Tower of Babel in the Classroom: Immigrants and Natives in Italian Schools

Published Version:
Tower of Babel in the Classroom: Immigrants and Natives in Italian Schools / Ballatore, Rosario Maria; Fort, Margherita; Ichino, Andrea. - In: JOURNAL OF LABOR ECONOMICS. - ISSN 0734-306X. - STAMPA. - 36:4(2018), pp. 885-921. [10.1086/697524]

Availability:
This version is available at: https://hdl.handle.net/11585/665651 since: 2021-03-01

Published:
DOI: http://doi.org/10.1086/697524

The Tower of Babel in the Classroom

Immigrants and Natives in Italian Schools

Rosario Maria Ballatore
Bank of Italy

Margherita Fort
Bologna

Andrea Ichino
Bologna and Eui

February 22, 2018

Abstract

We exploit rules of class formation to identify the causal effect of increasing the number of immigrants in a classroom on natives test scores, keeping class size and quality of the two types of students constant (Pure Ethnic Composition effect). We explain why this is a relevant policy parameter although it has been neglected so far. The PEC effect is sizeable and negative (16% of a st. dev.) on language and math scores. For first generation immigrants it is more negative (30% of a st. dev.). Estimates that cannot control for endogenous adjustments implemented by principals, are instead considerably smaller.

We would like to thank INVALSI (in particular Paolo Sestito, Piero Cipollone, Patrizia Falzetti and Roberto Ricci) and Gianna Barbieri (Italian Ministry of Education and University, MIUR) for giving us access to the many sources of restricted data that have been used in this paper. We are also grateful to Josh Angrist, Eric Battistin and Daniela Vuri who shared their data with us and gave us crucial suggestions. We received very useful comments from seminar participants at the European University Institute, the University of Bologna, the II International Workshop on Applied Economics of Education, the European Economic Association of Labour Economists, the Einaudi Institute for Economics and Finance, the Cage-Warwick Conference in Venice, ISER (University of Essex), CESifo, the University of Helsinki. Finally, we would like to thank Gordon Dahl, Elie Murard, Marco Paccagnella, Massimo Pietroni, Enrico Rettore, Lucia Rizzica, Paolo Sestito and Marco Tonello for the time they spent discussing this paper with us. We acknowledge financial support from MIUR- PRIN 2009 project 2009MAATFS_001. Margherita Fort is affiliated with CESifo and IZA. Andrea Ichino is a research fellow of CEPR, CESifo and IZA. The views here expressed are those of the authors and do not necessarily reflect those of the Bank of Italy.

1 Introduction

The integration of non-native children in schools is a potential problem that many countries are facing under the pressure of anecdotal evidence that generates increasing concerns in the population and among policy makers. We contribute to the literature that studies the causal effect of an immigrant inflow on native school performance[1], by comparing estimates for the case in which immigrants are added to classes keeping the number of natives constant and for the case in which, instead, additional resources allow the principal to keep total class size constant, thus replacing natives with new immigrants. We show how this comparison can be performed empirically and argue that it provides a more informative evaluation of the consequences of an immigrant inflow.

To clarify this point, let us abstract momentarily from other observables. We can then model the performance V_j^N of a native student in class j as:

$$V_j^N = \alpha + \beta N_j + \gamma I_j + U_j \tag{1}$$

where N_j and I_j are, respectively, the numbers of natives and immigrants in that class. A first possible way to look at the effect of an immigrant inflow on this outcome is to focus on γ, which measures the effect of adding one immigrant to class j keeping constant the number of natives. It is however possible that principals use additional resources to cope with the immigrant inflow, particularly if they expect non-native children to be on average more problematic and that smaller classes help the learning process. In this case, immigrants may

[1]See, for example, Hoxby (2000), Bossavie (2011), Tonello (2016) who exploit the variability in ethnic composition between adjacent cohorts within the same schools finding a weak negative inter-race peer effect on test scores, while the intra-race and intra-immigrants status peer effect is found to be more clearly negative and stronger. Gould et al. (2009) uses the mass immigration from the Soviet Union that occurred in Israel during the 1990s to identify the long run causal effect of having immigrants as classmates, finding a negative effect of immigration on the probability of passing the high-school matriculation exam. A similar global event is used for identification by Geay et al. (2013) to reach opposite conclusions: they focus on the inflow of non native speakers students in English schools induced by the Eastern Enlargement of the EU in 2005 and conclude, in contrast with the Israeli case, that a negative effect can be ruled out. Negative effects on natives performance are found by Jensen and Rasmussen (2011), who use the immigrants concentration at larger geographical areas as an instrumental variable for the share of immigrants in a school and by Brunello and Rocco (2013), who aggregate the data at the country level and exploit the within country variation over time in the share of immigrants in a school. On the contrary Hunt (2017), using variation across US states and years as well as instruments constructed on previous settlement patterns of immigrants, reports positive effects (though small) of immigrants' concentration on the probability that natives complete high-school.

displace natives, inasmuch as the additional resources allow principals to contain the increase in total class size that would be induced by new immigrants at constant resources. When this is a possibility, a relevant parameter for policy is $\delta = \gamma - \beta$, which measures the effect of an immigrant inflow net of any change in class size. If this effect were null, immigrants would be equivalent to natives and an inflow of immigrants should not be treated by principals differently then an inflow of natives. We call δ the Pure Ethnic Composition (PEC) effect.

The presence of non-ignorable unobservables in U_j is of course problematic for the identification of β and γ in equation (1). If we are interested only in γ, the relevant concern is that principals may choose the class in which to assign new immigrants taking into account the quality of the students already attending in the available classes. If, for example, $U_j = \lambda Q^N_j + \mu Q^I_j + \epsilon_j$, where Q^N_j and Q^I_j are, respectively, the unobservable qualities of natives and immigrants in class j, OLS estimates of γ would be inconsistent because of the correlation between I_j and these qualities. With a valid instrument for I_j we could estimate γ consistently, even in the likely event that N_j were correlated with Q^N_j and Q^I_j, because this parallel correlation would just be an inconsequential nuisance if we are interested only in γ. However, it would represent a problem if we want to obtain consistent estimates of both β and γ do derive the Pure Ethnic Composition effect δ.

To address these identification issues, we adapt to our contest the empirical strategy designed by Angrist and Lang (2004). Their goal is to estimate the effect of an increase in the number of disadvantaged students in schools of affluent districts in the Boston area, induced by the desegregation program run by the Metropolitan Council for Educational Opportunity (Metco). For identification, they exploit the fact that students from disadvantaged neighbourhoods are transferred by Metco to receiving schools on the basis of the available space generated by a "Maimonides-type rule" of class formation (Angrist and Lavy 1999) requiring principals to cap class size at 25 and to increase the number of classes whenever the enrollment of non-Metco students goes above the 25 thresholds or its multiples.

In the Italian context, students should pre-enroll in a given school during the month of February for the year that starts in the following September. Each principal manages an educational *institution* with multiple schools and tentatively decides in February the number of classes in each school according to the corresponding number of pre-enrolled students, following a similar "Maimonides rule" with a cap at 25. However, a crucial institutional detail for our empirical strategy is that some additional splitting of classes may occur in September if late enrollment requires further adjustments.

If principals knew the final student count when making assignments and creating classes in February, we would expect to find a monotone negative relation between native class size and the number of immigrants. However, because plans are made before the final student count is known, some late student arrivals cause large classes with no immigrants to split in two small classes with no immigrants. Consequently, we show that in our setting the relation between the number of immigrants in a class and predicted native class size is hump-shaped. As for the actual number of natives, instead, its relation with predicted native class size is monotone and positive. This institutional detail is helpful because it allows us to use this exogenous variation to construct multiple instruments for both N_j and I_j, which is what we need to identify β, γ and thus δ.

As in any other setting exploiting a Maimonides-type rule for identification, parents choose schools on the basis of their quality, preferences and expectations (including those regarding the presence of immigrants), but they cannot easily predict whether the realised enrollment in each given school reaches a class splitting threshold. This is true for both February and September enrollment decisions: in both cases, the comparison of student performance above and below class splitting thresholds is therefore likely not to be confounded by unobservables.[2] However, if enrollment and class formation occurred only in February, we would be able to identify only the effect of total class size on student performance, as in a standard setting in which a Maimonides-type-rule is used for identification. The interaction between February and September enrollment, instead, is helpful because it allows us to disentangle separate effects for the numbers of immigrants and natives in a class, as it induces a hump-shaped relationship between the number of immigrants in class and theoretical native class size.

We find that the Pure Ethnic Composition effect on native language test scores is negative and statistically significant (about 16% of a standard deviation), when all immigrants are considered. For first generation immigrants the estimate is slightly more negative (about

[2] See also a recent discussion of the use of Maimonides-type rules in Angrist et al. (2017b).

30% of a standard deviation). Results for math test scores are similar in size but less precise.[3] We compare this evidence with that obtained using a strategy that exploits variation between schools within the same educational institution (or classes within the same school), in the same spirit of Contini (2013) and Ohinata and van Ours (2013). Although these alternative estimates go a long way in controlling for unobservables, they cannot take care of the possibility that, within a grade of a given school (or institution), principals allocate immigrants and natives to classes taking their quality and class size into account. Using these strategies we obtain an estimate of γ that is significantly smaller in absolute size, suggesting that principals tend to allocate more immigrants to smaller classes or to classes with natives of better quality. Our results do not suffer from this problem because they are driven by the exogenous variation determined by a Maimonides-type rule in conjunction with the consequences of the interaction between February and September class splitting.

The paper is organised as follows. After a description of our data in Section 2, in Section 3 we describe the institutional setting. Section 4 discusses the assumptions on which our identification strategy is based. Results are presented and discussed in Section 5. Section 6 concludes.

2 The data

The data on test scores used in this paper are collected by the Italian National Institute for the Evaluation of the Education System (INVALSI). They originate from a standardised testing procedure that assesses both language (Italian) and mathematical skills of pupils in 2nd and 5th grade (primary school). We use the 2009-2010 wave, i.e. the first one in which all schools and students of the selected grades were required to take part in the assessment.[4]

We aggregate the data at the school level since the variation we exploit for identification is between schools managed by the same principal. We define this aggregation of schools as an institution. The outcomes on which we focus are the fractions of correct answers in

[3] In principle, our identification strategy would work also for the PEC effect of an immigrant inflow on the performance of immigrants themselves but, in the data at our disposal, there is effectively not enough information for this purpose and, thus, we focus our empirical analysis entirely on the performance of natives.

[4] In previous waves, the participation of individual schools to the test was voluntary. Only a very limited number of schools and of students within schools were sampled on a compulsory basis.

language and math of natives who were not absent on the day of the test. Following the INVALSI classification (see INVALSI 2010), we consider as natives those students who are born from at least one Italian parent, independently of the place of birth. Viceversa, students born from parents who are both non-Italian are classified as immigrants, again regardless of whether they are born in Italy or not.[5] We begin by adopting this classification, but we also replicate our analysis using a more stringent definition of "first generation" immigrant that attributes this status to children not born in Italy from parents who are both non-Italian. The complement of first generation immigrants, which we label as quasi-natives, is therefore made of natives (children with at least one Italian parent) and "second generation" immigrants (children who are born in Italy from parents who are both non-Italian). A drawback of this second classification is that first generation immigrant status is available only for students actually taking each test, not for all students enrolled.

Note that since language and math tests were held on different days and students could have missed none, one or both tests, regressions for the two outcomes are based on slightly different datasets. Descriptive statistics for the language sample are displayed in Table 1.[6]

In addition to test scores, the INVALSI dataset contains some individual socio-economic variables collected by school administrations for each student taking the test, among which: gender, previous attendance of nursery or kindergarten, highest educational level achieved by parents and parental occupational status. We aggregate this information for natives at the school level to construct the control variables that we include in our specifications, together with the share of native students for whom each of these variables is missing.

Starting from the universe of schools (16,828 belonging to 7,561 institutions in grade 2; 16,803 belonging 7,549 institutions in grade 5) we operate the following sample restrictions for the language sample.[7] We retain all schools in institutions in which at least one immigrant applies for a given grade (14,580 schools of 5,966 different institutions in grade 2; 14,603

[5] The numbers of natives and immigrants based on this definition, that were officially enrolled in each class at the beginning of a school year, are not contained in the standard files distributed by INVALSI, but were kindly provided to us in a separate additional file.

[6] Analogous statistics for the math sample and for both the language and math samples when immigrant status is defined as first generation only, can be found in the Online Appendix. The are very similar to the ones displayed in Table 1.

[7] The analogous description of how we constructed the data for the other samples, can be found in the Online Appendix.

schools of 6,010 different institutions in grade 5). We then restrict the attention to schools that enroll between 10 and 75 native students (11,574 schools of 5,466 different institutions in grade 2; 11,639 schools of 5,484 different institutions in grade 5). We remove records of students with missing test scores in language and we collapsed the data at the school level. We drop the few schools in which no native took the test and the schools in which at least one of the included covariate is missing (1,699 schools and 746 institutions in grade 2; 1,675 schools and 726 institutions in grade 5). We restrict our attention to the schools that are grouped together with other schools in educational institutions managed by a single principal (about 82%): as anticipated in the Introduction and further explained below in Section 3, our identification strategy cannot apply to "stand-alone" schools. This leaves us with 8,014 schools of 2,867 different institutions in grade 2 and 8,086 schools in 2,882 different institutions in grade 5.

The average enrollment of natives per school-grade is 28.2 while for immigrants it is 3.75. We measure school performance with the fraction of correct answers in each test. As expected, immigrants tend to perform worse than natives in language and math (respectively 0.54 and 0.55 for immigrants and 0.67 and 0.62 for natives), but the gap between ethnic groups is more sizeable in language. Natives perform relatively better in Italian than in mathematics and the opposite happens for immigrants. The gap between natives and immigrants in language tends to narrow across grades but remains relatively more stable in mathematics. Finally, the dispersion in the score distribution for both Italian and mathematics is lower among natives who are more homogeneous than immigrants. The fact that immigrants test scores are lower on average, has motivated the public opinion concern that immigrant inflows may reduce native performance. Changing the sample or the definition of immigrant status does not affect the above descriptive findings, although the gap between natives and first generation immigrants is somewhat larger in both language and mathematics test scores.

3 Variability for the identification of the PEC effect

In the month of February of each year Italian parents are invited to pre-enroll their children, for the following academic year, in one of the schools near where they live.[8] On the basis of this pre-enrollment information at the school level, principals forecast the number of classes they will need in the schools that they manage, being constrained by a "Maimoinides-type rule": no class should have more than 25 students (and less than 10), with a 10% margin of flexibility around these thresholds.

Principals decide also on a preliminary allocation of students across schools. While natives are typically assigned to the schools in which they pre-enroll, for immigrants the allocation is less straightforward. The instructions of the Ministry are that foreign students should be directed towards schools where, because of how classes are formed, there is sufficient space for immigrants and any potential disruption can thus be avoided.[9] After the February pre-enrollment phase, later in September final enrollment in schools may change because of new arrivals, family mobility and other contingencies. According to the Ministry[10], in the school year 2013-2014 approximately 35,000 students enrolled later than February, corresponding to approximately 6% of total enrollment.

As a result of this allocation mechanism, the average number of immigrants per class is a hump-shaped function of the average number of natives. This happens because a fraction of classes with few natives originates from the splitting of classes expected to be large in February and in which principals do not put immigrants to reduce disruption. Classes that are expected to be large in the number of natives and remain large have no or few immigrants, while the highest number of immigrants remains allocated to classes with an intermediate number of natives.[11]

[8]The official rules for class formation in Italy are contained the DL n. 331/1998 and the DPR n. 81/2009.

[9]See, for example, the "Circolare ministeriale" Number 4, comma 10.2, of January 15, 2009: "In order to avoid the problems and inconvenience deriving from the presence of students of foreign citizenship, principals are invited to make use of schools in their institutions to achieve a rational territorial distribution of these students. [...] In areas where institutions grouping multiple schools under the same principal are already present, the enrollment of foreign students must be handled in a controlled way so that their allocation across schools is less disruptive."

[10]We are grateful to Dr. Gianna Barbieri who gave us this aggregate information that concerns the enrollment in the first year of primary school.

[11]See Section A of the Online Appendix for a formal characterisation of this allocation mechanism.

This hump shape emerges clearly in our data, as shown in Figure 1 that plots the average number of natives per class (squares; left vertical axis) and of immigrants per class (circles; right vertical axis) for each level of theoretical class size based on native enrollment. The figure also plots fitted values of the two relationships (solid for immigrants and dashed for natives). Theoretical native class size in school s of institution k and grade g is calculated as a function of the corresponding final enrollment of natives N_{skg}, using the following "Maimonides-type" rule:

$$C^N_{skg} = \frac{N_{skg}}{\text{Int}\left(\frac{N_{skg}-1}{25}\right)+1} \tag{2}$$

Figure 1 shows that the average number of natives in a class is an increasing function of theoretical class size, as predicted by the conventional effects of a Maimonides-type rule. These conventional effects are highlighted in the left panel of Figure 2. The dashed line plots the theoretical class size C^N_{skg} as a function of the final enrollment of natives in each school, while the solid circles describe how the actual number of natives per class changes as a function of their enrollment.[12] The hollow circles describe instead the total actual class size, including immigrants, as a function of native enrollment. If the vertical distance between the hollow and the solid circles in the left panel of Figure 2 were constant at all levels of native enrollment, in Figure 1 we would not see a different shape of the numbers of natives and immigrants per class as functions of the theoretical class size C^N_{skg}. The right panel of Figure 2 plots this vertical distance (the connected hollow circles, which represent the actual number of immigrants per class) as a function of native enrollment, suggesting that this quantity is not constant. The same right panel of Figure 2 also plots the theoretically available space for immigrants (dashed line), defined as the maximum number of students in a class (25) minus the theoretical class size based on the number of natives C^N_{skg}. Note that there is a correspondence between the spikes of the space actually used for immigrants (i.e., their number per class) and the theoretically available space. Moreover the used space for immigrants tends to be relatively higher than expected for intervals of native enrollment that generate medium size classes. This result is due to the interaction between early/late

[12]We restrict the analysis to the sample of schools enrolling between 10 and 75 natives: in this range students are allocated to classes almost exactly according to what the rule prescribes. At higher levels of enrollment, the correspondence is less precise as it usually happens in this type of analysis.

enrollment and rules of class formation on the allocations of immigrants across schools and is responsible for the difference in the shapes displayed in Figure 1.

These non-collinear evolutions in the number of natives and immigrants, as a function of predicted native class size, offer the exogenous source of variation that we exploit for the identification of the effects of both the number of natives and immigrants in a class on native school performance. This is what we need to identify and estimate the PEC effect.[13]

4 Identification strategy

Abstracting for simplicity from other observables and from the aggregation of schools in institutions (to which we come back below), we would like to use the sources of variation described in the previous section to estimate the following linear approximation of the production function of native performance V^N_{jskg} (fraction of correct answers for a subject test) in class j of school s belonging to institution k and grade g:

$$V^N_{jskg} = \alpha + \beta N_{jskg} + \gamma I_{jskg} + \lambda Q^N_{jskg} + \mu Q^I_{jskg} + \epsilon_{jskg} \tag{3}$$

where N_{jskg} and I_{jskg} are, respectively, the relevant numbers of natives and immigrants. Students of the two ethnicities may have different qualities, denoted by Q^N_{jskg} and Q^I_{jskg}, that are observed by the principal but not by the econometrician. ϵ_{jskg} is noise. The parameter β (γ) measures the erosion of natives' performance due to an additional native (immigrant) for given number of immigrants (natives) and for given quality of the two ethnicities. We further assume that class performance increases with the quality of natives and immigrants, keeping constant their number, so that λ and μ are positive.

We are interested in two possible ways to characterise the effect of an immigrant inflow on student performance. The first one is offered by γ which measures the effect of an additional immigrant keeping constant the number of natives in a class. This effect is interesting, but taken in isolation, it does not tell us whether an additional immigrant has a different effect than an additional native. If principals think that these two effects differ because immigrants

[13]Similar patterns emerge in the Online Appendix when we use the math sample or both the language and math samples defining immigrant status as first generation only.

are more problematic students than natives, they will distort resources to contain the increase in class size deriving from an immigrant inflow. To understand whether this is a good use of scarce resources, we need to estimate a parameter that compares directly the effect of the two ethnic inflows. Denoting total class size with $C_{jskg} = N_{jskg} + I_{jskg}$, such a parameter is:

$$\delta = \left(\frac{dV^N_{jskg}}{dI_{jskg}} \right)_{C_{jskg}=\bar{C};Q^N_{jskg}=\bar{Q}^N;Q^I_{jskg}=\bar{Q}^I} = \gamma - \beta \tag{4}$$

and we call it the Pure Ethnic Composition effect. This is the effect of increasing *exogenously* the number of immigrants keeping class size constant (i.e., reducing the number of natives by the same amount) as well as keeping constant the quality of the two ethnicities in the class. If the PEC effect is null, immigrants are equivalent to natives and principals should not treat differently the two ethnic inflows.

The estimation of δ based on equation (3) is however problematic for two reasons. First, we need at least two non-collinear sources of variation for both N_{jskg} and I_{jskg} and, second, these sources of variation must not be correlated with the student qualities Q^N_{jskg} and Q^I_{jskg}. We claim that the institutional setting described in Section 3 offers what we need for this purpose, with the caveat that it generates exogenous variability across schools within institutions managed by the same principal, not at the level of classes within schools.

To see why, suppose first that enrollment and class formation were completely determined in February. In this case, parents decide in which school to apply based on the number and quality of the students that they expect to find there, among which immigrants, but they cannot predict whether the realised enrollment in each given school reaches a class splitting threshold. Therefore, as in a standard setting in which identification of class size effects is based on a Maimonides-type rule, the differences in student performance above and below class splitting thresholds can only depend on the differences in predicted class size induced by class splitting and are orthogonal to the quality of students that is continuous at these threshold.

Maimonides rule alone, however, would not be enough to generate two non collinear sources of variation, respectively for N_{jskg} and I_{jskg} and here is where the interaction between February and September enrollment helps. If principals knew in February the final student count, the class splitting due to Maimonides rule would produce collinear variations of both the numbers of immigrants and natives as a function of predicted native class size. But the late enrollment of some students causes some large classes planned in February with no (or few) immigrants, to split in two small classes with no (or few) immigrants when September arrives. This institutional feature generates a hump-shaped relation between immigrants and predictive native class size, while the actual number of natives grows monotonically with the same variable, as displayed in Figure 1.

5 Estimates of the Pure Ethnic Composition effect

Since our identification strategy is based on variation between schools of an institution, not between classes of a school, we aggregate the data at the school level and we estimate the following empirical counterpart of Equation (3)[14]

$$V_{skg}^{N} = \alpha + \beta N_{skg} + \gamma I_{skg} + \mu X_{skg} + \eta_{kg} + f(N_{sg}) + u_{skg}. \tag{5}$$

In this equation, the already defined dependent variable and regressors are now averages of the underlying variables at the class level, for each school s in institution k and grade g. X_{skg} is a set of predetermined control variables (averaged at the school level) for natives only[15], while η_{kg} denotes institution×grade fixed effects.[16] As in a standard setting in which identification is based on a Maimonides-type rule, a polynomial $f(N_{skg})$ in native enrollment at the school×grade level is included to control for the systematic and continuous components of the relationship between native enrollment and native performance.

The residual term u_{skg} includes the unobservable qualities of the two types of students, Q_{skg}^{N} and Q_{skg}^{I}, which are likely to be correlated with their numbers because of how principals allocate them across schools of their institution. To deal with this problem, we use the

[14]A similar equation could be defined for the performance of an average immigrant, but as explained in footnote 3, the data at our disposal have enough information to focus on the performance on natives only.

[15]Specifically, the school-level averages of the shares of mothers and fathers that have attended at most a lower secondary school, the shares of employed mothers and fathers, the share of pupils that attended kindergarten and the share of males in the class. All the specifications include also the school-level averages of the shares of native students in class that report missing values in each of these variables.

[16]See Di Liberto et al. (2015) for evidence on the effects of managerial practices on the performance of students in Italian schools.

identification strategy described in Section 4 and exploit the within-institution and across-schools variation of the average numbers of natives and immigrants that is exogenously generated by the interaction between early/late enrollment and rules of class formation.

Specifically, we adapt to our setting the empirical strategy designed by Angrist and Lang (2004) and use the set of instruments provided by the following indicator variables:

$$\Psi \in \{10(1 \leq C^N_{skg} < 11), 1(11 \leq C^N_{skg} < 12), ..., 1(24 \leq C^N_{skg} < 25\}, \tag{6}$$

These indicators are defined for each possibile level of the theoretical number of natives in a class, C^N_{skg}, predicted by equation (2) according to the rules of class formation as a function of native enrollment at the school×grade level.[17] With this approach, we can capture in the most flexible way the non-linearities and discontinuities generated by the rules of class formation that relate native enrollment to the numbers of natives and immigrants in a class.

5.1 IV estimates based on our identification strategy

In the first two columns of Table 2, grades 2 and 5 are pooled together and the outcome is the fraction of correct answers provided by natives in the language test score. In column (1), both the numbers of natives N_{skg} and the number of immigrants I_{skg} are instrumented with the indicator variables Ψ defined in equation (6). The coefficient β for natives is precisely estimated to be negative: keeping constant the number and the quality of immigrants, one additional native of average quality reduces the fraction of native correct answers in the language test score by 0.0019, about 2% of a standard deviation in test scores in the pooled sample.

Similarly negative but much larger in absolute size is the estimate of γ: keeping constant the number and quality of natives, one additional immigrant of average quality reduces the language test score of natives by 0.0176, roughly 18% of a standard deviation in languange test scores in the pooled sample. These IV estimates imply that the effect of swapping one native with one immigrant while keeping the quality of the two types of students as well

[17] Note that the number of natives in a class can potentially range between 1 and 25, but the minimum number is actually 10 in the samples that we use in our analysis because we focus on schools who enroll between 10 and 75 natives.

as class size constant, reduces the native language test score by 0.0158, roughly 16% of a standard deviation of the native performance. This is δ: the PEC effect for the language test score of natives when grades 2 and 5 are pooled.

For comparison, in column (2) of Table 2, we report results obtained instrumenting only the number of immigrants I_{skg} with the indicator variables Ψ defined in equation (6). Note that in this case the estimate of β could be inconsistent, if the number of natives in a school is correlated with the number and quality of students. However, the similarity of the estimates of β in columns (1) and (2) of Table 2 suggests that this concern is actually irrelevant in our context. Principals appear not to have much effective leeway in implementing adjustments concerning the number of natives between schools. The estimate of γ is instead smaller when only the number of immigrants is instrumented (-0.0104 instead of -0.0176), but it is in any case sizeable and statistically significant. Using this set of estimates, the PEC effect for the language test score of natives, when grades 2 and 5 are pooled, is -0.0085, which amounts to a more conservative loss of about 9% of a standard deviation in the native's language test score in the pooled sample.

The first two columns of Table 3, offer a very similar set of results for the case in which the dependent variable is the fraction of correct answers given by natives in the math test score. Also in this case both estimates of β are negative, have a similar magnitude and are statistically different from zero. The estimates of γ are similarly negative and precisely estimated, as well as substantially larger than those of β in absolute size. The PEC effect for math ranges between -0.0176 and -0.0098 respectively for the case in which both N_{skg} and I_{skg} are instrumented or the case in which this is done only for I_{skg}. These effects roughly correspond to losses of -16% and -9% of a standard deviation in the natives' math test score in the pooled sample.

In the remaining columns of Tables 2 and 3, grades 2 and 5 are analysed separately. Results are qualitatively and quantitatively similar although the estimates are less precise.

5.2 Benchmark estimates, based on different strategies

To interpret the size of these effects it is useful to compare them with the benchmark offered, in Table 4, by OLS estimates of equation (5) that exploit, for identification, the within institution variation across schools in the number of natives and immigrants. These results replicate in our context the approach followed by Contini (2013) and Ohinata and van Ours (2013), among others, who exploit the variation within schools and across classes since in their context the schools aggregation into institutions is irrelevant.[18]

The inclusion of institution fixed effects controls for a large set of unobservables, but cannot address the problem raised by the possibility that principals allocate students (in particular immigrants) across schools of their institution (or, analogously, across classes of their schools) taking into account the number of enrolled students and their quality. We therefore expect the estimates of β and γ to be smaller in size, if, for example, principals allocate more natives and/or immigrants to schools that are less crowded and in which students are of better quality.

This is indeed what we find. Pooling together both grades and focussing on the language test in column (1) of Table 4, the size of β is estimated to be about half of what we obtain in columns (1) or (2) of Table 2 (-0.0009 instead of -0.0019), while the size of γ is about four to two times smaller (-0.0038 instead of -0.0176 or -0.0104) depending on whether both natives and immigrants are instrumented or only immigrants. Therefore, also the corresponding PEC effect δ is smaller in size based on this identification strategy (-0.0029 instead of -0.0158 or -0.0085). These conclusions are qualitatively an quantitatively similar when we look separately at the two grades or when the outcome is the math test score, in the remaining columns of the relevant tables. Our results do not suffer from this attenuation bias because they are driven by the variation determined by a Maimonides-type rule in conjunction with the consequences of the interaction between February and September class splitting, which is exogenous with respect to principals' decisions.

5.3 First stages

Table 2 and 3 report the F-test on the joint significance of the instruments in the first stage regressions for all the specifications and we always reject the null in the pooled samples.

[18] The Online Appendix contains also results that exploit variation within schools and across classes, which is closer in spirit to Contini (2013) and Ohinata and van Ours (2013), but less comparable with the results reported in Tables 2 and 3, which aggregate the data at the level of schools and include institution fixed effects in the specification.

We also report the Sanderson-Windemeier (SW) F-test for two endogenous variables in the relevant columns and the p-value of the SW χ^2 statistics (Sanderson and Windmeijer 2016). First stages are presented in the Tables 5 and 6. The evidence provided by these statistics suggest that we do not face any problem of weak instruments when only the number of immigrants is considered as endogenous and we pool data from grades 2 and 5 to increase the sample size. Our instruments appear instead to be weak in the other cases.

To provide evidence in favor of the robustness of our estimates even in the presence of weak instruments, in Table 7 and 8 we report results from regressions in which we reduce the number of instruments to achieve exact identification.[19] For this exercise, in the case in which only the number of immigrants is instrumented we use as instrument the theoretical class size C^N_{skg} predicted by equation (2). When both the numbers of natives and immigrants are treated as endogenous we use, as a second instrument, a dummy variable that takes value one if predicted class size falls in its medium range of values, i.e. between the median (17.5) and the 75th percentile (20.5), and zero otherwise. This strategy is justified by the institutional setting described in Section 3, which produces a hump shaped relationship between the number of immigrants and theoretical class size (see Figure 1 and Section A of the Online Appendix).

In the language sample, when the two grades are pooled (see Table 7), we do no longer face a weak instruments problem and our qualitative results are confirmed. Point estimates are similar in size with respect to the over-identified model when only immigrants are treated as endogenous, while they are larger in absolute size when both natives and immigrants are instrumented. In the math sample, instead, estimates are less precise but similar in size to the over-identified case.[20]

Table 2 and 3 report also the p-values of the Hansen J test of over-identifying restrictions, which suggest that we cannot reject the null although the test has limited power in our setting.

[19]As suggested by Angrist and Pischke (2008) just-identified instrumental variable estimates are approximately unbiased.

[20]The first stage estimates corresponding to the instrumental variable specifications of Table 7 and 8 are reported the Online Appendix. Incidentally, they confirm that the institutional framework described in Section 3 implies a hump-shaped relation between the number of immigrants and theoretical class size (see Figure 1 and Section A of the Online Appendix).

5.4 Evidence on the validity of our identification strategy

A threat to our identification strategy is represented by the possibility that the parental decisions to enroll late in a given school depends on expectations concerning the quality and the number of students of the different ethnicities that will apply (or have applied) to the same school. To show that this threat is plausibly irrelevant in our case and thus that the variation displayed in Figures 1 and 2 is likely to be exogenous, we construct similar figures using as outcomes the average values of the control variables included in our regressions, net of institution×grade fixed effects. If our identification strategy is internally valid, we should observe that these covariates do not exhibit spikes nor hump shapes, and actually evolve smoothly over native enrollment in schools or over the corresponding predicted class size.

This is indeed the visual pattern emerging from Figures 3 and 4 where, using the language sample, we plot the residuals, after controlling for institution×grade fixed effects, of some relevant covariates (mother/father education, mother/father occupation, kindergarten attendance, gender and related dummies for missing information on these variables) against theoretical predicted class size or native enrollment respectively.[21] While in Figure 1 the number of natives is monotonically increasing in theoretical class size and the number of immigrants is a clear hump-shaped function of the same variables, in Figure 3, the residuals of these covariates, after controlling for institution×grade fixed effects, appear to be completely unrelated to theoretical class size. And while in Figure 2 the numbers of natives and immigrants follow relatively closely the rules of class formation, in Figure 4 the same residuals of covariates are smooth and relatively flat functions of native enrollment with no indication of discontinuities at the class splitting thresholds.

When we formally test whether our instruments affect covariates (see Table 9 for the language sample and the analogous table for the math sample in the Online Appendix), most coefficients associated to the instruments are estimated to be equal or very close to zero, consistently with the visual evidence described above. For the two parental education covariate and for maternal occupation, however, we detect a potentially disturbing non-negligible correlation with the instruments (see the p-value of the test for the joint significance of the instruments in the last line of Table 9). This finding is similar to what Angrist et al.

[21] The analogous evidence for the math sample is reported in the Online Appendix.

(2017a) detect using data on the INVALSI test scores in Italy and a similar identification strategy. They attribute it to the possibility of score manipulation that we will discuss below in Section 5.6. Since our results are robust to the presence of score manipulation, as shown in that section, we remain confident in the validity of our identification strategy.

As additional evidence on the validity of our identification strategy, we compare regressions that include and exclude covariates finding that the estimates of β, γ and δ are remarkably stable. The full set of results is reported in the Online Appendix, to save on space, but we provide a summary here: $\widehat{\beta}$ ranges between -0.0018/-0.0019 (language/mathemtics) including control variables and -0.0014/-0.0015 excluding them; $\widehat{\gamma}$ ranges between -0.0157/-0.0176 (language/mathemtics) including control variables and -0.0175/-0.0188 excluding them in the sample that pooles data from the 2nd and 5th grade and in all cases estimates are statistically significant. The PEC effect δ is always estimated at around -0.02. We interpret this evidence as indirect support for the internal validity of our identification strategy.

5.5 Evidence based on first generation immigrants

So far, in line with the INVALSI classification, we have defined immigrants as children born in Italy or elsewhere but from parents who are both non-Italian. If our finding of a non-negligible negative PEC effect is reasonable (ranging between 9% and 16% of a standard deviation in native's test scores in both in language and math, depending on whether one or two endogenous are considered), we should detect a larger in size PEC effect when a more restrictive definition of first generation immigrants is considered.

Tables 10 and 11 are based on a classification that defines first generation immigrants as children who, in addition to having both a non-Italian mother and father, are also born outside Italy. The complement of first generation immigrants, which we label as quasi-natives, is therefore made of natives (children with at least one Italian parent) and "second generation" immigrants (children who are born in Italy from parents who are both non-Italian).

When the language sample is considered, pooling together grades 2 and 5 in the first two columns of Table 10, the coefficient β for quasi-natives is essentially identical to the corresponding one estimated in Table 2 for natives (-0.0017 as opposed to -0.0019), while the coefficient γ for first generation immigrants is estimated to be substantially larger than the corresponding one for immigrants in the same two tables (-0.0285 and -0.0213 as opposed to -0.0176 and -0.0104, respectively for the specifications with two or one endogenous). As a result, the PEC effect δ for the language test score is estimated to be larger, as expected, when first generation immigrants are considered (between 22% and 30% of a standard deviation of the native language test score). As anticipated in Section 2 the possibility to look at the effect of first generation immigrants comes at the cost of lower quality data: while we gained access to restrictive administrative data on the number of immigrants in class, the indicator of the number of first generation immigrants in class is based on the information on students actually taking the test. As such, the latter indicator measures with error the number of students regularly attending classes and estimates based on this indicator are less precise.

A similar qualitative pattern emerges when the math sample is considered (Table 11) or when grades are split for both subjects (the remaining columns of Tables 10 and 11) but in these cases the estimates are imprecise.[22]

5.6 The possibility of test score manipulation

It has been recently suggested by Angrist et al. (2017a) that estimates of class size effects in Italy, based on rules of class formation, are heavily manipulated by teachers more as a result of shirking than because of self-interested cheating. These authors explore a variety of institutional and behavioural reasons why such manipulation is inhibited in larger classes, originating the appearance of more negative, but fictitious, effects of class size in that part of the country. This manipulation, as discussed by Angrist et al. (2017a), may induce correlation between observables and the rules of class formation.

In the light of this evidence it is possible that our estimates of the effects β and γ (and thus of their difference δ) just capture score manipulation. It is not immediately evident, however, why this manipulation should occur more frequently and intensively when class size changes because of immigrants as opposed to when it changes because of natives: i.e., why

[22]The first stage regressions corresponding to the instrumental variable estimates of Table 10 and 11 are reported in the Online Appendix.

$\gamma < \beta$ (being both negative) if manipulation were the only driving force of class size effects in Italy.

In any case, to address this issue, we show in Tables 12, respectively for the language and math test scores, that the IV-FE estimates of β and γ remain large in size, negative and statistically significant when we restrict the analysis to the sub-sample of institutions in which, according to the indicator constructed by Angrist et al. (2017a), score manipulation is likely to be minimal, if at all present.[23] The estimates of the Pure Ethnic Composition effects are therefore also similar to the corresponding ones in Tables 2 and 3 while remaining statistically significant.[24]

This evidence is relevant not only because it shows that our conclusions are largely unaffected by the score manipulation problem highlighted by Angrist et al. (2017a), but also because, in light of the above discussion in Section 5.4, they allow us to be confident in the internal validity of our identification strategy.

6 Conclusions

Anecdotal evidence of class disruption involving immigrants often generates concerns in the public opinion. These concerns, more than convincing estimates of the real dimension of the problem, typically drive educational authorities in the implementation of policies to address it. An example is the rule introduced by the Italian Ministry of Education, according to which no class should have more than 30% of immigrants: the reason why this threshold was chosen is unclear and certainly not based on experimental evidence.

This papers suggests that a useful parameter for policy makers should be the causal effect of substituting one native with one immigrant in a educational unit, net of the endogenous adjustments implemented by principals, in terms of number and quality of students, when confronted with immigrant inflows. This is what we call a Pure Ethnic Composition effect.

[23]The "cheating indicator" proposed by Angrist et al. (2017a) is based on evidence of an abnormally high performance of students in a class, an unusually small dispersion of test scores, an unusually low proportion of missing items and a high concentration in response patterns. It takes value one "for classes where score manipulation seems likely" and 0 otherwise. See Angrist et al. (2017a) for more details. We thank these authors for having shared with us the information that they constructed.

[24]The first stage estimates for these regressions are reported Online Appendix

We discuss the problems posed by the identification and estimation of this parameter and we compare it to estimates of the overall effect of an immigrant inflow, inclusive of the endogenous reactions of principals. We expect this alternative effect to be smaller in size and this is what we find.

Our results suggest that the PEC effect is sizeable: adding one immigrant to a class while taking away one native, and for given quality of students independently of their ethnicity, reduces native test scores in language and math by 16% of a standard deviation in the corresponding native test scores in our preferred specification. These losses are even larger in size when the PEC effect of fist generation immigrants is considered.

References

Angrist, Joshua D. and Kevin Lang. 2004. Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program. *American Economic Review* 94, no. 5:1613–34.

Angrist, Joshua D. and Victor Lavy. 1999. Using Maimonides Rule To Estimate The Effect of Class Size on Scholastic Achievement. *The Quarterly Journal of Economics* 114, no. 2:533–75.

Angrist, Joshua D. and Jorn-Steffen Pischke. 2008. *Mostly Harmless Econometrics*. Pricenton University Press.

Angrist, Joshua D., Erich Battistin, and Daniela Vuri. 2017a. In a Small Moment: Class Size and Moral Hazard in the Italian Mezzogiorno. *American Economic Journal: Applied Economics* 9, no. 4:216–49.

Angrist, Joshua D., Victor Lavy, and Jetson Leder-Luis, and Adi Shany. 2017b. Maimonides Rule Redux. NBER Working Paper no. 23486. National Bureau of Economic Research, Cambridge, MA.

Bossavie, Laurent. 2011. Does Immigration Affect the School Performance of Natives? Evidence From Microdata. Unpublished manuscript. Department of Economics, European University Institute.

Brunello, Giorgio and Lorenzo Rocco. 2013. The effect of immigration on the school performance of natives: Cross country evidence Using PISA test scores. *Economics of Education Review* 32:234–46.

Contini, Dalit. 2013. Immigrant background peer effects in Italian schools. *Social Science Research* 42, no. 4:1122–42.

Di Liberto, Adriana, Fabiano Schivardi, and Giovanni Sulis. 2015. Managerial practices and student performance. *Economic Policy* 30(84):683–728.

Geay, Charlotte, Sandra McNally, and Shqiponja Telhaj. 2013. Non-native Speakers of English in the Classroom: What Are the Effects on Pupil Performance. *Economic Journal* 123:F281–F307.

Gould, Eric. D., Victor Lavy, and M.Daniele Paserman. 2009. Does Immigration Affect the Long-Term Educational Outcomes of Natives? Quasi-Experimental Evidence. *Economic Journal* 119, no. 540:1243–69.

Hoxby, Caroline. 2000. Peer Effects in the Classroom: Learning from Gender and Race Variation. NBER Working paper no. 7867. National Bureau of Economic Research, Cambridge, MA.

Hunt, Jennifer. 2017. The Impact of Immigration on Educational Attainment of Natives. *Journal of Human Resources* 52, no. 4:1060–1118.

INVALSI. 2010. *Rilevazione degli Apprendimenti - SNV Prime Analisi*. `http://www.invalsi.it/download/rapporti/snv2010`.

Jensen, Peter and Astrid W. Rasmussen. 2011. The effect of immigrant concentration in schools on native and immigrant children's reading and math skills. *Economics of Education Review* 30, no. 6:1503–15.

Ohinata, Asako and Jan C. van Ours. 2013. How Immigrant Children Affect the Academic Achievement of Native Dutch Children. *Economic Journal* 123, no. 570:F308–F331.

Quintano, Claudio, Rosalia Castellano, and Sergio Longobardi. 2009. A fuzzy clustering approach to improve the accuracy of Italian student data. An experimental procedure to correct the impact of the outliers on assessment test scores. *Statistica & Applicazioni* VII, no. 2:149–71.

Sanderson, Eleanor and Frank Windmeijer. 2016. A weak instrument F-test in linear IV models with multiple endogenous variables. *Journal of Econometrics* 190, no. 2:212–21.

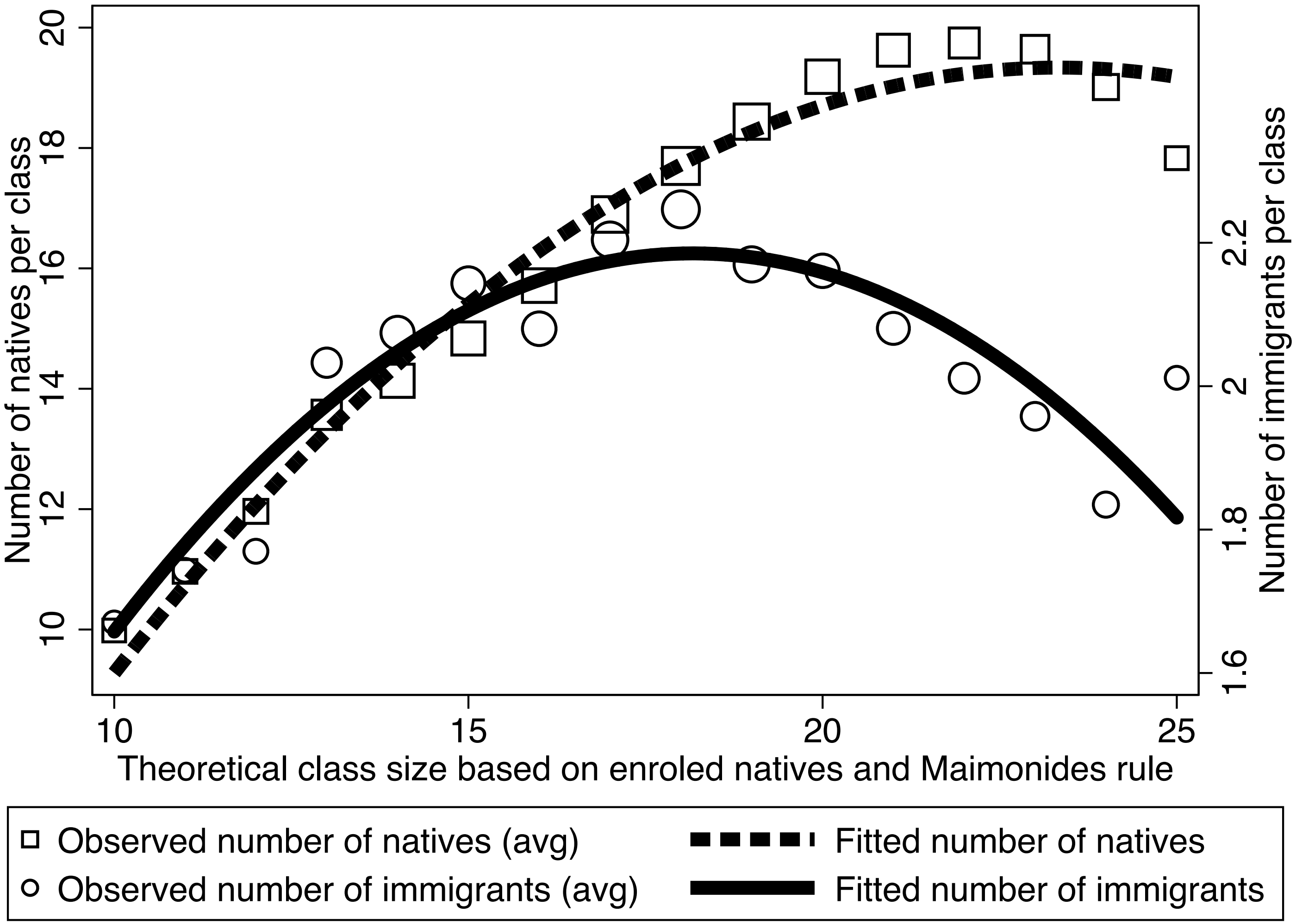
Tonello, Marco. 2016. Peer effects of non-native students on natives' educational outcomes: mechanisms and evidence. *Empirical Economics* 51, no. 1:383–414.

Table 1: Descriptive statistics for the language sample

	Pooled 2nd & 5th		2nd grade		5th grade	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
	Panel A. School characteristics					
Fraction of correct answers:						
language (natives)	0.69	0.10	0.67	0.11	0.71	0.08
mathematics (natives)	0.64	0.11	0.62	0.12	0.65	0.10
language (immigrants)	0.57	0.17	0.54	0.18	0.61	0.15
mathematics (immigrants)	0.57	0.15	0.55	0.16	0.58	0.15
Number of natives in class	16.6	3.8	16.5	3.8	16.7	3.8
Number of immigrants in class	2.0	2.0	2.1	2.0	2.1	1.99
Class size	18.7	4.0	18.6	3.9	18.7	4.0
Share (0-1) of natives in class with						
low educated mother	0.37	0.20	0.35	0.19	0.39	0.20
low educated father	0 .46	0.20	0.44	0.20	0.47	0.20
employed mother	0.62	0.22	0.63	0.22	0.62	0.22
employed father	0.96	0.07	0.96	0.07	0.96	0.07
Share (0-1) of natives in class who						
attended kindergarten	0.99	0.04	0.99	0.04	0.99	0.04
are male	0.51	0.12	0.51	0.12	0.51	0.12
Cheating propensity (Quintano et al. 2009)	0.04	0.14	0.04	0.13	0.05	0.15
Cheating indicator (Angrist et al. 2017a)	0.04	0.18	0.04	0.17	0.04	0.19
Enrollment (natives)	28.3	15.3	28.2	15.2	28.5	15.4
Enrollment (immigrants)	3.73	4.57	3.75	4.68	3.71	4.46
Average number of classes	1.68	0.77	1.68	0.77	1.68	0.78
Sample size (number of schools)	16,100		8,014		8,086	
	Panel B. Institution characteristics					
External monitored institutions	0.25	0.43	0.25	0.43	0.25	0.43
Average number of classes	4.78	1.71	4.76	1.71	4.79	1.71
Average number of schools	2.8	0.97	2.8	0.97	2.8	0.97
Sample size (number of institutions)	5,749		2,867		2,882	

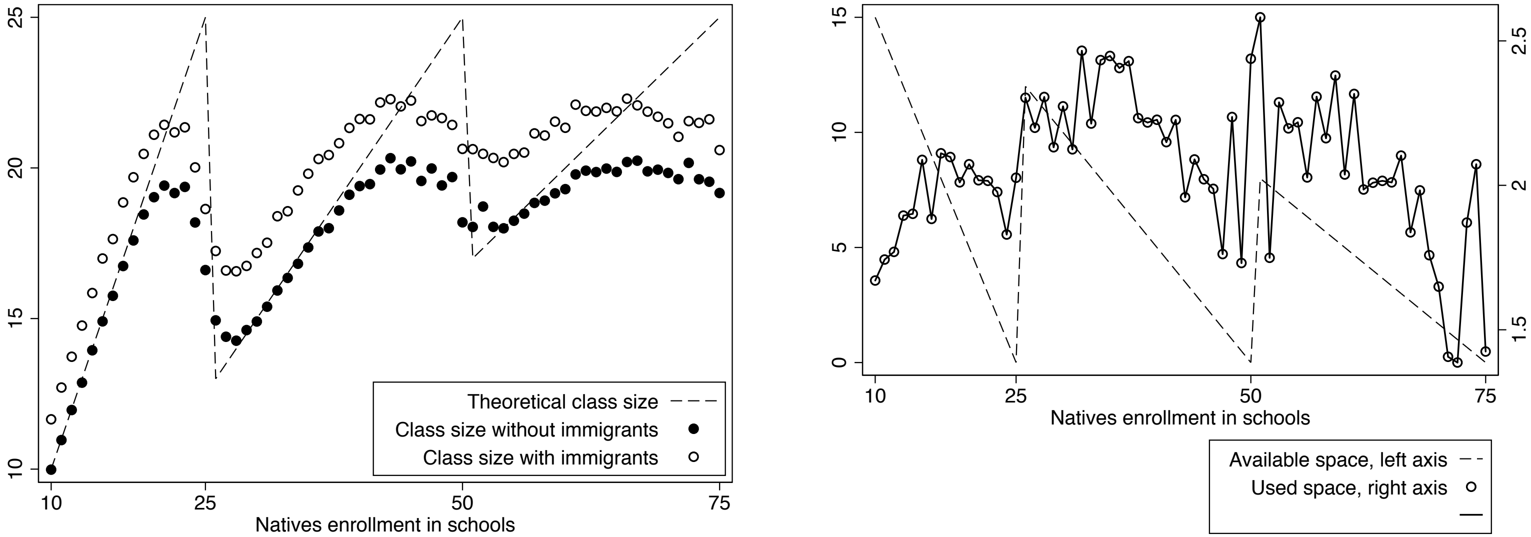
Notes: The unit of observation is a school in Panel A and an institution in Panel B. Institutions are groups of schools managed by the same principal. The family and individual background characteristics in Panel A are the school-average shares of natives in a class who have that specific characteristic over the total number of natives in the class. Children are defined as natives if they are born from at least one Italian parent, independently of the place of birth. Immigrants are instead children whose parents are both non-Italian, again independently of the place of birth. Missing values do not contribute to the computation of these shares. All regressions in the following tables include the school-average shares of missing values for each characteristic as an additional control. All these variables come from the school administration through the data file that we received from INVALSI, except for the cheating indicator that was computed by Angrist et al. (2017a) and kindly given to us by these authors. The Online Appendix contains analogous descriptive statistics for the math sample.

Figure 1: Natives and immigrants in a class as a function of theoretical class size based on native enrollment; language sample



Notes: In this figure, squares (left vertical axis) indicate the average number of natives per class in schools with the correspondent theoretical class size based on enrolled natives (horizontal axis). The dashed line is a quadratic fit of these averages. Circles (right vertical axis) indicate the average number of immigrants per class in schools with the correspondent theoretical class size based on enrolled natives (horizontal axis). The continuous line is a quadratic fit of these averages. The size of squares and circles is proportional to the number of schools used to compute the averages that they represent. The quadratic fitted lines have been estimated with weights equal to the number of schools for each value of theoretical class size based on enrolled natives (horizontal axis). The 2nd and 5th grades in each school are pooled. The Online Appendix contains an analogous figure for the math sample.

Figure 2: Natives and immigrants in a class and class size as a function of native enrollment; language sample.



Notes: The left panel reports the theoretical class size (dashed line), the class size without immigrants (solid circles) and the class size with immigrants (hollow circles) as a function of native enrollment in schools. In the right panel, the line connecing hollow circles represent the vertical distance between the hollow and solid circles of the left panel (the actual number of immigrants per class) as a function of native enrollment. The right panel also plots the theoretically available space for immigrants (dashed line), defined as the maximum number of students in a class (25) minus the theoretical class size based on the number of natives C_{sg}^{N}. The 2nd and 5th grades in each school are pooled. The Online Appendix contains an analogous figure for the math sample.

Table 2: IV-FE estimates of the effect of the number of natives and immigrants on the language test score of natives.

	Pooled 2nd & 5th		2nd grade		5th grade	
	Two	One	Two	One	Two	One
	endogenous		endogenous		endogenous	
	(1)	(2)	(3)	(4)	(5)	(6)
Number of natives: $\hat{\beta}$	-0.0019***	-0.0018***	-0.0022***	-0.0021***	-0.0014***	-0.0014***
	(0.0005)	(0.0005)	(0.0008)	(0.0008)	(0.0005)	(0.0005)
Number of immigrants: $\hat{\gamma}$	-0.0176**	-0.0104***	-0.0163	-0.0086**	-0.0108	-0.0108***
	(0.0078)	(0.0029)	(0.0135)	(0.0043)	(0.0073)	(0.0037)
Pure Ethnic Composition effect: $\hat{\delta}$	-0.0158**	-0.0085***	-0.0141	-0.0065*	-0.0094	-0.0093***
	(0.0077)	(0.0025)	(0.0131)	(0.0037)	(0.0072)	(0.0033)
Observations	16,100	16,100	8,014	8,014	8,086	8,086
Institution×grade FE	✓	✓	✓	✓	✓	✓
Polynomial in natives enrollment	✓	✓	✓	✓	✓	✓
School level controls	✓	✓	✓	✓	✓	✓
Hansen (p-value)	0.8871	0.8310	0.9120	0.9052	0.6432	0.7130
F stat (natives)	387.0621		185.8595		207.6623	
SW F stat (natives)	21.8812		4.2407		63.9004	
SW χ^2 p-value (natives)	0.00		0.00		0.00	
F stat (immigrants)	2.9252	18.0111	1.4659	9.5939	2.2578	9.0044
SW F stat (immigrants)	2.7954	18.0111	1.1612	9.5939	2.3781	9.0044
SW χ^2 p-value (immigrants)	0.00	0.00	0.34	0.00	0.00	0.00

Notes: The table reports in each column a different set of estimates of equation (5) for the language sample. The unit of observation is a school. The dependent variable is the fraction of correct answers of natives in the language test. The instruments are a set of 15 dummies, one for each level of the theoretical number of natives in a class predicted by equation (2) according to the rules of class formation as a function of native enrollment at the school×grade level. The omitted category corresponds to a number of natives in a class equal to 25. There are no school-grades in which the number of natives in a class is less than 10. All regressions include a 2nd order polynomial of natives enrollment at the school×grade level. The controls are aggregated at the school level and include the following set of family and individual covariates: the share of natives with mothers and fathers who attended, at most, a lower secondary school, the shares of natives with employed mothers and fathers, the share of natives who attended kindergarten and the share of male natives in the school. All regressions include also the share of native students who report missing values in each of these variables as well as institution×grade fixed effects. Robust standard errors clustered at the institution-grade level are reported in parentheses. A * denotes significance at 10%; a ** denotes significance at 5%; a *** denotes significance at 1%. The table reports also: i) the p-value of the Hansen test; ii) the value of the F test of null hypothesis that the coefficients of the instruments are all zero in each first stage equation; and iii) the Sanderson-Windemeier (SW) first stage F statistic of each individual endogenous regressor (in the case of a model with one endogenous regressor this coincides with the F-test on excluded instruments) to test for weak identification ; and iv) the p-value of Sanderson-Windemeier χ^2 statistic of each individual endogenous regressor to test for under-identification (Sanderson and Windmeijer 2016).

Table 3: IV-FE estimates of the effect of the number of natives and immigrants on the math test score of natives.

	Pooled 2nd & 5th		2nd grade		5th grade	
	Two	One	Two	One	Two	One
	endogenous		endogenous		endogenous	
	(1)	(2)	(3)	(4)	(5)	(6)
Number of natives: $\hat{\beta}$	-0.0019***	-0.0018***	-0.0022**	-0.0020**	-0.0014**	-0.0014**
	(0.0006)	(0.0005)	(0.0009)	(0.0009)	(0.0007)	(0.0007)
Number of immigrants: $\hat{\gamma}$	-0.0195**	-0.0117***	-0.0198	-0.0099**	-0.0090	-0.0117**
	(0.0091)	(0.0034)	(0.0148)	(0.0047)	(0.0092)	(0.0047)
Pure Ethnic Composition effect: $\hat{\delta}$	-0.0176**	-0.0098***	-0.0176	-0.0079**	-0.0076	-0.0103**
	(0.0089)	(0.0029)	(0.0143)	(0.0040)	(0.0090)	(0.0042)
Observations	16,091	16,091	8,006	8,006	8,085	8,085
Institution×grade FE	✓	✓	✓	✓	✓	✓
Polynomial in natives enrollment	✓	✓	✓	✓	✓	✓
School level controls	✓	✓	✓	✓	✓	✓
Hansen (p-value)	0.9241	0.9002	0.9375	0.9314	0.9204	0.9464
F stat (natives)	392.2346		191.4878		206.6452	
SW F stat (natives)	19.9219		3.8105		61.4364	
SW χ^2 p-value (natives)	0.00		0.00		0.00	
F stat (immigrants)	2.8739	17.7883	1.4590	9.4916	2.2368	8.8869
SW F stat (immigrants)	2.7227	17.7883	1.1203	9.4916	2.3530	8.8869
SW χ^2 p-value (immigrants)	0.00	0.00	0.33	0.00	0.00	0.00

Notes: The table reports in each column a different set of estimates of equation (5) for the math sample. The unit of observation is a school. The dependent variable is the fraction of correct answers of natives in the math test. The instruments are a set of 15 dummies, one for each level of the theoretical number of natives in a class predicted by equation (2) according to the rules of class formation as a function of native enrollment at the school×grade level. The omitted category corresponds to a number of natives in a class equal to 25. There are no school-grades in which the number of natives in a class is less than 10. All regressions include a 2nd order polynomial of natives enrollment at the school×grade level. The controls are aggregated at the school level and include the following set of family and individual covariates: the share of natives with mothers and fathers who attended, at most, a lower secondary school, the shares of natives with employed mothers and fathers, the share of natives who attended kindergarten and the share of male natives in the school. All regressions include also the share of native students who report missing values in each of these variables as well as institution×grade fixed effects. Robust standard errors clustered at the institution-grade level are reported in parentheses. A * denotes significance at 10%; a ** denotes significance at 5%; a *** denotes significance at 1%. The table reports also: i) the p-value of the Hansen test; ii) the value of the F test of null hypothesis that the coefficients of the instruments are all zero in each first stage equation; and iii) the Sanderson-Windemeier first stage F statistic of each individual endogenous regressor (in the case of a model with one endogenous regressor this coincides with the F-test on excluded instruments) to test for weak identification ; and iv) the p-value of Sanderson-Windemeier χ^2 statistic of each individual endogenous regressor to test for under-identification (Sanderson and Windmeijer 2016).

Table 4: OLS-FE estimates of the effect of the number of natives and immigrants on the language and math test scores of natives.

	Language			Mathematics		
	Pooled	2nd grade	5th grade	Pooled	2nd grade	5th grade
	(1)	(2)	(3)	(4)	(5)	(6)
Number of natives: $\hat{\beta}$	-0.0009***	-0.0013***	-0.0006*	-0.0010***	-0.0013***	-0.0007**
	(0.0002)	(0.0004)	(0.0003)	(0.0003)	(0.0004)	(0.0004)
Number of immigrants: $\hat{\gamma}$	-0.0038***	-0.0047***	-0.0028***	-0.0041***	-0.0044***	-0.0039***
	(0.0005)	(0.0008)	(0.0006)	(0.0006)	(0.0008)	(0.0007)
Confounded ethnic composition effect: $\hat{\delta}$	-0.0029***	-0.0034***	-0.0022***	-0.0031***	-0.0031***	-0.0031***
	(0.0005)	(0.0008)	(0.0006)	(0.0006)	(0.0008)	(0.0007)
Observations	16,100	8,014	8,086	16,091	8,006	8,085
Institution×grade FE	✓	✓	✓	✓	✓	✓
School level controls	✓	✓	✓	✓	✓	✓

Notes: The table reports in each column a different set of OLS estimates of equation (5) for the language and maths samples. The unit of observation is a school. The dependent variable is the fraction of correct answers of natives in the language (math) test. The controls are aggregated at the school level and include the following set of family and individual covariates: the share of natives with mothers and fathers who attended, at most, a lower secondary school, the shares of natives with employed mothers and fathers, the share of natives who attended kindergarten and the share of male natives in the school. All regressions include also the share of native students who report missing values in each of these variables as well as institution×grade fixed effects. Robust standard errors clustered at the institution-grade level are reported in parentheses.A * denotes significance at 10%; a ** denotes significance at 5%; a *** denotes significance at 1%. The Online Appendix contains an analogous table in which the unit of observation is a class and we include school-grade fixed effects, more in line with what is typically done in the literature (e.g., Contini 2013 and Ohinata and van Ours 2013) that exploit within school variation across classes in contexts in which the aggregation into institutions is irrelevant. In the Online Appendix we also report a Table where we replicate the insitution-grade fixed effects estimates on the class-level sample, and we compare all these estimates.

Table 5: First stage for the number natives and immigrants; language sample.

	Pooled 2nd & 5th			2nd grade			5th grade		
	Two endogenous		One endogenous	Two endogenous		One endogenous	Two endogenous		One endogenous
	N	I	I	N	I	I	N	I	I
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$1(10 \leq C_{sg}^{N} < 11)$	-6.03***	-0.13	-1.31***	-5.66***	-0.13	-1.35***	-6.44***	-0.15	-1.25***
	(0.29)	(0.14)	(0.15)	(0.41)	(0.21)	(0.21)	(0.42)	(0.20)	(0.21)
$1(11 \leq C_{sg}^{N} < 12)$	-5.14***	-0.12	-1.12***	-4.78***	-0.24	-1.27***	-5.55***	0.02	-0.93***
	(0.29)	(0.14)	(0.14)	(0.40)	(0.20)	(0.19)	(0.41)	(0.18)	(0.19)
$1(12 \leq C_{sg}^{N} < 13)$	-3.93***	-0.13	-0.90***	-3.56***	-0.23	-1.00***	-4.33***	-0.03	-0.77***
	(0.28)	(0.14)	(0.14)	(0.40)	(0.21)	(0.20)	(0.41)	(0.18)	(0.18)
$1(13 \leq C_{sg}^{N} < 14)$	-2.76***	-0.02	-0.55***	-2.43***	-0.13	-0.65***	-3.13***	0.10	-0.43***
	(0.29)	(0.12)	(0.11)	(0.40)	(0.17)	(0.16)	(0.41)	(0.16)	(0.16)
$1(14 \leq C_{sg}^{N} < 15)$	-2.32***	-0.06	-0.51***	-2.08***	-0.15	-0.60***	-2.59***	0.03	-0.41***
	(0.28)	(0.12)	(0.11)	(0.39)	(0.18)	(0.16)	(0.40)	(0.15)	(0.15)
$1(15 \leq C_{sg}^{N} < 16)$	-1.58***	-0.09	-0.40***	-1.35***	-0.12	-0.41***	-1.83***	-0.07	-0.38**
	(0.28)	(0.12)	(0.11)	(0.39)	(0.18)	(0.16)	(0.40)	(0.15)	(0.15)
$1(16 \leq C_{sg}^{N} < 17)$	-0.74***	-0.06	-0.20*	-0.58	-0.18	-0.31*	-0.95**	0.08	-0.09
	(0.27)	(0.12)	(0.11)	(0.38)	(0.17)	(0.16)	(0.39)	(0.15)	(0.15)
$1(17 \leq C_{sg}^{N} < 18)$	0.29	0.01	0.06	0.56	-0.19	-0.06	-0.02	0.21	0.20
	(0.27)	(0.11)	(0.10)	(0.38)	(0.17)	(0.15)	(0.39)	(0.15)	(0.14)
$1(18 \leq C_{sg}^{N} < 19)$	0.96***	0.01	0.20*	1.16***	-0.13	0.12	0.72*	0.17	0.29**
	(0.27)	(0.11)	(0.10)	(0.38)	(0.17)	(0.15)	(0.39)	(0.15)	(0.14)
$1(19 \leq C_{sg}^{N} < 20)$	1.64***	-0.06	0.26**	1.88***	-0.18	0.23	1.36***	0.08	0.31**
	(0.28)	(0.11)	(0.10)	(0.39)	(0.17)	(0.15)	(0.39)	(0.15)	(0.14)
$1(20 \leq C_{sg}^{N} < 21)$	2.25***	-0.13	0.31***	2.57***	-0.28*	0.27*	1.91***	0.03	0.36**
	(0.28)	(0.11)	(0.11)	(0.39)	(0.17)	(0.15)	(0.39)	(0.15)	(0.15)
$1(21 \leq C_{sg}^{N} < 22)$	2.73***	-0.23*	0.31***	2.94***	-0.37**	0.26*	2.49***	-0.07	0.35**
	(0.29)	(0.12)	(0.11)	(0.41)	(0.17)	(0.16)	(0.40)	(0.15)	(0.15)
$1(22 \leq C_{sg}^{N} < 23)$	2.57***	-0.30**	0.20*	3.01***	-0.44**	0.21	2.10***	-0.15	0.21
	(0.30)	(0.12)	(0.11)	(0.42)	(0.18)	(0.16)	(0.43)	(0.16)	(0.15)
$1(23 \leq C_{sg}^{N} < 24)$	2.47***	-0.31**	0.17	2.30***	-0.31*	0.19	2.61***	-0.31*	0.14
	(0.32)	(0.12)	(0.11)	(0.45)	(0.18)	(0.16)	(0.44)	(0.17)	(0.16)
$1(24 \leq C_{sg}^{N} < 25)$	1.72***	-0.37***	-0.04	1.94***	-0.52***	-0.10	1.45***	-0.19	0.05
	(0.32)	(0.13)	(0.12)	(0.46)	(0.19)	(0.16)	(0.46)	(0.17)	(0.16)
Institution×grade FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Polynomial in natives enrollment	✓	✓	✓	✓	✓	✓	✓	✓	
School level controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	16,100	16,100	16,100	8,014	8,014	8,014	8,086	8,086	8,086
F stat	387.06	2.93	18.01	185.86	1.47	9.59	207.66	2.26	9.00
SW F stat	21.88	2.80	18.01	4.24	1.16	9.59	63.90	2.38	9.00
SW χ^2 p-value	0.00	0.00	0.00	0.00	0.34	0.00	0.00	0.00	0.00

Notes: The table reports in each column a different first stage regression correspondent to the IV estimates of Table 2. The unit of observation is a school. The dependent variable is the average number of natives (immigrants) per a class in a school. The instruments are a set of 15 dummies, one for each level of the theoretical number of natives in a class predicted by equation (2) according to the rules of class formation as a function of native enrollment at the school×grade level. The omitted category corresponds to a number of natives in a class equal to 25. There are no school-grades in which the number of natives in a class is less than 10. All regressions include a 2nd order polynomial of natives enrollment at the school×grade level. The controls are aggregated at the school level and include the following set of family and individual covariates: the share of natives with mothers and fathers who attended, at most, a lower secondary school, the shares of natives with employed mothers and fathers, the share of natives who attended kindergarten and the share of male natives in the school. All regressions include also the share of native students who report missing values in each of these variables as well as institution×grade fixed effects. Robust standard errors clustered at the institution-grade level are reported in parentheses. A * denotes significance at 10%; a ** denotes significance at 5%; a *** denotes significance at 1%. The table reports also: i) the value of the F test of null hypothesis that the coefficients of the instruments are all zero in each first stage equation; and ii) the Sanderson-Windemeier first stage F statistic of each individual endogenous regressor (in the case of a model with one endogenous regressor this coincides with the F-test on excluded instruments) to test for weak identification ; and iii) the p-value of Sanderson-Windemeier χ^2 statistic of each individual endogenous regressor to test for under-identification (Sanderson and Windmeijer 2016).

Table 6: First stage for the number natives and immigrants; math sample.

	Pooled 2nd & 5th			2nd grade			5th grade		
	Two endogenous		One endogenous	Two endogenous		One endogenous	Two endogenous		One endogenous
	N	I	I	N	I	I	N	I	I
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$1(10 \leq C^N_{sg} < 11)$	-5.96***	-0.12	-1.28***	-5.61***	-0.11	-1.33***	-6.34***	-0.14	-1.22***
	(0.29)	(0.14)	(0.15)	(0.40)	(0.21)	(0.21)	(0.42)	(0.20)	(0.21)
$1(11 \leq C^N_{sg} < 12)$	-5.06***	-0.11	-1.09***	-4.71***	-0.24	-1.26***	-5.46***	0.03	-0.90***
	(0.29)	(0.14)	(0.14)	(0.40)	(0.20)	(0.19)	(0.42)	(0.18)	(0.19)
$1(12 \leq C^N_{sg} < 13)$	-3.85***	-0.13	-0.88***	-3.47***	-0.23	-0.99***	-4.24***	-0.03	-0.75***
	(0.29)	(0.14)	(0.14)	(0.40)	(0.21)	(0.20)	(0.42)	(0.18)	(0.18)
$1(13 \leq C^N_{sg} < 14)$	-2.68***	-0.02	-0.54***	-2.36***	-0.14	-0.65***	-3.03***	0.10	-0.41***
	(0.29)	(0.12)	(0.11)	(0.40)	(0.18)	(0.16)	(0.42)	(0.16)	(0.16)
$1(14 \leq C^N_{sg} < 15)$	-2.24***	-0.06	-0.49***	-1.99***	-0.15	-0.58***	-2.51***	0.04	-0.39**
	(0.28)	(0.12)	(0.11)	(0.39)	(0.18)	(0.16)	(0.40)	(0.15)	(0.15)
$1(15 \leq C^N_{sg} < 16)$	-1.49***	-0.09	-0.38***	-1.24***	-0.13	-0.40**	-1.75***	-0.07	-0.36**
	(0.28)	(0.12)	(0.11)	(0.38)	(0.18)	(0.16)	(0.40)	(0.15)	(0.15)
$1(16 \leq C^N_{sg} < 17)$	-0.65**	-0.06	-0.18*	-0.48	-0.19	-0.29*	-0.86**	0.08	-0.06
	(0.28)	(0.12)	(0.11)	(0.38)	(0.17)	(0.16)	(0.40)	(0.15)	(0.15)
$1(17 \leq C^N_{sg} < 18)$	0.38	0.00	0.08	0.66*	-0.20	-0.05	0.07	0.21	0.22
	(0.27)	(0.11)	(0.10)	(0.38)	(0.17)	(0.15)	(0.40)	(0.15)	(0.14)
$1(18 \leq C^N_{sg} < 19)$	1.05***	0.02	0.22**	1.24***	-0.13	0.14	0.82**	0.17	0.31**
	(0.28)	(0.11)	(0.10)	(0.38)	(0.17)	(0.15)	(0.40)	(0.15)	(0.15)
$1(19 \leq C^N_{sg} < 20)$	1.71***	-0.06	0.28***	1.96***	-0.19	0.24	1.44***	0.08	0.33**
	(0.28)	(0.11)	(0.10)	(0.39)	(0.17)	(0.15)	(0.40)	(0.15)	(0.14)
$1(20 \leq C^N_{sg} < 21)$	2.33***	-0.13	0.32***	2.65***	-0.29*	0.28*	2.00***	0.04	0.38**
	(0.28)	(0.11)	(0.11)	(0.39)	(0.17)	(0.15)	(0.40)	(0.15)	(0.15)
$1(21 \leq C^N_{sg} < 22)$	2.81***	-0.23*	0.32***	3.03***	-0.38**	0.27*	2.57***	-0.07	0.37**
	(0.29)	(0.12)	(0.11)	(0.41)	(0.17)	(0.16)	(0.41)	(0.15)	(0.15)
$1(22 \leq C^N_{sg} < 23)$	2.66***	-0.29**	0.22**	3.10***	-0.44**	0.23	2.18***	-0.14	0.23
	(0.30)	(0.12)	(0.11)	(0.42)	(0.18)	(0.16)	(0.43)	(0.16)	(0.15)
$1(23 \leq C^N_{sg} < 24)$	2.56***	-0.31**	0.19	2.40***	-0.32*	0.20	2.69***	-0.31*	0.15
	(0.32)	(0.12)	(0.11)	(0.45)	(0.18)	(0.16)	(0.45)	(0.17)	(0.16)
$1(24 \leq C^N_{sg} < 25)$	1.80***	-0.37***	-0.02	2.03***	-0.52***	-0.08	1.53***	-0.19	0.07
	(0.33)	(0.13)	(0.12)	(0.46)	(0.19)	(0.16)	(0.47)	(0.17)	(0.17)
Institution×grade FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Polynomial in natives enrollment	✓	✓	✓	✓	✓	✓	✓	✓	
School level controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	16,091	16,091	16,091	8,006	8,006	8,006	8,085	8,085	8,085
F stat	392.23	2.87	17.79	191.49	1.46	9.49	206.65	2.24	8.89
SW F stat	19.92	2.72	17.79	3.81	1.12	9.49	61.44	2.35	8.89
SW χ^2 p-value	0.00	0.00	0.00	0.00	0.33	0.00	0.00	0.00	0.00

Notes: The table reports in each column a different first stage regression correspondent to the IV estimates of Table 3. The unit of observation is a school. The dependent variable is the average number of natives (immigrants) per a class in a school. The instruments are a set of 15 dummies, one for each level of the theoretical number of natives in a class predicted by equation (2) according to the rules of class formation as a function of native enrollment at the school×grade level. The omitted category corresponds to a number of natives in a class equal to 25. There are no school-grades in which the number of natives in a class is less than 10. All regressions include a 2nd order polynomial of natives enrollment at the school×grade level. The controls are aggregated at the school level and include the following set of family and individual covariates: the share of natives with mothers and fathers who attended, at most, a lower secondary school, the shares of natives with employed mothers and fathers, the share of natives who attended kindergarten and the share of male natives in the school. All regressions include also the share of native students who report missing values in each of these variables as well as institution×grade fixed effects. Robust standard errors clustered at the institution-grade level are reported in parentheses. A * denotes significance at 10%; a ** denotes significance at 5%; a *** denotes significance at 1%. The table reports also: i) i) the value of the F test of null hypothesis that the coefficients of the instruments are all zero in each first stage equation; and ii) the Sanderson-Windemeier first stage F statistic of each individual endogenous regressor (in the case of a model with one endogenous regressor this coincides with the F-test on excluded instruments) to test for weak identification ; and iii) the p-value of Sanderson-Windemeier χ^2 statistic of each individual endogenous regressor to test for under-identification (Sanderson and Windmeijer 2016).

Table 7: IV-FE estimates with exact identification of the effect of the number of natives and immigrants on the language test score of natives.

	Pooled 2nd & 5th		2nd grade		5th grade	
	Two	One	Two	One	Two	One
	endogenous		endogenous		endogenous	
	(1)	(2)	(3)	(4)	(5)	(6)
Number of natives: $\hat{\beta}$	-0.0024***	-0.0018***	-0.0029***	-0.0020**	-0.0020***	-0.0017**
	(0.0006)	(0.0006)	(0.0010)	(0.0010)	(0.0007)	(0.0008)
Number of immigrants: $\hat{\gamma}$	-0.0269**	-0.0102**	-0.0336	-0.0082	-0.0223*	-0.0128**
	(0.0111)	(0.0043)	(0.0228)	(0.0059)	(0.0112)	(0.0062)
Pure Ethnic Composition effect: $\hat{\delta}$	-0.0245**	-0.0084**	-0.0306	-0.0062	-0.0203*	-0.0111**
	(0.0109)	(0.0037)	(0.0222)	(0.0050)	(0.0110)	(0.0056)
Observations	16,100	16,100	8,014	8,014	8,086	8,086
Institution×grade FE	✓	✓	✓	✓	✓	✓
Polynomial in natives enrollment	✓	✓	✓	✓	✓	✓
School level controls	✓	✓	✓	✓	✓	✓
F stat (natives)	1437.3606		699.7993		736.2048	
SW F stat (natives)	97.2990		19.7752		105.1318	
SW χ^2 p-value (natives)	0.00		0.00		0.00	
F stat (immigrants)	13.0868	115.5303	4.7531	75.4156	8.7353	42.4370
SW F stat (immigrants)	21.2063	115.5303	6.6758	75.4156	15.2886	42.4370
SW χ^2 p-value (immigrants)	0.00	0.00	0.01	0.00	0.00	0.00

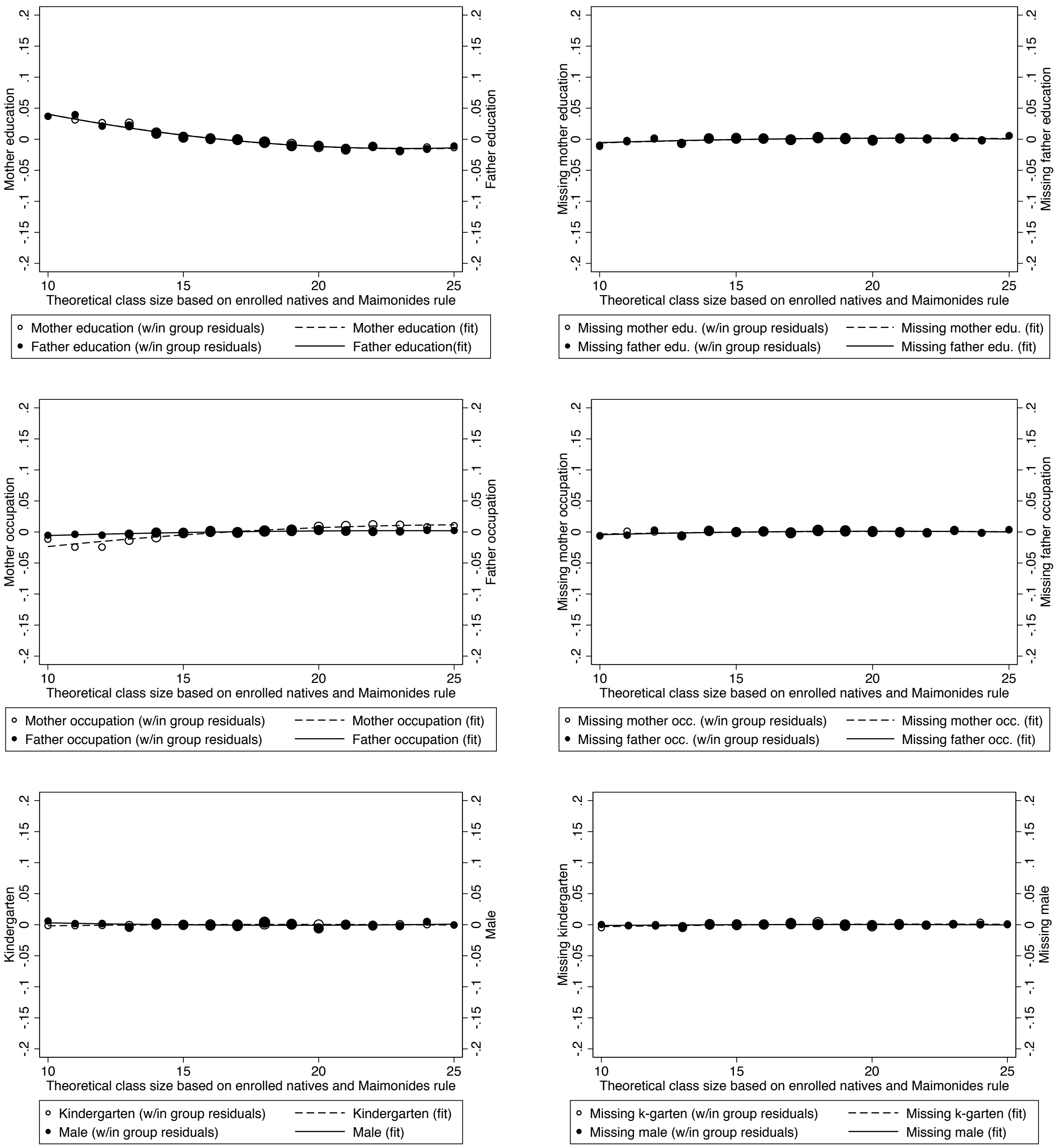
Notes: The table reports in each column a different set of estimates of equation (5) for the language samples. The unit of observation is a school. The dependent variable is the average test scores in language for natives students (i.e. fraction of correct answers). In the case of one endogenous variable, we use as instruments the theoretical number of natives in a class predicted by equation (2) according to the rules of class formation as a function of native enrollment at the school×grade level. In the case of two endogenous variables, we add to the predicted class size a dummy variable that takes the value one if the predicted class size falls in the medium range (i.e. between the median -17.5 native students - and the 75th percentile -20.5 native students-) and zero otherwise. All regressions include a 2nd order polynomial of natives enrollment at the school×grade level. The controls are aggregated at the school level and include the following set of family and individual covariates: the share of natives with mothers and fathers who attended, at most, a lower secondary school, the shares of natives with employed mothers and fathers, the share of natives who attended kindergarten and the share of male natives in the school. All regressions include also the share of native students who report missing values in each of these variables as well as institution×grade fixed effects. Robust standard errors clustered at the institution-grade level are reported in parentheses. A * denotes significance at 10%; a ** denotes significance at 5%; a *** denotes significance at 1%. The table reports also: i) the value of the F test of null hypothesis that the coefficients of the instruments are all zero in each first stage equation; and ii) the Sanderson-Windemeier first stage F statistic of each individual endogenous regressor (in the case of a model with one endogenous regressor this coincides with the F-test on excluded instruments) to test for weak identification ; and iii) the p-value of Sanderson-Windemeier χ^2 statistic of each individual endogenous regressor to test for under-identification (Sanderson and Windmeijer 2016).

Table 8: IV-FE estimates with exact identification of the effect of the number of natives and immigrants on math test scores of natives.

	Pooled 2nd & 5th		2nd grade		5th grade	
	Two	One	Two	One	Two	One
	endogenous		endogenous		endogenous	
	(1)	(2)	(3)	(4)	(5)	(6)
Number of natives: $\hat{\beta}$	-0.0017***	-0.0017**	-0.0018*	-0.0016	-0.0016**	-0.0017*
	(0.0006)	(0.0007)	(0.0010)	(0.0010)	(0.0008)	(0.0010)
Number of immigrants: $\hat{\gamma}$	-0.0118	-0.0104**	-0.0121	-0.0070	-0.0116	-0.0149*
	(0.0117)	(0.0049)	(0.0222)	(0.0063)	(0.0130)	(0.0079)
Pure Ethnic Composition effect: $\hat{\delta}$	-0.0101	-0.0087**	-0.0103	-0.0054	-0.0100	-0.0132*
	(0.0114)	(0.0043)	(0.0216)	(0.0053)	(0.0127)	(0.0070)
Observations	16,091	16,091	8,006	8,006	8,085	8,085
Institution×grade FE	✓	✓	✓	✓	✓	✓
Polynomial in natives enrollment	✓	✓	✓	✓	✓	✓
School level controls	✓	✓	✓	✓	✓	✓
F stat (natives)	1438.4585		699.9137		737.7495	
SW F stat (natives)	95.9696		18.7907		105.2888	
SW χ^2 p-value (natives)	0.00		0.00		0.00	
F stat (immigrants)	13.0614	115.8777	4.7519	75.2139	8.7703	42.7929
SW F stat (immigrants)	21.0931	115.8777	6.5515	75.2139	15.3253	42.7929
SW χ^2 p-value (immigrants)	0.00	0.00	0.01	0.00	0.00	0.00

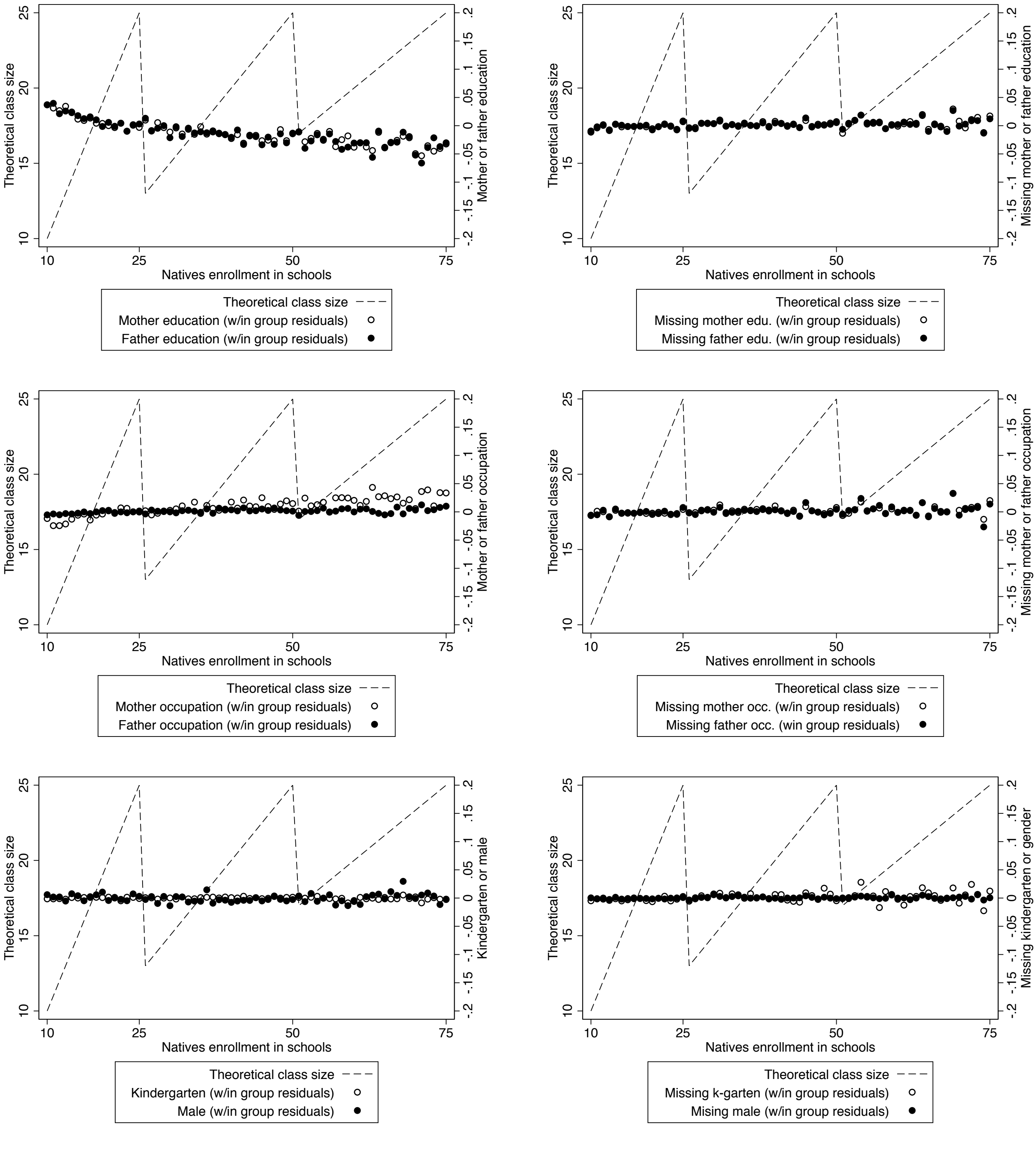
Notes: The table reports in each column a different set of estimates of equation (5) for the language samples. The unit of observation is a school. The dependent variable is the average test scores in mathematics for natives students (i.e. fraction of correct answers). In the case of one endogenous variable, we use as instruments the theoretical number of natives in a class predicted by equation (2) according to the rules of class formation as a function of native enrollment at the school×grade level. In the case of two endogenous variables, we add to the predicted class size a dummy variable that takes the value one if the predicted class size falls in the medium range (i.e. between the median -17.5 native students - and the 75th percentile -20.5 native students-) and zero otherwise. All regressions include a 2nd order polynomial of natives enrollment at the school×grade level. The controls are aggregated at the school level and include the following set of family and individual covariates: the share of natives with mothers and fathers who attended, at most, a lower secondary school, the shares of natives with employed mothers and fathers, the share of natives who attended kindergarten and the share of male natives in the school. All regressions include also the share of native students who report missing values in each of these variables as well as institution×grade fixed effects. Robust standard errors clustered at the institution-grade level are reported in parentheses. A * denotes significance at 10%; a ** denotes significance at 5%; a *** denotes significance at 1%. The table reports also: i) the value of the F test of null hypothesis that the coefficients of the instruments are all zero in each first stage equation; and ii) the Sanderson-Windemeier first stage F statistic of each individual endogenous regressor (in the case of a model with one endogenous regressor this coincides with the F-test on excluded instruments) to test for weak identification ; and iii) the p-value of Sanderson-Windemeier χ^2 statistic of each individual endogenous regressor to test for under-identification (Sanderson and Windmeijer 2016).

Figure 3: Covariates as a function of theoretical class size based on native enrollment; language sample.



Notes: In these panels, solid and hollow circles indicate the averages of the residual, after controlling for institution×grade fixed effects, of the covariates indicated in the legend and in the corresponding vertical axis. These covariates are included in equation (5), whose IV-FE estimates are reported Table 2. The solid and dashed lines represent quadratic fits of these averages. The size of circles is proportional to the number of schools used to compute the averages that they represent. The quadratic fitted lines have been estimated with weights equal to the number of schools for each value of theoretical class size based on enrolled natives (horizontal axis). The Online Appendix contains an analogous figure for the math sample.

Figure 4: Covariates as a function of native enrollment in schools; language sample.



Notes: In these panels, we plot, as a function of native enrollment in schools, the theoretical class size (dashed line) and the averages (solid or hollow circles) of the residual, after controlling for institution×grade fixed effects, of the covariates indicated in the legend and in the corresponding vertical axis. These covariates are included in equation (5), whose IV-FE estimates are reported Table 2. The Online Appendix contains an analogous figure for the math sample.

Table 9: OLS estimates of the effect of instruments on covariates; language sample.

Outcomes	Education		Employed		Kindergarten	Male	Missing					
							Education		Employed		Kindergarten	Male
	Mother	Father	Mother	Father			Mother	Father	Mother	Father		
Regressors												
$1(10 \leq C_{sg}^N < 11)$	0.03***	0.01	-0.00	-0.01	-0.00	-0.00	-0.02***	-0.02***	-0.01*	-0.01**	0.00	0.00*
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
$1(11 \leq C_{sg}^N < 12)$	0.02**	0.02**	-0.02**	-0.00	-0.00	-0.01	-0.01	-0.01*	0.00	-0.01**	0.01	0.00
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
$1(12 \leq C_{sg}^N < 13)$	0.02	-0.01	-0.02**	-0.01*	-0.00	-0.01	-0.00	-0.00	-0.00	-0.00	0.01	-0.00
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
$1(13 \leq C_{sg}^N < 14)$	0.02**	0.00	-0.01	-0.00	-0.00	-0.01	-0.02***	-0.02***	-0.01**	-0.02***	-0.00	-0.00
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
$1(14 \leq C_{sg}^N < 15)$	0.01	0.00	-0.01	-0.00	0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
$1(15 \leq C_{sg}^N < 16)$	-0.00	-0.01	0.00	-0.00	-0.00	-0.01	-0.01	-0.00	-0.00	-0.01	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
$1(16 \leq C_{sg}^N < 17)$	-0.00	-0.01	0.00	0.00	0.00	-0.01	-0.01*	-0.01*	-0.01	-0.01*	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
$1(17 \leq C_{sg}^N < 18)$	0.00	-0.01	-0.00	-0.00	0.00	-0.00	-0.01*	-0.01*	-0.01*	-0.01*	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
$1(18 \leq C_{sg}^N < 19)$	-0.00	-0.01	-0.00	-0.00	0.00	0.00	-0.01	-0.01	-0.01	-0.01	0.00	-0.00
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
$1(19 \leq C_{sg}^N < 20)$	-0.01	-0.02**	0.00	0.00	0.00	-0.00	-0.01**	-0.01*	-0.01	-0.01	-0.00	-0.00
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
$1(20 \leq C_{sg}^N < 21)$	-0.01	-0.01	0.01	0.00	0.00	-0.01	-0.01**	-0.01**	-0.01	-0.01**	-0.00	-0.00
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
$1(21 \leq C_{sg}^N < 22)$	-0.01*	-0.02**	0.01	-0.00	0.00	-0.00	-0.01**	-0.01**	-0.01*	-0.01***	-0.00	-0.00
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
$1(22 \leq C_{sg}^N < 23)$	-0.00	-0.01	0.01	-0.00	-0.00	-0.00	-0.01	-0.01	-0.01*	-0.01	-0.00	-0.00
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
$1(23 \leq C_{sg}^N < 24)$	-0.02**	-0.03***	0.00	-0.00	0.00	-0.01	-0.01*	-0.01*	-0.01*	-0.01*	-0.00	-0.00
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
$1(24 \leq C_{sg}^N < 25)$	-0.00	-0.01	0.00	-0.00	0.00	0.00	-0.02**	-0.02***	-0.01**	-0.02***	0.00	-0.00
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
Observations							16,100					
Joint significance of instruments (p-value)	0.0002	0.0029	0.0250	0.3137	0.4441	0.6061	0.1464	0.0957	0.2791	0.0879	0.5625	0.3145

Notes: The table reports in each column the estimates of a system of equations (Seemingly Unrelated Regression Estimates) with one equation for each control variable included in the OLS (Table 4) and IV (Table 2) regression of the paper. The unit of observation is a school. The dependent variable in each column is the within-group (institution-grade) residual of the covariate indicated in the heading of each column, i.e. the observed value of the variable in the school minus the institution-grade average of the same variable. The controls include the following set of within-group (institution-grade) residuals of family and individual covariates: the share of natives with mothers and fathers who attended, at most, a lower secondary school, the shares of natives with employed mothers and fathers, the share of natives who attended kindergarten and the share of male natives in class as well as the share of native students who report missing values in each of these variables. The instruments are a set of 15 dummies, one for each level of the theoretical number of natives in a class predicted by equation (2) according to the rules of class formation as a function of native enrollment at the school×grade level. The omitted category corresponds to a number of natives in a class equal to 25. There are no school-grades in which the number of natives in a class is less than 10. All regressions include a 2nd order polynomial of natives enrollment at the school×grade level. Robust standard errors clustered at the institution-grade level are reported in parentheses. A * denotes significance at 10%; a ** denotes significance at 5%; a *** denotes significance at 1%. The table also report the p-value for the test of joint significance of the instruments in all equation in the system and equation by equation. The Online Appendix contains an analogous table for the math sample.

Table 10: IV-FE estimates of the effect of the number of quasi-natives and first generation immigrants on the language test score of natives.

	Pooled 2nd & 5th		2nd grade		5th grade	
	Two	One	Two	One	Two	One
	endogenous		endogenous		endogenous	
	(1)	(2)	(3)	(4)	(5)	(6)
Number of quasi-natives: $\hat{\beta}$	-0.0017***	-0.0018***	-0.0018**	-0.0020***	-0.0015***	-0.0013**
	(0.0005)	(0.0004)	(0.0008)	(0.0007)	(0.0005)	(0.0005)
Number of immigrants (1st generation): $\hat{\gamma}$	-0.0285**	-0.0213**	-0.0328	-0.0176	-0.0090	-0.0163**
	(0.0143)	(0.0090)	(0.0285)	(0.0187)	(0.0108)	(0.0081)
Pure Ethnic Composition effect: $\hat{\delta}$	-0.0268*	-0.0195**	-0.0309	-0.0156	-0.0075	-0.0150*
	(0.0142)	(0.0086)	(0.0283)	(0.0181)	(0.0107)	(0.0077)
Observations	14,533	14,533	7,037	7,037	7,496	7,496
Institution×grade FE	✓	✓	✓	✓	✓	✓
Polynomial in natives enrollment	✓	✓	✓	✓	✓	✓
School level controls	✓	✓	✓	✓	✓	✓
Hansen (p-value)	0.9146	0.9036	0.9681	0.9570	0.5892	0.6439
F stat (natives)	305.1284		141.6049		166.5021	
SW F stat (natives)	61.5715		16.6815		75.1120	
SW χ^2 p-value (natives)	0.00		0.00		0.00	
F stat (immigrants)	2.2261	5.2617	1.0771	2.1167	2.1725	4.0935
SW F stat (immigrants)	2.2866	5.2617	1.0807	2.1167	2.2725	4.0935
SW χ^2 p-value (immigrants)	0.00	0.00	0.37	0.01	0.00	0.00

Notes: The table reports in each column a different set of estimates of equation (5) for the language samples. The unit of observation is a school. The dependent variable is the average test scores in language for natives students (i.e. fraction of correct answers). Quasi-natives are native students and second generation immigrants. The instruments are a set of 15 dummies, one for each level of the theoretical number of natives in a class predicted by equation (2) according to the rules of class formation as a function of native enrollment at the school×grade level. The omitted category corresponds to a number of natives in a class equal to 25. There are no school-grades in which the number of natives in a class is less than 10. All regressions include a 2nd order polynomial of natives enrollment at the school×grade level. The controls are school-level averages of the following set class-level covariates covariates: the share of natives with mothers and fathers who attended, at most, a lower secondary school, the shares of natives with employed mothers and fathers, the share of natives who attended kindergarten and the share of male natives in the class. All regressions include also the share of native students who report missing values in each of these variables as well as institution×grade fixed effects. Robust standard errors clustered at the institution-grade level are reported in parentheses. A * denotes significance at 10%; a ** denotes significance at 5%; a *** denotes significance at 1%. The table reports also: i) the p-value of the Hansen test; ii) the value of the F test of null hypothesis that the coefficients of the instruments are all zero in each first stage equation; and iii) the Sanderson-Windemeier first stage F statistic of each individual endogenous regressor (in the case of a model with one endogenous regressor this coincides with the F-test on excluded instruments) to test for weak identification ; and iv) the p-value of Sanderson-Windemeier χ^2 statistic of each individual endogenous regressor to test for under-identification (Sanderson and Windmeijer 2016).

Table 11: IV-FE estimates of the effect of the number of quasi-natives and first generation immigrants on the math test score of natives.

	Pooled 2nd & 5th		2nd grade		5th grade	
	Two	One	Two	One	Two	One
	endogenous		endogenous		endogenous	
	(1)	(2)	(3)	(4)	(5)	(6)
Number of quasi-natives: $\hat{\beta}$	-0.0015**	-0.0016***	-0.0014	-0.0015*	-0.0015**	-0.0013**
	(0.0005)	(0.0005)	(0.0009)	(0.0008)	(0.0007)	(0.0006)
Number of immigrants (1st generation): $\hat{\gamma}$	-0.0289*	-0.0223*	-0.0301	-0.0160	-0.0143	-0.0205**
	(0.0163)	(0.0103)	(0.0303)	(0.0204)	(0.0132)	(0.099)
Pure Ethnic Composition effect: $\hat{\delta}$	-0.0275*	-0.0207**	-0.0287	-0.0145	-0.0128	-0.0191**
	(0.0162)	(0.0099)	(0.0300)	(0.0198)	(0.0131)	(0.095)
Observations	14,524	14,524	7,030	7,030	7,494	7,494
Institution×grade FE	✓	✓	✓	✓	✓	✓
Polynomial in natives enrollment	✓	✓	✓	✓	✓	✓
School level controls	✓	✓	✓	✓	✓	✓
Hansen (p-value)	0.6681	0.6942	0.9245	0.9316	0.9113	0.9350
F stat (natives)	307.0714		144.0655		165.5012	
SW F stat (natives)	55.5993		12.2310		84.9578	
SW χ^2 p-value (natives)	0.00		0.00		0.00	
F stat (immigrants)	2.1894	5.2005	1.0697	2.0345	2.2018	4.1820
SW F stat (immigrants)	2.2437	5.2005	1.0555	2.0345	2.3121	4.1820
SW χ^2 p-value (immigrants)	0.00	0.00	0.39	0.01	0.00	0.00

Notes: The table reports in each column a different set of estimates of equation (5) for the language samples. The unit of observation is a school. The dependent variable is the average test scores in mathematics for natives students (i.e. fraction of correct answers). Quasi-natives are native students and second generation immigrants. The instruments are a set of 15 dummies, one for each level of the theoretical number of natives in a class predicted by equation (2) according to the rules of class formation as a function of native enrollment at the school×grade level. The omitted category corresponds to a number of natives in a class equal to 25. There are no school-grades in which the number of natives in a class is less than 10. All regressions include a 2nd order polynomial of natives enrollment at the school×grade level. The controls are school-level averages of the following set class-level covariates covariates: the share of natives with mothers and fathers who attended, at most, a lower secondary school, the shares of natives with employed mothers and fathers, the share of natives who attended kindergarten and the share of male natives in the class. All regressions include also the share of native students who report missing values in each of these variables as well as institution×grade fixed effects. Robust standard errors clustered at the institution-grade level are reported in parentheses. A * denotes significance at 10%; a ** denotes significance at 5%; a *** denotes significance at 1%. The table reports also: i) the p-value of the Hansen test; ii) the value of the F test of null hypothesis that the coefficients of the instruments are all zero in each first stage equation; and iii) the Sanderson-Windemeier first stage F statistic of each individual endogenous regressor (in the case of a model with one endogenous regressor this coincides with the F-test on excluded instruments) to test for weak identification ; and iv) the p-value of Sanderson-Windemeier χ^2 statistic of each individual endogenous regressor to test for under-identification (Sanderson and Windmeijer 2016).

Table 12: IV-FE estimates of the effect of the number of natives and immigrants on language and math test scores of natives, in institutions in which score manipulation is unlikely to have occurred according to Angrist et al. (2017a)

	Language		Mathematics	
	Two	One	Two	One
	endogenous		endogenous	
	(1)	(2)	(3)	(4)
Number of natives: $\hat{\beta}$	-0.0013***	-0.0012***	-0.0011**	-0.0011**
	(0.0005)	(0.0004)	(0.0005)	(0.0005)
Number of immigrants: $\hat{\gamma}$	-0.0152**	-0.0086***	-0.0151*	-0.0101***
	(0.0072)	(0.0026)	(0.0077)	(0.0029)
Pure Ethnic Composition effect: $\hat{\delta}$	-0.0139**	-0.0074***	-0.0139*	-0.0090***
	(0.0070)	(0.0023)	(0.0075)	(0.0025)
Observations	14,880	14,880	14,798	14,798
Institution×grade FE	✓	✓	✓	✓
Polynomial in natives enrollment	✓	✓	✓	✓
School level controls	✓	✓	✓	✓
Hansen (p-value)	0.8323	0.7670	0.5475	0.5584
F stat (natives)	352.8103		359.9060	
SW F stat (natives)	22.9880		25.0160	
SW χ^2 p-value (natives)	0.00		0.00	
F stat (immigrants)	2.8168	17.4742	3.0897	17.7891
SW F stat (immigrants)	2.7093	17.4742	2.9888	17.7891
SW χ^2 p-value (immigrants)	0.00	0.00	0.00	0.00

Notes: The table reports in each column a different set of estimates of equation (5) for the language and mathematics sub-samples in which we restrict the attention to schools in which for all classes the cheating indicator computed by Angrist et al. (2017a) signals no cheating (institutions where we do not have at least two schools that meet this condition are also dropped). The unit of observation is a school. The dependent variable is the average test scores in language [col. (1) and (2)] and mathmatics [col. (4) and (5)] for natives students (i.e. fraction of correct answers). The instruments are a set of 15 dummies, one for each level of the theoretical number of natives in a class predicted by equation (2) according to the rules of class formation as a function of native enrollment at the school×grade level. The omitted category corresponds to a number of natives in a class equal to 25. There are no school-grades in which the number of natives in a class is less than 10. All regressions include a 2nd order polynomial of natives enrollment at the school×grade level. The controls are aggregated at the school level and include the following set of family and individual covariates: the share of natives with mothers and fathers who attended, at most, a lower secondary school, the shares of natives with employed mothers and fathers, the share of natives who attended kindergarten and the share of male natives in the school. All regressions include also the share of native students who report missing values in each of these variables as well as institution×grade fixed effects. Robust standard errors clustered at the institution-grade level are reported in parentheses. A * denotes significance at 10%; a ** denotes significance at 5%; a *** denotes significance at 1%. The table reports also: i) the p-value of the Hansen test; ii) the value of the F test of null hypothesis that the coefficients of the instruments are all zero in each first stage equation; and iii) the Sanderson-Windemeier first stage F statistic of each individual endogenous regressor (in the case of a model with one endogenous regressor this coincides with the F-test on excluded instruments) to test for weak identification ; and iv) the p-value of Sanderson-Windemeier χ^2 statistic of each individual endogenous regressor to test for under-identification (Sanderson and Windmeijer 2016).