic	Cubic
Cheating-corrected Language score	Cubic	Cubic
Students' characteristics	Yes	Yes
Family background	Yes	Yes
Immigrant*controls	Yes	Yes
Bullying, victimization and socialization controls	Yes	Yes
LLM FE	Yes	Yes
Year FE	Yes	Yes
Observations	445,869	446,596
R-squared	0.379	0.388
No. LLM	556	556

Note: FE-IV estimates. The dependent variable is on top of each column. We control for LLM and year fixed effects and we focus on the waves 2013/14-2014/15. Standard Errors are robust to heteroskedasticity and are clustered at local labor market level (shown in brackets). Significance at the 10% level is represented by *, at the 5% level by **, and at the 1% level by ***.

Table A8: FE-IV estimates. Placebo on low achieving natives. Math and language score – 5^{th} graders

	(1) FE-IV Math score Natives	(2) FE-IV Language score Natives
Panel (a): 2SLS		
Share of Immigrants*Low-achieving	-0.265	-0.116
	(0.196)	(0.172)
Low-achieving	-0.538***	-0.541***
	(0.077)	(0.157)
Cheating-corrected Math score	0.029***	0.017***
	(0.001)	(0.004)
Cheating-corrected Language score	0.029***	0.042
	(0.002)	(0.002)
Share of Immigrants	0.611	0.645
	(0.786)	(0.777)
Share of Immigrants in class	0.025	0.099***
	(0.031)	(0.063)
Share of Immigrants in class*Low-achieving	-0.068*	-0.016
	(0.037)	(0.034)
Panel (b): First stage		
Instrument	0.408***	0.408***
	(0.029)	(0.027)
F-stat	105.09	105.30
p-value	0.000	0.000
Instrument*Low-achieving	0.428***	0.428***
	(0.037)	(0.037)
F-stat	65.23	65.25
p-value	0.000	0.000
Cheating-corrected Math score	Cubic	Cubic
Cheating-corrected Language score	Cubic	Cubic
Students' characteristics	Yes	Yes
Family background	Yes	Yes
Low-achieving*controls	Yes	Yes
LLM FE	Yes	Yes
Year FE	Yes	Yes
Observations	1,116,715	1,118,896
R-squared	0.373	0.367
No. LLM	577	577

Note: FE-IV estimates. The dependent variable is measured by the teacher-assigned grades in Math (columns 1) and Language (columns 2). We control for LLM and year fixed effects and we focus on the waves 2012/13-2016/17. Standard Errors are robust to heteroskedasticity and are clustered at local labor market level (shown in brackets). Significance at the 10% level is represented by *, at the 5% level by **, and at the 1% level by ***.

Table A9: FE-IV estimates. I and II generation of immigrants and population size – 5^{th} graders

	(1) FE-IV Math score	(2) FE-IV Language score	(3) FE-IV Math score	(4) FE-IV Language score	(5) FE-IV Math score	(6) FE-IV Language score	(7) FE-IV Math score	(8) FE-IV Language score
	pop<med	pop<med	pop<med	pop<med	pop>med	pop>med	pop>med	pop>med
Panel (a): 2SLS								
Share of Immigrants*Immigrant	-1.067**	-0.757			-0.752	0.192		
(I generation)	(0.471)	(0.472)			(0.903)	(0.821)		
Share of Immigrants*Immigrant			-1.584***	-1.687***			0.906	0.005
(II generation)			(0.474)	(0.422)			(1.129)	(0.611)
Immigrant (I generation)	-0.143***	-0.301***			-0.296***	-0.411***		
	(0.041)	(0.041)			(0.081)	(0.071)		
Immigrant (II generation)			-0.029	-0.126***			-0.231**	-0.273***
			(0.041)	(0.038)			(0.108)	(0.059)
Cheating-corrected Math score	0.049***	0.032***	0.053***	0.034***	0.053***	0.038***	0.054***	0.037***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)
Cheating-corrected Language score	0.051***	0.063***	0.049***	0.063***	0.046***	0.058***	0.046***	0.058***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Share of Immigrants	1.351	1.277	1.497	1.137	0.447	0.557	0.434	0.665
	(1.253)	(1.241)	(1.246)	(1.206)	(1.617)	(1.351)	(1.518)	(1.307)
Share of Immigrants in class	0.119	0.132*	0.192***	0.167***	0.231***	0.245***	0.148***	0.217***
	(0.092)	(0.043)	(0.062)	(0.061)	(0.054)	(0.057)	(0.052)	(0.049)
Share of Immigrants in class*Immigrant	-0.089	-0.007	-0.109*	-0.056	-0.258***	-0.219***	-0.212***	-0.162***
	(0.086)	(0.074)	(0.065)	(0.059)	(0.067)	(0.065)	(0.065)	(0.055)
Panel (b): First stage								
Instrument	0.386***	0.386***	0.386***	0.386***	0.403***	0.403***	0.405***	0.405***
	(0.031)	(0.031)	(0.031)	(0.031)	(0.037)	(0.037)	(0.037)	(0.037)
F-stat	80.75	80.88	79.24	79.33	74.68	74.94	82.44	82.50
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Instrument*Immigrant	0.344***	0.344***	0.293***	0.293***	0.385***	0.385***	0.348***	0.348***
	(0.033)	(0.033)	(0.029)	(0.029)	(0.046)	(0.046)	(0.042)	(0.042)
F-stat	56.75	56.86	66.41	66.67	49.61		49.63	53.43
53.65								
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Cheating-corrected Math score	Cubic	Cubic	Cubic	Cubic	Cubic	Cubic	Cubic	Cubic
Cheating-corrected Language score	Cubic	Cubic	Cubic	Cubic	Cubic	Cubic	Cubic	Cubic
Students' characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family background	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Non-cognitive skills	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Immigrant*controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
LLM FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	569,729	570,786	593,820	594,926	573,287	574,472	595,943	597,176
R-squared	0.386	0.381	0.391	0.387	0.361	0.363	0.364	0.367
No. LLM	568	568	568	568	294	294	294	294

Note: FE-IV estimates. The dependent variable is measured by the teacher-assigned grades in Math (columns 1, 3, 5 and 7) and Language (columns 2, 4, 6 and 8). We control for LLM and year fixed effects and we focus on the waves 2012/13-2016/17. Standard Errors are robust to heteroskedasticity and are clustered at local labor market level (shown in brackets). Significance at the 10% level is represented by *, at the 5% level by **, and at the 1% level by ***.

Table A10: FE-IV estimates. Math and language score – 8^{th} graders

	(1) FE-IV Math score	(2) FE-IV Math score	(3) FE-IV Math score	(4) FE-IV Language score	(5) FE-IV Language score	(6) FE-IV Language score
Panel (a): 2SLS						
Share of Immigrants*Immigrant	-0.675**	-0.784**	-0.826***	-0.934***	-1.013***	-1.045***
	(0.501)	(0.308)	(0.302)	(0.264)	(0.259)	(0.254)
Immigrant	-0.044	-0.038	-0.034	-0.051	-0.055	-0.059
	(0.086)	(0.085)	(0.085)	(0.069)	(0.069)	(0.069)
Cheating-corrected Math score	0.068***	0.023***	0.054***	0.033***	0.008***	0.032***
	(0.001)	(0.001)	(0.006)	(0.001)	(0.001)	(0.002)
Cheating-corrected Language score	0.034***	0.029***	0.009***	0.047***	0.041***	0.017***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
Share of Immigrants	0.456	0.447	0.493	0.891	0.879	0.901
	(0.805)	(0.802)	(0.799)	(0.649)	(0.639)	(0.639)
Share of Immigrants in class	0.601***	0.599***	0.597***	0.437***	0.437***	0.434***
	(0.044)	(0.045)	(0.069)	(0.072)	(0.071)	(0.069)
Share of Immigrants in class*Immigrant	-0.584***	-0.592***	-0.591***	-0.412***	-0.418***	-0.417***
	(0.068)	(0.067)	(0.067)	(0.061)	(0.061)	(0.059)
Panel (b): First stage						
Instrument	0.269***	0.269***	0.269***	0.269***	0.269***	0.269***
	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
F-stat	50.01	50.02	50.02	49.84	49.85	49.86
p-value	0.000	0.000	0.000	0.000	0.000	0.000
Instrument*Immigrant	0.242***	0.242***	0.242***	0.242***	0.242***	0.242***
	(0.032)	(0.032)	(0.032)	(0.018)	(0.018)	(0.018)
F-stat	87.01	87.16	87.15	87.44	87.64	87.62
p-value	0.000	0.000	0.000	0.000	0.000	0.000
Cheating-corrected Math score	Linear	Quadratic	Cubic	Linear	Quadratic	Cubic
Cheating-corrected Language score	Linear	Quadratic	Cubic	Linear	Quadratic	Cubic
Students' characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Family background	Yes	Yes	Yes	Yes	Yes	Yes
Immigrant*controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,271,040	1,271,040	1,271,040	1,303,867	1,303,867	1,303,867
R-squared	0.418	0.423	0.425	0.434	0.436	0.437

Note: FE-IV estimates. The dependent variable is measured by the teacher-assigned grade in Math (columns 1-3) and Language (columns 4-6). We control for LLM and year fixed effects and we focus on the waves 2012/13-2016/17. Significance at the 10% level is represented by *, at the 5% level by **, and at the 1% level by ***.