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

April 19, 2016

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 the 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. We show that the PEC effect is sizeable and negative (-1.6% for language and math) and does not vanish when children grow between age 7 and 10. Conventional estimates are instead smaller because they are confounded by the endogenous adjustments, in terms of quantity and quality of students, implemented by principals when confronted with an immigrant inflow.

JEL Classification: C36, I20, I24, J15
Keywords: Education, Immigration, Integration.

*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 the editor and referees, as well as 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 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. There exists a literature going beyond anecdotes, that uses different kinds of exogenous sources of variation to identify and estimate the causal effect of an inflow of immigrants in a classroom, but this literature does not distinguish between class composition and class size effects.

This paper aims at clarifying the importance of this distinction and proposes that a relevant parameter for policy should be the causal effect on school performance of a change in the ethnic composition of a class due to an immigrant inflow, net of two possible confounding components: first, the endogenous adjustments of class size and students' quality generated by principals' reactions to this inflow and, second, the mechanical class size change, for given number of natives, associated to this inflow even in the absence of reactions by principals. We call it the Pure Ethnic Composition (PEC) effect.

The first of these two confounding components originates from principals' expectations that non-native children are on average more problematic and that smaller classes or classes with less problematic children (of any ethnicity) help the learning process. Independently of whether these expectations are correct, principals will react to an inflow of immigrants by reducing the dimension of classes in their schools and by sorting students across classes on the basis of their quality, incurring in possibly significant costs for the necessary educational inputs and organisational adjustments. In this context, class size and class composition are the results of joint decisions of the principal, constrained by the available budget. For example, suppose that a smaller class improves student performance and that the effects of immigrants is zero but principals think it is negative. In this situation, they would reduce class size, when immigrants are expected, wasting valuable resources. At the same time, the econometrician would estimate a positive effect of an immigrants inflow when in fact it is zero, as the induced class size reduction is the only responsible of the positive estimated effect. A similarly distorted conclusion would be reached if the principal reacts to an immigrants inflow by increasing the quality of natives in the class. For a correct design of educational policies dealing with immigrants inflows, it is therefore necessary to estimate the effects of these inflows controlling for principals' reactions.

To this end, 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 rule" of class formation like the one used by Angrist and Lavy (1999) to estimate the causal effect of class size in Israel. The analogous rule prevailing in the Boston Area requires principals to cap class size at 25 and to increase the number of classes whenever the enrolment of non-Metco students goes above the 25 thresholds or its multiples.

In the Italian context, students should pre-enrol 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, additional splitting of classes occurs in September if late enrolment requires any further adjustment. Principals are also in charge of the distribution of natives and immigrants across the schools that they manage. While natives are typically allocated to the school where they enrol and more classes are formed in that school if the cap at 25 students per class is binding, for non-native students the procedure is less straightforward. Principals are explicitly instructed by the Ministry of Education to put immigrants in the schools of their jurisdiction where, because of how pre-enrolment has taken place, classes are smaller and there is more space. The explicit assumption is that this allocation would help reducing the disruption that immigrants may cause in large classes.

At the end of the entire procedure, as a result of the interaction between these instructions, early/late enrolment and the 25 cap on class size, the average number of immigrants per class in the different schools administered by a principal is a hump-shaped function of the average number of natives. This happens because a fraction of classes that in the end have few natives originates from the splitting of classes expected to be large in February and in which principals do not plan to put any immigrant to reduce disruption. Therefore, in classes that are finally small, the number of immigrants is a weighted average of the immigrants per class allocated to the originally small classes and of the zero (or very few) immigrants allocated to the classes that become small just because of splitting induced by September enrollment. At the opposite side of the spectrum, classes that initially are expected to be large in the number of natives and remain large, have no or few immigrants as well to avoid disruption, while the highest number of immigrants remains allocated to classes with an intermediate number of natives. Moreover, when final class splitting occurs in September, the principal has no time left to adjust the quality of immigrants and natives across the schools of her jurisdiction. Thus, final class splitting generates changes in the numbers of immigrants and natives across classes that are orthogonal to the quality of the two types of students in each class.

Thanks to this institutional setting we are therefore able to compare classes that have different numbers of natives and immigrants, for given children's quality, only because of the interaction between rules of class formation and early/late enrolment. This comparison is not confounded by the endogenous reactions of principals and offers what we need to identify and estimate the causal effects on natives test scores of one additional immigrant keeping natives constant (in quantity and quality) and of one additional native keeping immigrants constant (in quantity and quality). The final step to obtain the PEC effect consists in taking the difference between these two causal effects in order to remove the mechanical class size consequences associated with adding immigrants or natives to a class, while keeping constant the other component.

We find that the Pure Ethnic Composition effect on native performance is negative and statistically significant (-1.6% for language and math respectively, in our preferred specification) and does not vanish when children grow between age 7 and 10.[1] When we use instead a more conventional identification strategy that exploits within-school variation, as

[1]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.

for example in Contini (2011) and Ohinata and Van Ours (2011), the estimated effect of an immigrant inflow on native performance is smaller because it incorporates the class size and quality composition adjustments implemented by principals who fear the disruption caused by such inflow. Note that this conventional approach is no less valid or less interesting than ours: it just aims at a different parameter. Our claim is that the PEC effect is at least as relevant for policy decisions concerning how to handle immigrant inflows as what the conventional literature identifies.

Moreover, we conjecture that the failure to control for the endogenous reactions of principals affecting class size and class composition might be responsible for the large variability of estimates, in terms of sign and size, that one can find in the existing literature. For example, Hoxby (2000), Bossavie (2011), Tonello (forthcoming) exploit the variability in ethnic composition between adjacent cohorts within the same schools under the assumption that the variability between subsequent cohorts is random when the data are aggregated at the school-cohort level. Results based on this approach suggest 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 C. (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 (2012), 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.

Independently of how convincing their identification strategies are, these papers do not aim for the PEC effect and are not interested in controlling for the possibility that class size and class composition are endogenously adjusted by principals as a reaction to an immigrant inflow. What they aim for is the overall effect of such inflow, inclusive of all its indirect consequences, among which endogenous re-equilibrating changes in class size and quality. Such an overall effect is certainly an interesting parameter, but, if changes in class size and quality offset the consequences of immigrant inflows, the variability of these estimates should not come as a surprise.

The paper is organised as follows. We start in Section 2 with a discussion of the problems posed by the identification of the PEC effect. Then, after a description of our data in Section 3, we describe, in Section 4, conventional results that exploit variation within each institution and grade for identification. In Section 5 we present an alternative approach that allows us to identify and estimate the PEC effect. Results are presented in Section 6, where we also show that they are robust to the possibility that test scores are manipulated in some Italian regions. Section 7 concludes.

2 Problems in the estimation of the PEC effect

Consider an educational institution, with classes denoted by j, that is managed by a principal who is interested in improving students' performance V_j. N_j and I_j are, respectively, the numbers of natives and immigrants allocated to class j. Students of the two ethnicities may have a different quality, denoted by Q_j^n and Q_j^i in each class. Note, therefore, that within each ethnic group, students' quality may be heterogeneous across classes. Abstracting for simplicity from observable controls, we want to estimate the following linear approximation of the educational production function:

$$V_j = \alpha + \beta N_j + \gamma I_j + \lambda Q_j^n + \mu Q_j^i + \epsilon_j \tag{1}$$

were ϵ_j is noise. The parameter β (γ) measures the erosion of students' 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.

Denoting total class size with $C_j = N_j + I_j$, the effect of a pure ethnic composition change that we want to identify is:

$$\delta = \left(\frac{dV_j}{dI}\right)_{C_j=\bar{C};Q_j^n=\bar{Q}^n;Q_j^i=\bar{Q}^i} = \gamma - \beta \tag{2}$$

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.

The problem in estimating equation (1) is that the principal has some control on the number of immigrants and natives in each class as well as on their quality. Specifically, she may allocate students of better quality in larger classes and students of worse quality in smaller classes. And if immigrants are expected to have low quality, she may put more immigrants together with the best natives and fewer immigrants together with the worse. This setting would not be problematic if the econometrician could observe not only N_j and I_j but also Q_j^n and Q_j^i. When the quality of natives and immigrants is instead not observed by the econometrician, it is immediate to see that OLS estimation of equation (1) cannot deliver a consistent estimate of the PEC effect. Denoting with $\hat{\beta}_{ols}$ and $\hat{\gamma}_{ols}$ the OLS estimands of the population regression of V_j on N_j and I_j, they can be written as:

$$\begin{aligned}\hat{\beta}_{ols} &= \beta + \lambda\pi_{Q^nN} + \mu\pi_{Q^iN} \qquad (3)\\ \hat{\gamma}_{ols} &= \gamma + \lambda\pi_{Q^nI} + \mu\pi_{Q^iI}\end{aligned}$$

where π_{Q^hN} and π_{Q^hI}, for $h \in \{n, i\}$, denote the parameters of the projection of Q^h on N and I. Since all these parameters are likely to be positive, we expect the OLS estimates of β and γ to be upward inconsistent. As a consequence also the estimation of the PEC effect will be inconsistent

$$\hat{\delta}_{ols} = \hat{\gamma}_{ols} - \hat{\beta}_{ols} \neq \delta \tag{4}$$

although the expected direction of the inconsistency is not obvious.

As an introduction to the Instrumental Variable identification strategies that we describe and implement in Sections 5 and 6 to obtain a consistent estimate of the PEC effect, here we discuss the implications of two hypothetical controlled experiments that those IV strategies approximate. The first experiment assumes that we are interested only in the effect of an immigrant inflow. Given this goal, the presence of natives in equation (1) can be considered as a nuisance that does not create problems for the estimation of γ and we can limit ourselves to randomly assign immigrants to each class. Note that in this hypothetical experiment the principal maintains the possibility to take decisions concerning natives as well as immigrants that are not randomly assigned, and therefore there is some degree of non-compliance with the assignment scheme. The actual number of immigrants in a class can be decomposed in the randomly assigned component and in a residual which captures the effect of non-exogenous determinants of the allocation of immigrants to classes:

$$I_j = \hat{I}_j + r^i_j \tag{5}$$

Note that also the quality of immigrants can be decomposed in a similar way as a result of the random assignment:

$$Q^i_j = \hat{Q}^i_j + p^i_j \tag{6}$$

where $\hat{Q}^i_j$, like $\hat{I}_j$, is orthogonal all other variables. Substituting (5) and (6) in (1), the estimated equation is:

$$V_j = \alpha + \beta N_j + \gamma \hat{I}_j + \gamma r^i_j + \lambda Q^n_j + \mu \hat{Q}^i_j + \mu p^i_j + \epsilon_j \tag{7}$$

and the corresponding OLS estimands are

$$\begin{aligned} \hat{\beta}^1_{2sls} &= \beta + \gamma \sigma_{r^i_j N} + \lambda \sigma_{Q^n N} + \mu \sigma_{p^i_j N} \\ \hat{\gamma}^1_{2sls} &= \gamma \end{aligned} \tag{8}$$

where $\sigma_{r^i_j N}$, $\sigma_{Q^n N}$ and $\sigma_{p^i_j N}$ denote the parameters of the projections of r^i_j on Q^n and p^i_j on N, respectively.

The random assignment of immigrants to classes ensures that γ is estimated consistently because $\hat{I}_j$ is orthogonal to the unobservable components $r^i_j, Q^n_j, \hat{Q}^i_j, p^i_j$ and ϵ_j. β is instead estimated inconsistently because in this experiment the principal maintains the possibility of adjusting the number of natives N_j and their quality Q_j^n, to the non random component of the numbers of immigrants r_j^i and to the non-random component of the quality of immigrants p_j^i. We can expect $\sigma_{r_j^i N}$ to be negative if principals prefer to allocate fewer natives where the number of immigrants is larger (or viceversa) because they assume immigrants to be on average more disruptive. On the other hand, $\sigma_{Q^n N}$ and $\sigma_{p_j^i N}$ are likely to be positive if principals feel that they can allocate more natives in classes in which the quality of both types of students is better (or viceversa). The opposite signs of these effects, suggest that the direction and size of the inconsistency in not a priori determined since the different components can at least in part cancel out reciprocally. This strategy is fine if we are interested only in γ, but may be problematic if we are interested in the PEC effect, for which consistent estimates of both parameters are needed:

$$\hat{\delta}^1_{2sls} = \hat{\gamma}^1_{2sls} - \hat{\beta}^1_{2sls} \neq \delta \tag{9}$$

Consider instead an experiment in which some natives and some immigrants are randomly allocated to classes when principals have no time left to change their previous decisions concerning the non-random components of the two types of students. Therefore, also the number of natives

$$N_j = \hat{N}_j + r_j^n \tag{10}$$

and their quality

$$Q_j^n = \hat{Q}_j^n + p_j^n \tag{11}$$

can be decomposed in a random part that is orthogonal to all other unobservables and in a residual that captures the decisions of the principal. Substituting (5), (6), (10) and (11) in (1), the estimated equation is:

$$V_j = \alpha + \beta\hat{N}_j + \gamma\hat{I}_j + \beta r_j^n + \gamma r_j^i + \lambda\hat{Q}_j^n + \lambda p_j^n + \mu\hat{Q}_j^i + \mu p_j^i + \epsilon_j \tag{12}$$

In this case both β and γ are estimated consistently, because $\hat{N}_j$ and $\hat{I}_j$ are both orthogonal to all unobservable components

$$\begin{aligned} \hat{\beta}_{2sls} &= \beta \\ \hat{\gamma}_{2sls} &= \gamma, \end{aligned} \tag{13}$$

and their difference delivers the PEC effect

$$\hat{\delta}_{2sls} = \hat{\gamma}_{2sls} - \hat{\beta}_{2sls} = \delta. \tag{14}$$

The Instrumental Variable strategy that we describe in Section 5 approximates, with observational data, the above hypothetical controlled experiments.

3 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.[2]

We aggregate the data at the level of a class in a school since the regressors of main interest (class size and class composition) are defined at the class level. The outcomes on which we focus are the average fractions of correct answers in language and math of natives who are not absent on the day of the test in each class. Following international classification criteria (see PISA, 2009), INVALSI considers as natives those students who are born in Italy from Italian parents. Viceversa, students born from non-Italian parents are classified as immigrants regardless of whether they are born in Italy or not. The numbers of natives and immigrants officially enrolled in each class at the beginning of a school year (key variables in our analysis) are not contained in the standard files distributed by INVALSI, but were kindly provided to us in a separate additional file.

[2]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.

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 "language and "math" outcomes are based on slightly different datasets. Descriptive statistics for the first of these two samples are displayed in Table 1, while those for the second, which are very similar, can be found in Table A–1 of the Appendix.

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 class level to construct the control variables that we include in our specifications, together with the share of native students in a class for whom each of these variables is missing.

Starting from the universe of the 17,040 Italian schools, we operate the following sample restrictions. First we drop outliers by restricting the analysis to schools in which more than 10 and less than 160 students are enrolled in the 2nd or in the 5th grade[3]: this leaves us with 15,398 schools. We then drop the relatively few "stand-alone" schools that are not grouped together with other schools in educational institutions managed by a single principal; as anticipated in the Introduction and further explained below in Section 5, our identification strategy cannot apply to "stand-alone" schools. Of the remaining 12,405 schools, we drop also the 430 schools belonging to institutions in which no immigrant applies to any school for a given grade. Finally, we retain all the institutions in which we have at least two classes in the same grade (possibly in different schools) with no missing data on the variables required for the analysis. In the end, our sample is constituted by 12,859 second grade classes and 13,084 fifth grade classes belonging (respectively for the two grades) to 7,387 and 7,496 schools.[4] These schools belong, again respectively for the two grades, to 2,734 and 2,776 institutions.

The average enrolment of natives per school-grade is 30.41 while for immigrants it is 3.7. As expected, immigrants tend to perform worse than natives in reading and math, but the gap between ethnic groups is more sizeable in language. Natives perform relatively

[3]Angrist et al. (2014) operate a similar restriction (see page 7 in their paper).

[4]Note that schools with both 2nd and 5th grade classes are counted in both these groups and this explains why the sum of 7,387 and 7,496 is larger than the total number of schools in our data. Similarly for the analogous numbers concerning institutions in the next sentence.

better in Italian than in math and unsurprisingly the opposite happens for immigrants. The gap between natives and immigrants in reading tends to narrow across grades but remains relatively more stable in math. Finally, the dispersion in the score distribution for both Italian and math 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 reduce native performance.

4 Conventional evidence

Allowing for different schools, institutions and grades, as required by our data, and focusing on native performance, the specification of equation (1) that we estimate is:[5]

$$V^N_{jskg} = \alpha + \beta N_{jskg} + \gamma I_{jskg} + \mu X_{jskg} + \eta_{kg} + u_{jskg} \tag{15}$$

where V^N_{jskg} is the (log) of the average test score of natives in class j belonging to school s of institution k in grade g; N_{jskg} (I_{jskg}) is the number of natives (immigrants) in class j; X_{jskg} is a set of predetermined control variables defined at the class level for natives only[6], while η_{kg} denotes institution×grade fixed effects. The residual term $u_{jskg} = Q^n_{jskg} + Q^i_{jskg} + \epsilon_{jskg}$ contains the unobservable qualities of the two types of students in class $jskg$, which may be correlated with their numbers. This specification exploits, for identification, the within-institution-grade variation across classes as, for example, in Contini (2011) and Ohinata and Van Ours (2011), which takes care of a large set of relevant confounders but fails to address the problem generated by the unobservability of Q^n_{jskg} and Q^i_{jskg}.

Using the data described in Table 1 and in Section 3, the estimates of β, γ and δ in equations (15) based on this source of variation are reported in Table 2 for the language and math test scores of natives, pooling together the 2nd and the 5th grades as well as separately by grade. The effect β of an additional native, for given number of immigrants, is estimated to be very small and positive but statistically significant only in 5th grade, where

[5]A similar equation could be defined for the performance of an average immigrant, but as explained in footnote 1, the data at our disposal have enough information to focus on the performance on natives only.

[6]Specifically, 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/or nursery) and the share of males in the class. All the specifications include also the shares of native students that report missing values in each of these variables.

it reaches the level of 0.09% and 0.12% respectively for Language and Math. The effect γ of an additional immigrant, for given number of natives, is instead estimated to be negative, larger in size and statistically different from zero, ranging between - 0.34% for Language in 5th grade and - 0.64% for the same subject in 2nd grade. The implied estimate of $\delta = \gamma - \beta$, is therefore negative, ranging between - 0.37% for Math in 2nd grade and -0.57% for Language in 2nd grade, and is always statistically different from zero.

These estimates are potential inconsistent for the reasons discussed in Section 2 with specific reference to equations (3) and (4). In the next section we turn to an identification strategy that should eliminate these inconsistencies, if they exist.

5 A strategy to identify the PEC effect

In the month of February of each year Italian parents are invited to pre-enrol their children, for the following academic year, in one of the schools near where they live.[7] On the basis of this pre-enrolment 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-enrol, 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. For example, the "Circolare ministeriale" Number 4, comma 10.2, of January 15, 2009, says (our translation): "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 enrolment of

[7]The official rules for class formation in Italy are contained the DL n. 331/1998 and the DPR n. 81/2009 of the Ministry of Education and Research.

foreign students must be handled in a controlled way so that their allocation across schools is less disruptive."

After the February pre-enrolment phase, later in September final enrolment in schools may change because of new arrivals, family mobility and other contingencies. According to the Ministry[8], in the school year 2013-2014 approximately 35000 students enrolled later than February, corresponding to approximately 6% of total enrolment.

To clarify how these events and procedures generate exogenous differential variations in the number of immigrants and natives per class, let's consider a simplified example. Suppose that in grade g of school s predicted average class size $\bar{C}^N_{sg}$ at the school level, based on February native pre-enrolment and rules of class formation, can take three equally likely values: $H > M > L = H/2$. The principal knows that if $\bar{C}^N_{sg} = H$ for a class in February, with probability π that class will be split in September because of late enrolment in the corresponding school, originating two small classes each one with (approximately) $L = H/2$ natives. In the other two cases, instead, there is no risk of splitting.

Each principal manages three otherwise similar schools with different predicted average class sizes and has to allocate a total of I immigrants who enrol in February or September. Let's also assume, again for simplicity, that each school has one class. Since there is a probability $1 - \pi$ (with $0 < \pi < 1$) that the class expected to be large in February, will remain large (late enrolment insufficient to cause splitting in September), the principal will not plan to put immigrants in that class to avoid possible disruption. In the other two classes, instead, predicted class size based on native enrolment is low enough that immigrants cause no disruption and can be randomly distributed. Therefore, the average number of immigrants in the three types of classes, as anticipated in February, is:

$$I_{sg} = \begin{cases} 0 & \text{if} \quad \bar{C}^N_{sg} = H \\ \frac{I}{2} & \text{if} \quad \bar{C}^N_{sg} = M \\ \frac{I}{2} & \text{if} \quad \bar{C}^N_{sg} = L \end{cases} \tag{16}$$

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

In September, however, the schools with the high predicted class size will split their class with probability π. Therefore, the allocation of immigrants based on the *final* number of natives per class C_{sg}^N, after late enrolment has occurred, is,

$$I_{sg} = \begin{cases} \approx 0 & \text{if} \quad C_{sg}^N \approx H \\ \frac{I}{2} & \text{if} \quad C_{sg}^N \approx M \\ \approx \frac{I}{2}\frac{1}{(1+2\pi)} & \text{if} \quad C_{sg}^N \approx L \end{cases} \tag{17}$$

where the size of the three types of classes is now approximately H, M or L because of late enrolment. The average number of immigrants per class remains approximately zero in high-sized classes and does not change in medium- sized classes, while in the remaining group it is an average of the $\frac{I}{2}$ immigrants allocated to each one of the originally small classes and of the 0 (or very few) immigrants in the classes that become small because of September splitting.[9]

As a result of this allocation mechanism, since $\frac{1}{1+2\pi} < 1$, 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 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.

This hump shape emerges clearly in our data, as shown in Figure 1 that plots the average number of natives per class (circles - left vertical axis) and of immigrants per class (squares - right vertical axis) for each level of theoretical class size based on native enrolment. The figure also plots fitted values of the two relationships (solid for immigrants and dashed for

[9] Suppose that there are R principals, and therefore R classes of each type in February (given that each principal manages one class of each type). The number of high-sized classes that split is πR and they originate $2\pi R$ small classes. Therefore after splitting, the number of small classes is $R + 2\pi R$. Each one of the R originally small classes have $I/2$ immigrants, while the new $2\pi R$ small classes have ≈ 0 immigrants. Thus the final average number of immigrants in small classes is given by

$$\approx \frac{R\frac{I}{2} + 2\pi R0}{R + 2\pi R} \approx \frac{I}{2}\frac{1}{(1+2\pi)} < \frac{I}{2}.$$

The number of intermediate size classes remains R in September, each one with $I/2$ immigrants. High-sized classes have ≈ 0 immigrants and their final number is $R(1-\pi)$.

natives). Theoretical class size is calculated as a function of final enrolment of natives N_{sg} in school s and grade g, using the following "Maimonides-type" rule:

$$C_{sg}^{N} = \frac{N_{sg}}{\text{Int}\left(\frac{N_{sg}-1}{25}\right) + 1} \tag{18}$$

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 panels of Figure 2. The dashed line plots the theoretical class size C_{sg}^{N} as a function of the final enrolment of natives in each school, while the dark dots describe how the actual number of natives per class changes as a function of their enrolment. The top left panel zooms on levels of native enrolment ranging between 10 and 75, showing that in this range natives are allocated to classes almost exactly according to what the rule prescribes. At higher levels of enrolment, which can be seen in the bottom left panel covering the entire range, the correspondence is less precise as it usually happens in this type of analysis. The light dots describe instead, in both panels, the total actual class size, including immigrants, as a function of native enrolment. If the vertical distance between the light and the dark dots in the left panels of Figure 2 where constant at all levels of native enrolments, 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_{sg}^{N}. The right panels of Figure 2 plot this vertical distance (the connected light dots, which represent the actual number of immigrants per class) as a function of native enrolment, suggesting that it is not constant. The same panels also plot 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}. 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, not only in the 10-75 range of native enrolment. Moreover, and this is more evidente in the 10-75 range, the used space for immigrants tends to be relatively higher than expected for intervals of native enrolment that generate medium size classes. This result is due to the interaction between early/late enrolment 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 native enrolment (or theoretical class size), offer an exogenous source of variation that may be exploited for identification. Approximating the hypothetical setting of the controlled experiments described in Section 2, we can therefore estimate effects of adding one native or one immigrant to a class on the performance of natives that are not confounded by the endogenous reactions of principals in terms of class size and quality adjustments. This is what we need to identify and estimate the PEC effect.

6 New evidence on the PEC effect

We use the sources of exogenous variation described in the previous section for the IV estimation of the following equation

$$V_{jskg} = \alpha + \beta N_{jskg} + \gamma I_{jskg} + \mu X_{jskg} + \eta_{kg} + f(N_{sg}) + u_{jskg}, \tag{19}$$

which differs from equation (15) because a polynomial in native enrolment at the school×grade level is included to control for the systematic and continuous components of the relationship between native enrolment and native performance.

As in Angrist and Lang (2004), the set of instruments is provided by the following indicator variables:

$$\Psi \in \{1(1 \leq C^N_{sg} < 2), 1(2 \leq C^N_{sg} < 3), ..., 1(24 \leq C^N_{sg} < 25\}, \tag{20}$$

These indicators are defined for each possibile level of the theoretical number of natives in a class, C^N_{sg}, predicted by equation (18) according to the rules of class formation as a function of native enrolment at the school×grade level.[10] 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 enrolment to the numbers of natives and immigrants in a class.

[10]Note that the number of natives in a class can potentially range between 1 and 25, but the minimum number is actually higher in some of the sub-samples that we use in our analysis. See the footnotes to Tables 5, 6, A–2, A–3, A–4.

In the first two columns of Table 3, the 2nd and 5th grades are pooled together and the outcome is the average language test score of natives. In column (1), both the numbers of natives N_{jskg} and of immigrants I_{jskg} are instrumented with the indicator variables Ψ. This IV strategy approximates, in our observational context, the setting of the controlled experiment discussed in Section 2 in which some natives and immigrants are randomly assigned to classes and principals have no control on the quantity and quality of the randomly allocated students. This is what the institutional framework described in Section 5 implies because, after the final splitting of some classes occurs in September, principals have no time and possibility left for further adjustments of the quantity and quality of natives and immigrants across classes of different schools. While in the correspondent column (1) of Table 2, based on the same pooled sample and for the same outcome, the OLS estimate of β is positive but not statistically significant, the IV estimate in column 1 of Table 3 is negative and precise: keeping constant the number and the quality of immigrants, one additional native of average quality reduces the language test score of natives by 0.19 percent.

Much larger in size than in column (1) of Table 2 and similarly negative is the IV 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 1.77 percent. This finding is particularly remarkable given that the corresponding estimate of Table 2 is as small as -0.49 percent. Comparing equations (3) and (13), we interpret these differences as due to the fact that the IV estimates of β and γ are not confounded by the endogenous class size and quality adjustments implemented by principals when confronted with immigrant and native inflows. 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 as class size constant, is -1.58 percent. This is δ: the PEC effect for the language test score of natives when the 2nd and 5th grades are pooled.

In column (2) of Table 3, only the number of immigrants I_{jskg} is instrumented with the indicator variables Ψ and, with this procedure, we approximate the controlled experiment in which only immigrants are assigned randomly to classes. As indicated by equation (8), the estimate of β may be inconsistent in this case. The similarity of the estimates of β in columns 1 and 2 of Table 3 suggests that principals do not have much leeway in implementing adjustments concerning natives and/or that the different sources of inconsistency cancel each others (i.e., $\gamma\sigma_{r_j^i N} + \lambda\sigma_{Q^n N} + \mu\sigma_{p_j^i N} \approx 0$, in equation (8)). The estimates of γ is slightly smaller in column 2 (-1.46% instead of -1.77%), 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 the 2nd and 5th grades are pooled is -1.25%.

The first two columns of Table 4, offer a very similar set of results for the math test score of natives. In this case both estimates of β are similar and negative, but not statistically different from zero, while γ is precisely estimated ranging between -1.22% and -1.68%. Therefore the PEC effect for math ranges between -1.09% and -1.58%.

In the remaining columns of both tables the two grades are analysed separately and results are qualitatively similar although occasionally the estimates are less precise. Interestingly, the absolute size and the statistical significance of the estimates of γ and δ are typically larger for the 5th grade than for the 2nd grade. Using our preferred specification in which both the numbers of natives and immigrants are instrumented, the PEC effect increases from -1.25% (not statistically different from zero) at age 7 to -1.71% at age 10 in the case of language and from -1.21% (not statistically different from zero) at age 7 to -1.69% at age 10 in the case of math. These estimates suggest that the negative effect of swapping a native with an immigrant while keeping the quality of the two types of students as well as class size constant, does not fade away with age at least during elementary education.

Table 3 and 4 report also the p-values of the Hansen J test of over-identifying restrictions, which suggest that we cannot reject the null. The observed value of the F-test on the joint significance of the instruments in the first stage regression is reported for all specifications and we never reject the null. We also report the Angrist-Pischke test for two endogenous in the relevant columns. 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.

It has been recently suggested by Angrist et al. (2014) that estimates of class size effects in Italy, based on rules of class formation, are heavily manipulated by teachers in the Southern regions of the country, 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 in the South, originating the appearance of more negative, but fictitious, effects of class size in that part of the country.

In the light of this evidence it is possible that our estimates of the effects of β and γ (and thus of their difference δ) just capture score manipulation in the South. 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 $\gamma < \beta$ (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 7 and 8, respectively for the language and math test scores, that our results are essentially unchanged when we restrict the analysis to different sub-samples in which, according to Angrist et al. (2014), score manipulation is likely to be minimal, if at all present. In each of these tables, the first three columns report estimates in which both the numbers of immigrants and natives are instrumented, while in the last three columns only the number of immigrants is treated as endogenous. Given the purpose of this exercise, however, here we focus only on the sample that pools together the 2nd and 5th grades.[11] In columns (1) and (4) of both Tables, only schools in the north and centre of the country are considered.[12] The estimates of the Pure Ethnic Composition effects are slightly smaller than in Tables 3 and 4 but still statistically significant.[13]

In columns (2) and (5) we further restrict the sample to classes in the north and centre in which, according to the "cheating indicator" proposed by Angrist et al. (2014)[14], cheating is less likely to have occurred. The estimates of the PEC effect continue to be large, negative and statistically significant.

[11] Separate results by grade are available from the authors.

[12] North and centre are defined according to the definition of ISTAT (the Italian central institute for statistics) and include the following Italian regions: Emilia-Romagna, Friuli Venezia Giulia, Lazio, Liguria, Lombardia, Marche, Piemonte, Toscana, Umbria, Valle d'Aosta, Veneto and Trentino Alto Adige.

[13] The first stage regressions for these regressions and for the remaining ones commented below, are reported in the Appendix Tables A–2 and A–3.

[14] This indicator 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. (2014) for more details. We thank these authors for having shared with us the information that they constructed.

Finally, in columns (3) and (6) of both tables Tables 7 and 8, we consider only schools belonging to northern and central institutions in which an external monitor was sent by INVALSI[15] and results remain essentially unchanged. We therefore conclude that our analysis of the effect of immigrant inflows has general validity and is largely unaffected, at least in the north, by the score manipulation problem highlighted by Angrist et al. (2014) in Southern Italian regions.

7 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 policy parameter is the causal effect of substituting one native with one immigrant in a class 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. We discuss the problems posed by the identification and estimation of this parameter and we show that it has been neglected in the existing literature. This is not to say that estimates of the overall effect of an immigrant inflow, inclusive of the endogenous reactions of principals, are not interesting. We just propose that estimates of the PEC effect are necessary as well if principals want to calibrate correctly their reactions to an immigrant inflow, avoiding waste of resources.

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 performance in both language and math by approximately 1.6% in our preferred specification and this effect does not vanish when children grow between age 7

[15] In these institutions external inspectors where randomly assigned to classrooms during the INVALSI test scores, as explained in Lucifora and Tonello (2015), with the following specific tasks: i) invigilate students during the test; ii) provide specific information on the test administration; iii) compute and send results and documentation to INVALSI within a couple of days.

and 10. This estimate is larger (in absolute terms) than the one obtained with conventional identification strategies previously exploited in the literature, precisely because these conventional estimates are confounded by the adjustments of class size and quality implemented by principals who fear the disruption caused by immigrant inflows.

References

Angrist, J., Battistin, E., Vuri, D., 2014. In A Small Moment: Class Size and Moral Hazard in the Mezzogiorno. Working Paper 20173, NBER.

Angrist, J., Lang, K., December 2004. Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program. American Economic Review 94 (5), 1613–1634.

Angrist, J., Lavy, V., May 1999. Using Maimonides Rule To Estimate The Effect of Class Size on Scholastic Achievement. The Quarterly Journal of Economics 114 (2), 533–575.

Bossavie, L., 2011. Does Immigration Affect the School Performance of Natives? Evidence From Microdata. Mimeo, EUI.

Brunello, G., Rocco, L., 2013. The Effect of Immigration on the School Performance of Natives: Cross Country Evidence using PISA Test Scores. Economics of Education Review 32, 234–246.

Contini, D., 2011. Immigrant Background Peer Effect in Italian Schools. Working paper, IRVAPP.

Geay C., M. T., 2013. Non-Native Speakers of English in the Classroom: What Are the Effects on Pupil Performance. Economic Journal 123 (August), F281–F307.

Gould, E. D., Lavy, V., Paserman, M., October 2009. Does Immigration Affect the Long-Term Educational Outcomes of Natives? Quasi-Experimental Evidence. Economic Journal 119 (540), 1243–1269.

Hoxby, 2000. Peer Effect in the Classroom: Learning from Gender and Race Variation. Working paper 7876, NBER.

Hunt, J., 2012. The Impact of Immigration on Educational Attainment of Natives. Working paper 18047, NBER.

Jensen, P., Rasmussen, A., 2011. The Effect of Immigrant Concentration in Schools on Native and Immigrant Children's Reading and Math Skils. Economics of Education Review, 1503–1515.

Lucifora, C., Tonello, M., 2015. Cheating and Social Interaction: Evidence from a Randomized Experiment in a National Evaluation Program. Journal of Economic Behavior and Organization 115, 45–66.

Ohinata, A., Van Ours, J., December 2011. How immigrant children affect the academic achievement of native duch children. Discussion Papers 6212, IZA.

PISA, 2009. Pisa 2006 Technical Report. Tech. rep., OECD.

Quintano, C., Castellano, R., Longobardi, S., 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 (2), 149–171.

Tonello, M., forthcoming. Peer effects of non native students on natives' educational outcomes: Mechanisms and evidence. Empirical Economics.

Table 1: Descriptive statistics for the language sample

	2nd grade		5th grade	
	Mean	S.D.	Mean	S.D.
	Panel A. Class characteristics			
Fraction of correct answers:				
language (natives)	0.67	0.11	0.71	0.08
mathematics (natives)	0.61	0.11	0.65	0.10
language (immigrants)	0.53	0.19	0.60	0.15
mathematics (immigrants)	0.54	0.16	0.58	0.16
Number of natives in class	16.37	4.09	16.65	4.02
Number of immigrants in class	3.08	2.21	3.02	2.17
Class size	19.44	3.84	19.67	3.89
Percentage (0-100) of natives in class with				
low educated father	0.42	0.20	0.44	0.21
low educated mother	0.32	0.19	0.37	0.20
employed father	0.97	0.07	0.96	0.07
employed mother	0.67	0.21	0.66	0.21
Percentage (0-100) of natives in class who				
attended nurseries	0.32	0.23	0.27	0.22
attended kindergarten	0.99	0.04	0.99	0.04
are male	0.51	0.13	0.51	0.13
Cheating propensity (Quintano et al., 2009)	0.03	0.11	0.03	0.13
Cheating indicator (Angrist et al., 2014)	0.02	0.16	0.03	0.17
Sample size (number of classes)	12,859		13,084	
	Panel B. School characteristics			
Enrolment (natives)	30.40	19.90	30.83	20.09
Enrolment (immigrants)	3.94	4.81	3.88	4.56
Enrolment	34.34	21.76	34.71	21.96
Average number of classes	1.74	0.92	1.75	0.92
Sample size (number of schools)	7,387		7,496	
	Panel C. Institution characteristics			
External monitored institutions	0.23	0.42	0.23	0.42
Average number of classes	4.70	1.85	4.71	1.84
Average number of schools	2.70	1.05	2.70	1.05
Sample size (number of institutions)	2,734		2,776	

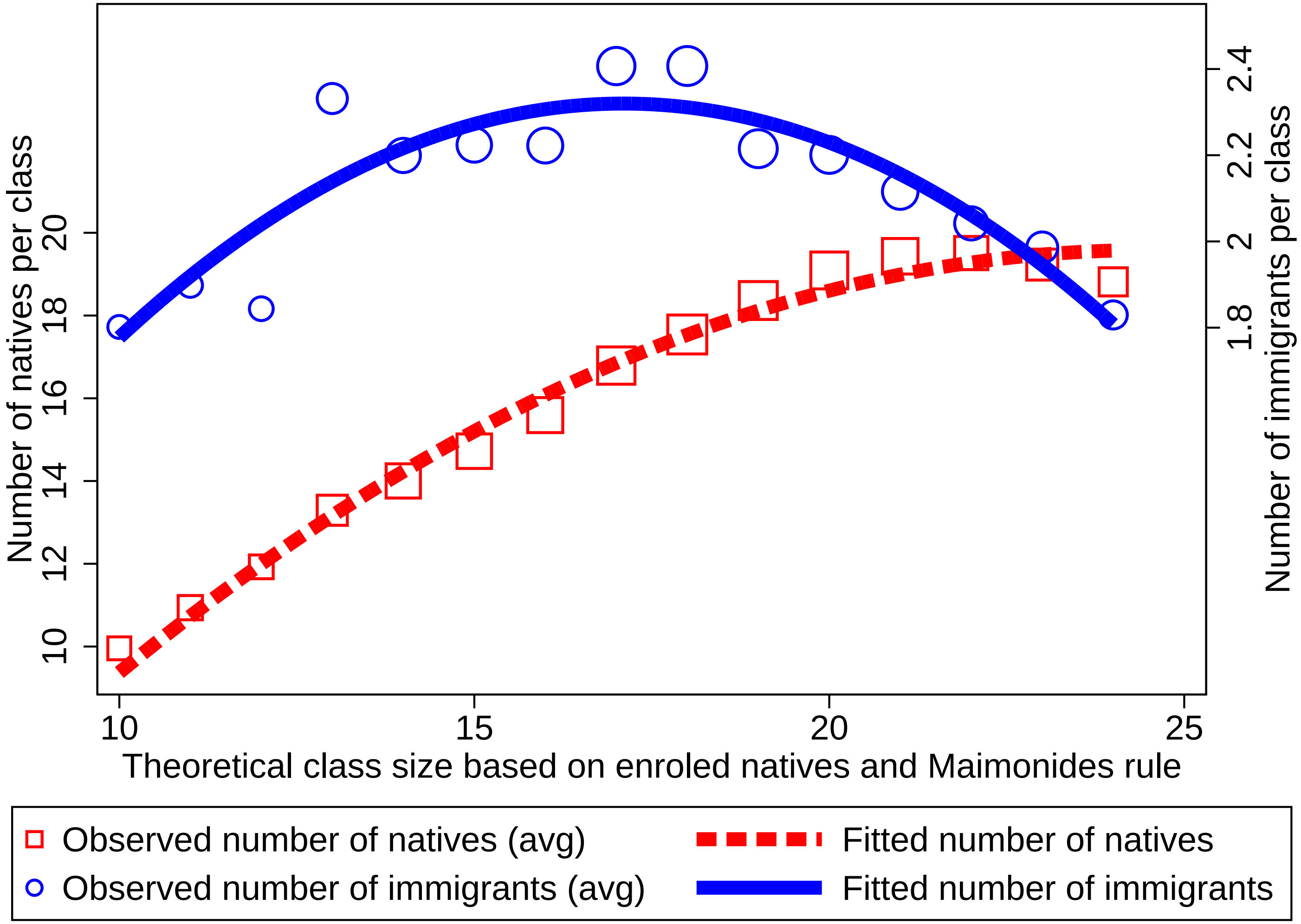
Notes: The unit of observation is a class in Panel A, a school in Panel B and an institution in Panel C. Institutions are groups of schools managed by the same principal. The family and individual background characteristics in Panel A are the shares of natives in a class who have that specific characteristic over the total number of natives in the class. Missing values do not contribute to the computation of these shares. All regressions in the following tables include the 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. (2014) and kindly given to us by these authors.

Table 2: OLS-FE estimates of the effect of the number of natives and immigrants on the test scores of natives; language and math samples.

	Language			Mathematics		
	Pooled	2nd grade	5th grade	Pooled	2nd grade	5th grade
	(1)	(2)	(3)	(4)	(5)	(6)
Number of natives: $\hat{\beta}$	0.0001	-0.0007	0.0009**	0.0002	-0.0009	0.0012**
	(0.0003)	(0.0005)	(0.0004)	(0.0004)	(0.0006)	(0.0005)
Number of immigrants: $\hat{\gamma}$	-0.0049***	-0.0064***	-0.0034***	-0.0042***	-0.0046***	-0.0039***
	(0.0006)	(0.0010)	(0.0007)	(0.0007)	(0.0011)	(0.0009)
Confounded ethnic composition effect: $\hat{\delta}$	-0.0050***	-0.0057***	-0.0043***	-0.0044***	-0.0037***	-0.0052***
	(0.0006)	(0.0009)	(0.0006)	(0.0006)	(0.0010)	(0.0008)
Observations	25,943	12,859	13,084	25,936	12,854	13,082
Institution×grade FE	✓	✓	✓	✓	✓	✓
Class level controls	✓	✓	✓	✓	✓	✓

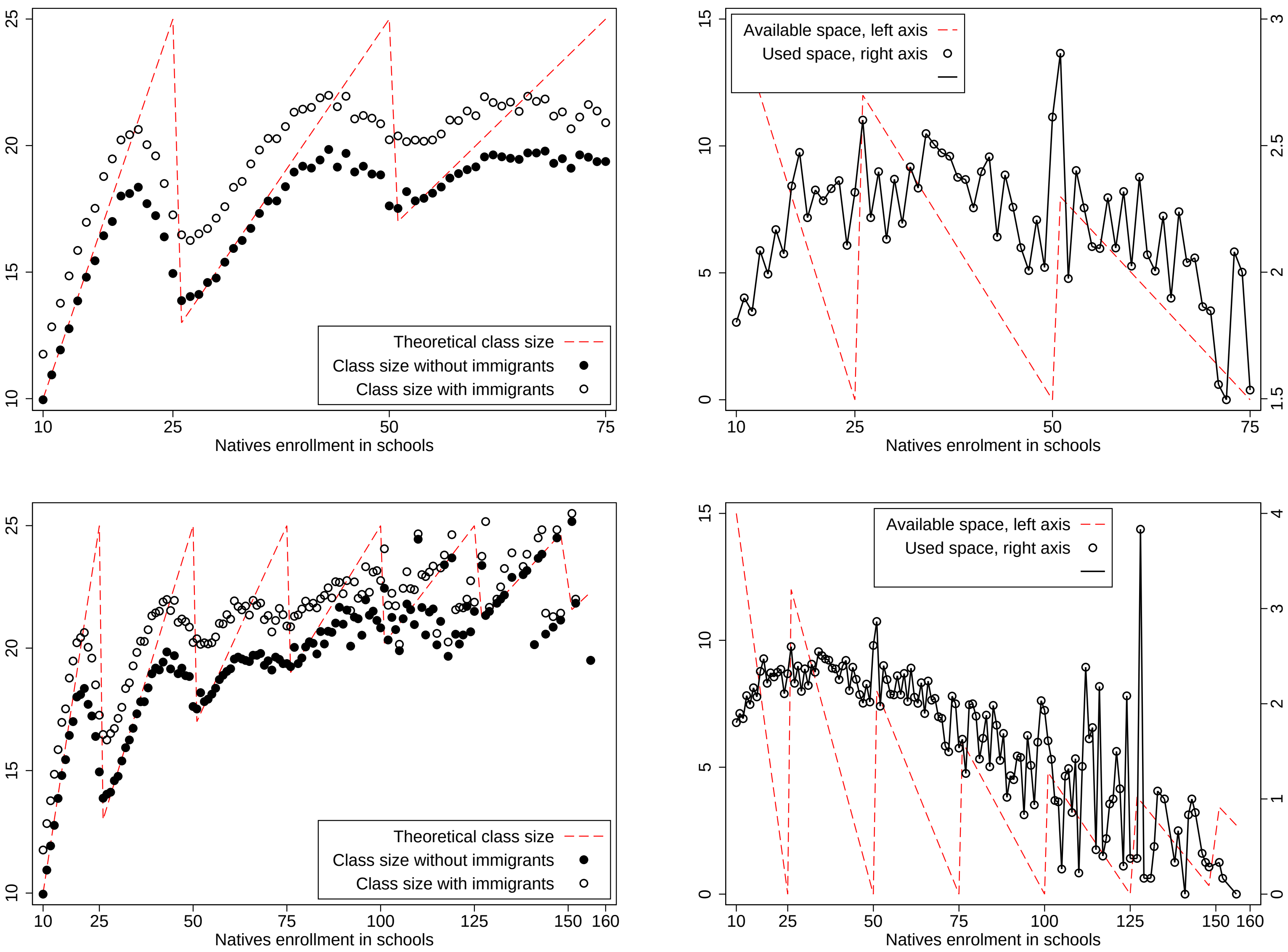
Notes: The table reports in each column a different regression based on the language and maths samples (described respectively in Tables 1 and A–1). The unit of observation is a class. The dependent variable is the log of the average test scores in language (math) for natives students. The controls are aggregated at the class 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/or nursery) 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. Standard errors are clustered at institution×grade level. A * denotes significance at 10%; a ** denotes significance at 5%; a *** denotes significance at 1%.

Figure 1: Number of natives and immigrants in a class as a function of theoretical class size based on native enrolment; language sample, 2nd and 5th grade.



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).

Figure 2: Number of natives and immigrants in a class and class size as a function of native enrolment; language sample, 2nd and 5th grades.



Notes: The left panels report the theoretical class size (dashed line), the class size without immigrants (dark dots) and the class size with immigrants (light dots) as a function of native enrollment in schools. In the right panels, the line connecing light dots represent the vertical distance between the light and dark dots of the left panels (the actual number of immigrants per class) as a function of native enrolment. The right panels also plot 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 top panels zoom on the first two splitting thresholds by restricting the horizontal axis to the 10-75 range, while the bottom panels use the entire 10-160 range.

Table 3: IV-FE estimates of the effect of the number of natives and immigrants on language native test scores; language sample.

	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**	-0.0021***	-0.0025**	-0.0026**	-0.0011	-0.0014
	(0.0007)	(0.0007)	(0.0012)	(0.0012)	(0.0009)	(0.0008)
Number of immigrants: $\hat{\gamma}$	-0.0177***	-0.0146***	-0.0150*	-0.0134***	-0.0182***	-0.0147***
	(0.0054)	(0.0030)	(0.0084)	(0.0047)	(0.0063)	(0.0037)
Pure Ethnic Composition effect: $\hat{\delta}$	-0.0158***	-0.0125***	-0.0125	-0.0108***	-0.0171***	-0.0133***
	(0.0052)	(0.0024)	(0.0080)	(0.0037)	(0.0061)	(0.0030)
Observations	25,943	25,934	12,859	12,859	13,084	13,084
Institution×grade FE	✓	✓	✓	✓	✓	✓
Polynomial in natives enrolment	✓	✓	✓	✓	✓	✓
Class level controls	✓	✓	✓	✓	✓	✓
Hansen (p-value)	0.7155	0.7038	0.7172	0.7621	0.7178	0.7312
F stat (natives)	382.5187		185.0871		231.9330	
AP stat (natives)	187.9826		86.9943		123.9195	
F stat (immigrants)	299.1926	89.7842	90.9724	41.9249	86.9107	38.0785
AP stat (immigrants)	88.0100	89.7842	41.7902	41.9249	37.1094	38.0785

Notes: The table reports in each column a different set of estimates of equation (19) for the language samples. The unit of observation is a class. The dependent variable is the log of the average test scores in language for natives students. The controls are aggregated at the class 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/or nursery) 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. Standard errors are clustered at institution×grade level. 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 Angrist-Pischke first stage 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).

Table 4: IV-FE estimates of the effect of the number of natives and immigrants on math native test scores; math sample.

	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: $\hat{\beta}$	-0.0010	-0.0013	-0.0013	-0.0016	-0.0005	-0.0007
	(0.0009)	(0.0009)	(0.0014)	(0.0014)	(0.0012)	(0.0011)
Number of immigrants: $\hat{\gamma}$	-0.0168***	-0.0122***	-0.0134	-0.0078	-0.0174**	-0.0150***
	(0.0056)	(0.0036)	(0.0083)	(0.0053)	(0.0069)	(0.0047)
Pure Ethnic Composition effect: $\hat{\delta}$	-0.0158***	-0.0109***	-0.0121	-0.0062	-0.0169**	-0.0143***
	(0.0053)	(0.0028)	(0.0078)	(0.0041)	(0.0066)	(0.0037)
Observations	25,936	25,936	12,854	12,854	13,082	13,082
Institution×grade FE	✓	✓	✓	✓	✓	✓
Polynomial in natives enrolment	✓	✓	✓	✓	✓	✓
Class level controls	✓	✓	✓	✓	✓	✓
Hansen (p-value)	0.9723	0.9565	0.7784	0.7974	0.8663	0.8840
F stat (natives)	221.7615		141.5156		194.8793	
AP stat (natives)	193.8379		88.6086		118.0768	
F stat (immigrants)	46.4591	35.7836	48.5942	28.1533	76.9678	42.3300
AP stat (immigrants)	40.8966	35.7836	35.3377	28.1533	41.3258	42.3300

Notes: The table reports in each column a different set of estimates of equation (19) for the math samples. The unit of observation is a class. The dependent variable is the log of the average test scores in math for natives students. The controls are aggregated at the class 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/or nursery) 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. Standard errors are clustered at institution×grade level. 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 Angrist-Pischke first stage 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).

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

	Pooled 2nd & 5th grades			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)
$1(3 \leq C^N_{sg} < 4)$	-11.30***	6.64***	3.63***	-12.14***	6.50***	3.14***	-10.34***	6.80***	4.15***
	(0.73)	(0.29)	(0.41)	(0.48)	(0.29)	(0.36)	(0.90)	(0.64)	(0.82)
$1(4 \leq C^N_{sg} < 5)$	-8.91***	3.77***	1.40***	-8.83***	3.65***	1.21***	-9.13***	4.03***	1.69***
	(0.30)	(0.13)	(0.14)	(0.42)	(0.22)	(0.22)	(0.43)	(0.19)	(0.19)
$1(5 \leq C^N_{sg} < 6)$	-9.48***	5.18***	2.61***	-8.52***	4.36***	1.93***	-10.52***	6.17***	3.46*
	(0.52)	(0.94)	(0.93)	(0.75)	(0.73)	(0.71)	(0.53)	(1.81)	(1.84)
$1(6 \leq C^N_{sg} < 7)$	-8.14***	3.83***	1.66**	-7.62***	3.61***	1.50*	-8.75***	4.04***	1.79
	(0.63)	(0.67)	(0.68)	(0.72)	(0.73)	(0.78)	(1.12)	(1.26)	(1.22)
$1(7 \leq C^N_{sg} < 8)$	-7.94***	2.20***	0.09	-7.46***	2.07***	0.00	-8.34***	2.37***	0.23
	(0.36)	(0.42)	(0.42)	(0.58)	(0.70)	(0.69)	(0.48)	(0.51)	(0.52)
$1(8 \leq C^N_{sg} < 9)$	-6.90***	1.23***	-0.61**	-6.74***	0.98***	-0.89***	-7.00***	1.52***	-0.28
	(0.32)	(0.23)	(0.24)	(0.42)	(0.30)	(0.31)	(0.49)	(0.38)	(0.40)
$1(9 \leq C^N_{sg} < 10)$	-6.14***	0.38*	-1.26***	-5.84***	0.10	-1.52***	-6.42***	0.64**	-1.01***
	(0.30)	(0.20)	(0.20)	(0.41)	(0.28)	(0.28)	(0.43)	(0.28)	(0.29)
$1(10 \leq C^N_{sg} < 11)$	-5.30***	-0.31**	-1.73***	-4.95***	-0.40*	-1.77***	-5.67***	-0.25	-1.70***
	(0.28)	(0.16)	(0.16)	(0.39)	(0.22)	(0.22)	(0.41)	(0.23)	(0.23)
$1(11 \leq C^N_{sg} < 12)$	-4.34***	-0.20	-1.35***	-4.06***	-0.34	-1.46***	-4.66***	-0.05	-1.24***
	(0.28)	(0.16)	(0.15)	(0.39)	(0.23)	(0.21)	(0.40)	(0.21)	(0.22)
$1(12 \leq C^N_{sg} < 13)$	-3.21***	-0.35**	-1.21***	-2.97***	-0.52**	-1.35***	-3.45***	-0.17	-1.06***
	(0.28)	(0.16)	(0.15)	(0.39)	(0.23)	(0.22)	(0.40)	(0.21)	(0.21)
$1(13 \leq C^N_{sg} < 14)$	-2.02***	-0.12	-0.66***	-1.77***	-0.24	-0.73***	-2.28***	0.00	-0.58***
	(0.27)	(0.13)	(0.13)	(0.38)	(0.20)	(0.18)	(0.39)	(0.18)	(0.18)
$1(14 \leq C^N_{sg} < 15)$	-1.49***	-0.18	-0.58***	-1.36***	-0.35*	-0.73***	-1.63***	-0.01	-0.43**
	(0.26)	(0.13)	(0.12)	(0.37)	(0.19)	(0.18)	(0.38)	(0.17)	(0.17)
$1(15 \leq C^N_{sg} < 16)$	-0.77***	-0.22*	-0.43***	-0.67*	-0.34*	-0.53***	-0.87**	-0.11	-0.34*
	(0.26)	(0.13)	(0.13)	(0.36)	(0.19)	(0.18)	(0.38)	(0.18)	(0.17)
$1(16 \leq C^N_{sg} < 17)$	-0.00	-0.24*	-0.24**	0.05	-0.46**	-0.45**	-0.05	-0.02	-0.04
	(0.26)	(0.13)	(0.12)	(0.36)	(0.19)	(0.17)	(0.38)	(0.17)	(0.17)
$1(17 \leq C^N_{sg} < 18)$	0.92***	-0.05	0.19	1.05***	-0.29	-0.00	0.78**	0.19	0.39**
	(0.26)	(0.12)	(0.12)	(0.36)	(0.18)	(0.16)	(0.37)	(0.17)	(0.16)
$1(18 \leq C^N_{sg} < 19)$	1.53***	-0.07	0.34***	1.60***	-0.30	0.14	1.45***	0.16	0.53***
	(0.26)	(0.13)	(0.12)	(0.36)	(0.19)	(0.17)	(0.38)	(0.17)	(0.17)
$1(19 \leq C^N_{sg} < 20)$	2.12***	-0.16	0.40***	2.25***	-0.34*	0.28*	2.01***	0.02	0.54***
	(0.26)	(0.12)	(0.12)	(0.37)	(0.18)	(0.17)	(0.38)	(0.16)	(0.16)
$1(20 \leq C^N_{sg} < 21)$	2.57***	-0.21*	0.47***	2.65***	-0.33*	0.41**	2.47***	-0.08	0.55***
	(0.27)	(0.12)	(0.12)	(0.37)	(0.18)	(0.17)	(0.38)	(0.17)	(0.16)
$1(21 \leq C^N_{sg} < 22)$	3.02***	-0.32**	0.48***	3.04***	-0.54***	0.30*	3.01***	-0.10	0.67***
	(0.28)	(0.13)	(0.12)	(0.39)	(0.19)	(0.17)	(0.39)	(0.17)	(0.17)
$1(22 \leq C^N_{sg} < 23)$	2.82***	-0.29**	0.46***	3.19***	-0.52***	0.36**	2.45***	-0.05	0.58***
	(0.29)	(0.13)	(0.12)	(0.40)	(0.19)	(0.17)	(0.42)	(0.18)	(0.18)
$1(23 \leq C^N_{sg} < 24)$	2.47***	-0.36***	0.29**	2.35***	-0.48**	0.17	2.59***	-0.25	0.42**
	(0.30)	(0.13)	(0.13)	(0.43)	(0.19)	(0.18)	(0.42)	(0.18)	(0.18)
$1(24 \leq C^N_{sg} < 25)$	1.91***	-0.41***	0.10	1.96***	-0.67***	-0.13	1.87***	-0.14	0.34*
	(0.30)	(0.14)	(0.13)	(0.42)	(0.20)	(0.18)	(0.43)	(0.18)	(0.18)
Institution×grade FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Polynomial in natives enrolment	✓	✓	✓	✓	✓	✓	✓	✓	
Class level controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	25,943	25,943	25,943	12,859	12,859	12,859	13,084	13,084	13,084
F stat	382.52	299.19	64.64	185.09	90.97	36.69	231.93	86.91	38.08
AP stat	187.98	88.01	64.64	86.99	41.79	36.69	123.92	37.11	38.08

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 class. The dependent variable is the number of natives (immigrants) in a class. The instruments are a set of 25 dummies, one for each level of the theoretical number of natives in a class predicted by equation (18) according to the rules of class formation as a function of native enrolment 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 3. All regressions include a 2nd order polynomial of natives enrolment at the school×grade level. The controls are aggregated at the class 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/or nursery) 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. Standard errors are clustered at the institution-grade level. 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 Angrist-Pischke first stage 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).

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

	Pooled 2nd & 5th grades			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)
$1(3 \leq C_{sg}^N < 4)$	-11.38***	6.71***	3.69***	-12.15***	6.55***	3.19***	-10.53***	6.93***	4.23***
	(0.73)	(0.28)	(0.41)	(0.47)	(0.29)	(0.36)	(0.98)	(0.60)	(0.81)
$1(4 \leq C_{sg}^N < 5)$	-9.13***	4.43***	2.00***	-9.43***	4.59***	1.98***	-8.43***	4.13***	1.96***
	(0.60)	(0.55)	(0.44)	(0.60)	(0.71)	(0.60)	(0.43)	(0.19)	(0.20)
$1(5 \leq C_{sg}^N < 6)$	-9.35***	5.15***	2.66***	-8.45***	4.40***	2.06***	-10.34***	6.09***	3.43*
	(0.48)	(0.95)	(0.94)	(0.68)	(0.75)	(0.72)	(0.51)	(1.83)	(1.84)
$1(6 \leq C_{sg}^N < 7)$	-8.20***	3.82***	1.64**	-7.87***	3.66***	1.48*	-8.57***	3.98***	1.78
	(0.63)	(0.67)	(0.69)	(0.74)	(0.73)	(0.80)	(1.10)	(1.27)	(1.23)
$1(7 \leq C_{sg}^N < 8)$	-7.93***	2.21***	0.10	-7.45***	2.09***	0.03	-8.33***	2.36***	0.23
	(0.37)	(0.42)	(0.42)	(0.60)	(0.70)	(0.69)	(0.49)	(0.51)	(0.52)
$1(8 \leq C_{sg}^N < 9)$	-6.89***	1.24***	-0.60**	-6.76***	1.00***	-0.87***	-6.98***	1.52***	-0.27
	(0.32)	(0.23)	(0.24)	(0.42)	(0.30)	(0.30)	(0.49)	(0.38)	(0.40)
$1(9 \leq C_{sg}^N < 10)$	-6.16***	0.39*	-1.25***	-5.85***	0.13	-1.49***	-6.45***	0.65**	-1.01***
	(0.30)	(0.20)	(0.20)	(0.41)	(0.28)	(0.28)	(0.43)	(0.28)	(0.29)
$1(10 \leq C_{sg}^N < 11)$	-5.32***	-0.29*	-1.71***	-5.00***	-0.35	-1.73***	-5.66***	-0.25	-1.70***
	(0.28)	(0.16)	(0.16)	(0.39)	(0.22)	(0.22)	(0.41)	(0.23)	(0.23)
$1(11 \leq C_{sg}^N < 12)$	-4.35***	-0.19	-1.35***	-4.09***	-0.31	-1.44***	-4.64***	-0.05	-1.25***
	(0.28)	(0.16)	(0.15)	(0.39)	(0.22)	(0.21)	(0.40)	(0.21)	(0.22)
$1(12 \leq C_{sg}^N < 13)$	-3.21***	-0.34**	-1.20***	-2.98***	-0.50**	-1.32***	-3.45***	-0.17	-1.06***
	(0.28)	(0.16)	(0.15)	(0.39)	(0.23)	(0.22)	(0.40)	(0.21)	(0.21)
$1(13 \leq C_{sg}^N < 14)$	-2.02***	-0.11	-0.64***	-1.78***	-0.22	-0.71***	-2.25***	0.01	-0.57***
	(0.27)	(0.13)	(0.13)	(0.38)	(0.20)	(0.18)	(0.39)	(0.18)	(0.18)
$1(14 \leq C_{sg}^N < 15)$	-1.50***	-0.17	-0.57***	-1.38***	-0.32*	-0.70***	-1.63***	-0.01	-0.43**
	(0.26)	(0.13)	(0.12)	(0.37)	(0.19)	(0.18)	(0.38)	(0.17)	(0.17)
$1(15 \leq C_{sg}^N < 16)$	-0.76***	-0.22*	-0.42***	-0.67*	-0.32*	-0.50***	-0.85**	-0.12	-0.34*
	(0.26)	(0.13)	(0.12)	(0.36)	(0.19)	(0.18)	(0.38)	(0.18)	(0.17)
$1(16 \leq C_{sg}^N < 17)$	-0.00	-0.23*	-0.23*	0.04	-0.44**	-0.43**	-0.05	-0.03	-0.04
	(0.26)	(0.13)	(0.12)	(0.36)	(0.19)	(0.17)	(0.38)	(0.17)	(0.17)
$1(17 \leq C_{sg}^N < 18)$	0.92***	-0.04	0.20*	1.05***	-0.28	0.01	0.79**	0.19	0.39**
	(0.26)	(0.12)	(0.12)	(0.36)	(0.18)	(0.16)	(0.38)	(0.17)	(0.17)
$1(18 \leq C_{sg}^N < 19)$	1.53***	-0.07	0.34***	1.59***	-0.28	0.16	1.46***	0.16	0.53***
	(0.26)	(0.13)	(0.12)	(0.36)	(0.18)	(0.17)	(0.38)	(0.17)	(0.17)
$1(19 \leq C_{sg}^N < 20)$	2.13***	-0.16	0.41***	2.25***	-0.33*	0.29*	2.02***	0.02	0.54***
	(0.26)	(0.12)	(0.12)	(0.36)	(0.18)	(0.16)	(0.38)	(0.16)	(0.16)
$1(20 \leq C_{sg}^N < 21)$	2.57***	-0.21*	0.47***	2.65***	-0.32*	0.41**	2.48***	-0.09	0.55***
	(0.27)	(0.12)	(0.12)	(0.37)	(0.18)	(0.17)	(0.38)	(0.17)	(0.17)
$1(21 \leq C_{sg}^N < 22)$	3.03***	-0.32**	0.49***	3.04***	-0.54***	0.30*	3.01***	-0.10	0.67***
	(0.28)	(0.13)	(0.12)	(0.39)	(0.19)	(0.17)	(0.39)	(0.17)	(0.17)
$1(22 \leq C_{sg}^N < 23)$	2.84***	-0.28**	0.47***	3.20***	-0.51***	0.38**	2.47***	-0.05	0.58***
	(0.29)	(0.13)	(0.12)	(0.40)	(0.19)	(0.17)	(0.42)	(0.18)	(0.18)
$1(23 \leq C_{sg}^N < 24)$	2.48***	-0.36***	0.30**	2.35***	-0.47**	0.18	2.60***	-0.24	0.43**
	(0.30)	(0.13)	(0.13)	(0.43)	(0.19)	(0.18)	(0.42)	(0.18)	(0.18)
$1(24 \leq C_{sg}^N < 25)$	1.92***	-0.40***	0.11	1.93***	-0.65***	-0.12	1.92***	-0.15	0.35*
	(0.30)	(0.14)	(0.13)	(0.42)	(0.20)	(0.18)	(0.43)	(0.19)	(0.18)
Institution×grade FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Polynomial in natives enrolment	✓	✓	✓	✓	✓	✓	✓	✓	
Class level controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	25,936	25,936	25,936	12,854	12,854	12,854	13,082	13,082	13,082
F stat	221.76	46.46	35.78	141.52	48.59	28.15	194.88	76.97	42.33
AP stat	193.84	40.90	35.78	88.61	35.34	28.15	118.08	41.33	42.33

Notes: The table reports in each column a different first stage regression correspondent to the IV estimates of Table 4. The unit of observation is a class. The dependent variable is the number of natives (immigrants) in a class. The instruments are a set of 25 dummies, one for each level of the theoretical number of natives in a class predicted by equation (18) according to the rules of class formation as a function of native enrolment 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 3. All regressions include a 2nd order polynomial of natives enrolment at the school×grade level. The controls are aggregated at the class 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/or nursery) 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. Standard errors are clustered at the institution-grade level. 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 Angrist-Pischke first stage 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).

Table 7: IV-FE estimates of the effect of the number of natives and immigrants on language native test scores; north & centre, language sample, pooled 2nd and 5th grades.

	Two Endogenous			One endogenous		
	Baseline specification	Classes with cheating indicator = 0	Externally monitored institutions	Baseline specification	Classes with cheating indicator = 0	Externally monitored institutions
	(1)	(2)	(3)	(4)	(5)	(6)
Number of natives: $\hat{\beta}$	-0.0026***	-0.0022***	-0.0033**	-0.0027***	-0.0024***	-0.0033**
	(0.0008)	(0.0008)	(0.0015)	(0.0008)	(0.0008)	(0.0014)
Number of immigrants: $\hat{\gamma}$	-0.0155***	-0.0172***	-0.0180**	-0.0144***	-0.0142***	-0.0170***
	(0.0053)	(0.0055)	(0.0075)	(0.0028)	(0.0028)	(0.0047)
Pure Ethnic Composition effect: $\hat{\delta}$	-0.0129**	-0.0149***	-0.0147**	-0.0117***	-0.0118***	-0.0137***
	(0.0051)	(0.0052)	(0.0072)	(0.0021)	(0.0021)	(0.0035)
Observations	19,001	18,636	4,730	19,001	18,636	4,730
Institution×grade FE	✓	✓	✓	✓	✓	✓
Polynomial in natives enrolment	✓	✓	✓	✓	✓	✓
Class level controls	✓	✓	✓	✓	✓	✓
Hansen (p-value)	0.3731	0.1972	0.4754	0.4196	0.2004	0.4754
F stat (natives)	243.6070	230.4025	92.5995	n.a.	n.a.	n.a.
AP stat (natives)	130.6553	125.4315	49.2872	n.a.	n.a.	n.a.
F stat (immigrants)	159.9033	156.4903	13.1686	57.3668	58.4058	11.8310
AP stat (immigrants)	58.3941	57.4936	8.9089	57.3668	58.4058	11.8310

Notes: The table reports in each column a different regression for schools in the north and centre of the country. The unit of observation is a class and the dependent variable is the log of the average test scores in language for natives students. Columns (1) and (4) report results for the baseline specification restricted to north & center. Columns (2) and (5) report results using only north & center classes where the cheating indicator computed by Angrist et al. (2014) signals no cheating (institutions where we do not have at least two classes that meet this condition are also dropped). Columns (3) and (6) display estimates for the sub-sample of north & center institutions in which INVALSI sent an external monitor. All regressions include a 2nd order polynomial of natives enrolment at the school×grade level. The controls are aggregated at the class 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/or nursery) 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. Standard errors are clustered at the institution×grade level. 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 Angrist-Pischke first stage 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).

Table 8: IV-FE estimates of the effect of the number of natives and immigrants on math native test scores; north & centre, math sample, pooled 2nd and 5th grades.

	Two Endogenous			One endogenous		
	Baseline specification	Classes with cheating indicator = 0	Externally monitored institutions	Baseline specification	Classes with cheating indicator = 0	Externally monitored institutions
	(1)	(2)	(3)	(4)	(5)	(6)
Number of natives: $\hat{\beta}$	-0.0008	-0.0010	-0.0019	-0.0010	-0.0012	-0.0020
	(0.0009)	(0.0009)	(0.0017)	(0.0009)	(0.0009)	(0.0016)
Number of immigrants: $\hat{\gamma}$	-0.0129**	-0.0155***	-0.0160*	-0.0096***	-0.0107***	-0.0150***
	(0.0055)	(0.0055)	(0.0084)	(0.0032)	(0.0032)	(0.0055)
Pure Ethnic Composition effect: $\hat{\delta}$	-0.0121**	-0.0145***	-0.0141*	-0.0086***	-0.0095***	-0.0130***
	(0.0052)	(0.0052)	(0.0079)	(0.0024)	(0.0024)	(0.0042)
Observations	19,005	18,697	4,733	19,005	18,697	4,733
Institution×grade FE	✓	✓	✓	✓	✓	✓
Polynomial in natives enrolment	✓	✓	✓	✓	✓	✓
Class level controls	✓	✓	✓	✓	✓	✓
Hansen (p-value)	0.8778	0.8316	0.6936	0.8834	0.7868	0.7368
F stat (natives)	165.0909	169.2722	103.9646	n.a.	n.a.	n.a.
AP stat (natives)	137.0163	135.5308	50.3710	n.a.	n.a.	n.a.
F stat (immigrants)	39.6502	34.4168	14.3539	32.3398	39.9397	12.4153
AP stat (immigrants)	33.2788	28.7020	9.5160	32.3398	39.9397	12.4153

Notes: The table reports in each column a different regression for schools in the north and centre of the country. The unit of observation is a class and the dependent variable is the log of the average test scores in math for natives students. Columns (1) and (4) report results for the baseline specification, restricted to north & center. Columns (2) and (5) report results using only north & center classes where the cheating indicator computed by Angrist et al. (2014) signals no cheating (institutions where we do not have at least two classes that meet this condition are also dropped). Columns (3) and (6) display estimates for the sub-sample of north & center institutions in which INVALSI sent an external monitor. All regressions include a 2nd order polynomial of natives enrolment at the school×grade level. The controls are aggregated at the class 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/or nursery) 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. Standard errors are clustered at the institution×grade level. 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 Angrist-Pischke first stage 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).

8 Appendix

Table A–1: Descriptive statistics for the math sample

	2nd grade		5th grade	
	Mean	S.D.	Mean	S.D.
	Panel A. Class characteristics			
Fraction of correct answers:				
language (natives)	0.67	0.11	0.71	0.08
mathematics (natives)	0.61	0.11	0.65	0.10
language (immigrants)	0.53	0.19	0.61	0.15
mathematics (immigrants)	0.54	0.16	0.58	0.15
Number of natives in class	16.37	4.09	16.65	4.02
Number of immigrants in class	3.07	2.21	3.02	2.17
Class size	19.44	3.84	19.67	3.89
Percentage (0-100) of natives in class with				
low educated father	0.42	0.20	0.44	0.21
low educated mother	0.32	0.19	0.37	0.20
employed father	0.97	0.07	0.96	0.07
employed mother	0.67	0.21	0.66	0.21
Percentage (0-100) of natives in class who				
attended nurseries	0.32	0.23	0.27	0.22
attended kindergarten	0.99	0.04	0.99	0.04
are male	0.51	0.13	0.51	0.13
Cheating propensity (Quintano et al., 2009)	0.03	0.12	0.03	0.12
Cheating indicator (Angrist et al., 2014)	0.03	0.16	0.03	0.18
Sample size (number of classes)	12,854		13,082	
	Panel B. School characteristics			
Enrolment (natives)	30.38	19.89	30.81	20.05
Enrolment (immigrants)	3.93	4.81	3.88	4.55
Enrolment	34.34	21.75	34.68	21.92
Average number of classes	1.74	0.92	1.74	0.92
Sample size (number of schools)	7,386		7,497	
	Panel C. Institution characteristics			
External monitored institutions	0.23	0.42	0.23	0.42
Average number of classes	4.70	1.85	4.72	1.84
Average number of schools	2.70	1.05	2.70	1.06
Sample size (number of institutions)	2,733		2,776	

Notes: The unit of observation is a class in Panel A, a school in Panel B and an institution in Panel C. Institutions are groups of schools managed by the same principal. The family and individual background characteristics in Panel A are the shares of natives in a class who have that specific characteristic over the total number of natives in the class. Missing values do not contribute to the computation of these shares. All regressions in the following tables include the 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. (2014) and kindly given to us by these authors.

Table A–2: First stage for the number of natives and immigrants (two endogenous variable model); north & centre, language sample, pooled 2nd and 5th grades.

	Baseline		Classes with cheating indicator = 0		Externally monitored institutions	
	Natives (1)	Immigrants (2)	Natives (3)	Immigrants (4)	Natives (5)	Immigrants (6)
$1(3 \leq C_{sg}^N < 4)$	-10.91***	6.53***	-10.94***	6.54***	-	-
	(0.75)	(0.32)	(0.75)	(0.32)	-	-
$1(4 \leq C_{sg}^N < 5)$	-8.61***	3.96***	-8.66***	3.96***	-10.03***	3.65***
	(0.35)	(0.18)	(0.35)	(0.18)	(0.66)	(0.45)
$1(5 \leq C_{sg}^N < 6)$	-8.84***	5.11***	-8.89***	5.31***	-10.67***	4.87**
	(0.63)	(0.96)	(0.68)	(1.04)	(0.95)	(2.40)
$1(6 \leq C_{sg}^N < 7)$	-7.52***	4.07***	-7.54***	4.09***	-9.49***	3.86***
	(0.68)	(0.76)	(0.70)	(0.80)	(0.63)	(1.05)
$1(7 \leq C_{sg}^N < 8)$	-7.57***	2.26***	-7.64***	2.24***	-8.46***	2.71***
	(0.40)	(0.45)	(0.40)	(0.46)	(0.72)	(0.88)
$1(8 \leq C_{sg}^N < 9)$	-6.57***	1.23***	-6.53***	1.17***	-7.53***	0.33
	(0.35)	(0.27)	(0.36)	(0.27)	(0.61)	(0.47)
$1(9 \leq C_{sg}^N < 10)$	-5.71***	0.39	-5.72***	0.41	-6.57***	0.94
	(0.33)	(0.25)	(0.34)	(0.25)	(0.60)	(0.59)
$1(10 \leq C_{sg}^N < 11)$	-4.86***	-0.35*	-4.87***	-0.35*	-5.79***	-0.51
	(0.32)	(0.20)	(0.32)	(0.20)	(0.57)	(0.41)
$1(11 \leq C_{sg}^N < 12)$	-3.96***	-0.22	-3.99***	-0.21	-4.27***	-0.71*
	(0.32)	(0.20)	(0.32)	(0.20)	(0.61)	(0.42)
$1(12 \leq C_{sg}^N < 13)$	-2.70***	-0.39**	-2.75***	-0.41**	-3.28***	-0.72**
	(0.31)	(0.20)	(0.31)	(0.20)	(0.59)	(0.36)
$1(13 \leq C_{sg}^N < 14)$	-1.72***	-0.16	-1.78***	-0.14	-2.36***	-0.38
	(0.30)	(0.17)	(0.30)	(0.17)	(0.57)	(0.35)
$1(14 \leq C_{sg}^N < 15)$	-1.19***	-0.24	-1.22***	-0.23	-1.81***	-0.48
	(0.29)	(0.16)	(0.29)	(0.16)	(0.55)	(0.36)
$1(15 \leq C_{sg}^N < 16)$	-0.33	-0.28*	-0.35	-0.31*	-1.38**	-0.61*
	(0.30)	(0.17)	(0.30)	(0.17)	(0.56)	(0.36)
$1(16 \leq C_{sg}^N < 17)$	0.40	-0.34**	0.36	-0.33**	-0.37	-0.21
	(0.29)	(0.16)	(0.29)	(0.16)	(0.55)	(0.34)
$1(17 \leq C_{sg}^N < 18)$	1.24***	-0.10	1.19***	-0.09	0.67	-0.25
	(0.29)	(0.16)	(0.29)	(0.16)	(0.53)	(0.33)
$1(18 \leq C_{sg}^N < 19)$	1.89***	-0.10	1.87***	-0.11	1.40***	-0.22
	(0.29)	(0.16)	(0.29)	(0.16)	(0.53)	(0.33)
$1(19 \leq C_{sg}^N < 20)$	2.49***	-0.25	2.45***	-0.25	1.33**	-0.25
	(0.29)	(0.16)	(0.29)	(0.16)	(0.55)	(0.33)
$1(20 \leq C_{sg}^N < 21)$	2.88***	-0.26	2.87***	-0.28*	1.91***	-0.34
	(0.30)	(0.16)	(0.30)	(0.16)	(0.56)	(0.32)
$1(21 \leq C_{sg}^N < 22)$	3.32***	-0.45***	3.24***	-0.41**	3.28***	-0.84**
	(0.31)	(0.16)	(0.31)	(0.16)	(0.58)	(0.34)
$1(22 \leq C_{sg}^N < 23)$	2.88***	-0.38**	2.83***	-0.37**	2.35***	-0.47
	(0.32)	(0.17)	(0.32)	(0.17)	(0.60)	(0.35)
$1(23 \leq C_{sg}^N < 24)$	2.59***	-0.52***	2.56***	-0.52***	1.32**	-0.55
	(0.35)	(0.17)	(0.35)	(0.17)	(0.62)	(0.34)
$1(24 \leq C_{sg}^N < 25)$	2.04***	-0.50***	2.04***	-0.51***	1.50**	-0.51
	(0.34)	(0.18)	(0.35)	(0.18)	(0.62)	(0.36)
Institution×grade FE	✓	✓	✓	✓	✓	✓
Polynomial in natives enrolment	✓	✓	✓	✓	✓	✓
Class level controls	✓	✓	✓	✓	✓	✓
Observations	19,001	19,001	18,636	18,636	4,730	4,730
F stat	219.17	102.76	208.71	102.04	92.58	13.17
AP stat	136.60	51.19	131.12	50.54	49.29	8.91

Notes: The table reports in each column a different first stage regression correspondent to the IV estimates in columns in columns (1), (2), (3) of Table 7. The unit of observation is a class. The dependent variable is the number of natives (immigrants) in a class. The instruments are a set of 25 dummies, one for each level of the theoretical number of natives in a class predicted by equation (18) according to the rules of class formation as a function of native enrolment at the school×grade level. The omitted category corresponds to a number of natives in a class equal to 25. There are no schools in which the number of natives in a class is equal to 1 in both grades and less than 4 in grade 5 The omitted category corresponds to a predicted class size equal to 25.There are no classes with only one native. Missing coefficients in one or more of the dummy variables indicate that there are no schools with the corresponding level of predicted class size. All regressions include a 2nd order polynomial of natives enrolment at the school×grade level. The controls are aggregated at the class 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/or nursery) 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. In columns (3) and (4) the analysis is restricted to classes in which the cheating indicator constructed by Angrist and Lang (2004) is equal to zero. In columns (5) and (6) it is instead restricted to classes in institutions in which INVALSI sent an external monitor. Standard errors are clustered at the institution×grade level. 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 Angrist-Pischke first stage 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).

Table A–3: First stage for the number of natives and immigrants (two endogenous variable model); north & centre, math sample, pooled 2nd and 5th grades.

	Baseline		Classes with cheating indicator = 0		Externally monitored institutions	
	Natives (1)	Immigrants (2)	Natives (3)	Immigrants (4)	Natives (5)	Immigrants (6)
$1(3 \leq C_{sg}^N < 4)$	-11.02***	6.61***	-11.80***	6.53***	-	-
	(0.75)	(0.31)	(0.51)	(0.33)	-	-
$1(4 \leq C_{sg}^N < 5)$	-8.79***	4.57***	-8.91***	4.62***	-9.97***	3.66***
	(0.55)	(0.55)	(0.57)	(0.56)	(0.66)	(0.44)
$1(5 \leq C_{sg}^N < 6)$	-9.01***	5.19***	-9.10***	5.21***	-10.98***	4.98**
	(0.49)	(0.97)	(0.49)	(0.98)	(1.06)	(2.50)
$1(6 \leq C_{sg}^N < 7)$	-7.63***	4.08***	-7.73***	4.09***	-9.65***	3.90***
	(0.69)	(0.77)	(0.70)	(0.77)	(0.61)	(1.03)
$1(7 \leq C_{sg}^N < 8)$	-7.57***	2.28***	-7.69***	2.37***	-8.56***	2.76***
	(0.40)	(0.45)	(0.41)	(0.45)	(0.73)	(0.89)
$1(8 \leq C_{sg}^N < 9)$	-6.57***	1.24***	-6.68***	1.24***	-7.54***	0.31
	(0.35)	(0.27)	(0.36)	(0.28)	(0.60)	(0.47)
$1(9 \leq C_{sg}^N < 10)$	-5.72***	0.42*	-5.82***	0.46*	-6.65***	1.00*
	(0.33)	(0.25)	(0.34)	(0.25)	(0.60)	(0.58)
$1(10 \leq C_{sg}^N < 11)$	-4.88***	-0.33	-5.00***	-0.31	-5.87***	-0.48
	(0.32)	(0.20)	(0.32)	(0.20)	(0.57)	(0.41)
$1(11 \leq C_{sg}^N < 12)$	-3.98***	-0.21	-4.07***	-0.17	-4.37***	-0.68*
	(0.32)	(0.20)	(0.32)	(0.20)	(0.60)	(0.41)
$1(12 \leq C_{sg}^N < 13)$	-2.71***	-0.37*	-2.83***	-0.39**	-3.37***	-0.68*
	(0.31)	(0.20)	(0.32)	(0.20)	(0.59)	(0.37)
$1(13 \leq C_{sg}^N < 14)$	-1.72***	-0.15	-1.82***	-0.11	-2.37***	-0.38
	(0.30)	(0.17)	(0.31)	(0.17)	(0.57)	(0.35)
$1(14 \leq C_{sg}^N < 15)$	-1.21***	-0.21	-1.30***	-0.19	-1.90***	-0.44
	(0.29)	(0.16)	(0.30)	(0.17)	(0.55)	(0.36)
$1(15 \leq C_{sg}^N < 16)$	-0.34	-0.27	-0.40	-0.27	-1.41**	-0.60*
	(0.30)	(0.17)	(0.30)	(0.17)	(0.56)	(0.36)
$1(16 \leq C_{sg}^N < 17)$	0.39	-0.32**	0.30	-0.28*	-0.43	-0.18
	(0.29)	(0.16)	(0.29)	(0.16)	(0.55)	(0.34)
$1(17 \leq C_{sg}^N < 18)$	1.23***	-0.09	1.15***	-0.04	0.64	-0.23
	(0.29)	(0.16)	(0.29)	(0.16)	(0.53)	(0.33)
$1(18 \leq C_{sg}^N < 19)$	1.89***	-0.10	1.82***	-0.09	1.35**	-0.20
	(0.29)	(0.16)	(0.30)	(0.16)	(0.53)	(0.33)
$1(19 \leq C_{sg}^N < 20)$	2.50***	-0.24	2.41***	-0.21	1.28**	-0.22
	(0.29)	(0.16)	(0.30)	(0.16)	(0.55)	(0.33)
$1(20 \leq C_{sg}^N < 21)$	2.87***	-0.26	2.81***	-0.24	1.85***	-0.31
	(0.30)	(0.16)	(0.30)	(0.16)	(0.56)	(0.32)
$1(21 \leq C_{sg}^N < 22)$	3.32***	-0.44***	3.24***	-0.41**	3.21***	-0.82**
	(0.31)	(0.16)	(0.32)	(0.16)	(0.58)	(0.34)
$1(22 \leq C_{sg}^N < 23)$	2.89***	-0.38**	2.78***	-0.36**	2.29***	-0.45
	(0.32)	(0.17)	(0.33)	(0.17)	(0.60)	(0.35)
$1(23 \leq C_{sg}^N < 24)$	2.60***	-0.52***	2.49***	-0.50***	1.29**	-0.54
	(0.35)	(0.17)	(0.35)	(0.17)	(0.62)	(0.34)
$1(24 \leq C_{sg}^N < 25)$	2.04***	-0.49***	1.97***	-0.47***	1.44**	-0.49
	(0.35)	(0.18)	(0.35)	(0.18)	(0.62)	(0.36)
Institution×grade FE	✓	✓	✓	✓	✓	✓
Polynomial in natives enrolment	✓	✓	✓	✓	✓	✓
Class level controls	✓	✓	✓	✓	✓	✓
Observations	19,005	19,005	18,697	18,697	4,733	4,733
F stat	165.08	39.65	169.26	34.42	103.94	14.35
AP stat	137.02	33.28	135.53	28.70	50.37	9.52

Notes: The table reports in each column a different first stage regression correspondent to the IV estimates in columns (1), (2), (3) of Table 8. The unit of observation is a class. The dependent variable is the number of natives (immigrants) in a class. The instruments are a set of 25 dummies, one for each level of the theoretical number of natives in a class predicted by equation (18) according to the rules of class formation as a function of native enrolment at the school×grade level. The omitted category corresponds to a number of natives in a class equal to 25. There are no schools in which the number of natives in a class is less than 3 in both grades and less than 4 in grade 5 All regressions include a 2nd order polynomial of natives enrolment at the school×grade level. The controls are aggregated at the class 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/or nursery) 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. In columns (3) and (4) the analysis is restricted to classes in which the cheating indicator constructed by Angrist and Lang (2004) is equal to zero. In columns (5) and (6) it is instead restricted to classes in institutions in which INVALSI sent an external monitor. Standard errors are clustered at the institution×grade level. 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 Angrist-Pischke first stage 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).

Table A–4: First stage for the number of immigrants (one endogenous variable model); north & centre, language or math samples, pooled 2nd and 5th grades

	Baseline		Classes with cheating indicator = 0		Externally monitored institutions	
	Lang. sample (1)	Math sample (2)	Lang. sample (3)	Math sample (4)	Lang. sample (5)	Math sample (6)
$1(3 \leq C_{sg}^N < 4)$	3.10***	3.15***	3.12***	2.82***	-	-
	(0.41)	(0.41)	(0.40)	(0.31)	-	-
$1(4 \leq C_{sg}^N < 5)$	1.25***	1.81***	1.25***	1.82***	0.43	0.40
	(0.19)	(0.44)	(0.18)	(0.45)	(0.41)	(0.42)
$1(5 \leq C_{sg}^N < 6)$	2.33**	2.36**	2.53**	2.34**	1.41	1.40
	(0.95)	(0.97)	(1.03)	(0.97)	(2.24)	(2.20)
$1(6 \leq C_{sg}^N < 7)$	1.71**	1.68**	1.73**	1.66**	0.77	0.78
	(0.75)	(0.77)	(0.78)	(0.77)	(1.03)	(1.00)
$1(7 \leq C_{sg}^N < 8)$	-0.12	-0.10	-0.15	-0.04	-0.01	-0.04
	(0.46)	(0.45)	(0.47)	(0.46)	(0.87)	(0.86)
$1(8 \leq C_{sg}^N < 9)$	-0.83***	-0.82***	-0.87***	-0.86***	-2.14***	-2.11***
	(0.28)	(0.28)	(0.29)	(0.29)	(0.48)	(0.48)
$1(9 \leq C_{sg}^N < 10)$	-1.40***	-1.38***	-1.38***	-1.37***	-1.16**	-1.19**
	(0.25)	(0.25)	(0.25)	(0.25)	(0.54)	(0.54)
$1(10 \leq C_{sg}^N < 11)$	-1.88***	-1.86***	-1.87***	-1.88***	-2.39***	-2.39***
	(0.20)	(0.20)	(0.21)	(0.21)	(0.38)	(0.38)
$1(11 \leq C_{sg}^N < 12)$	-1.47***	-1.46***	-1.46***	-1.45***	-2.10***	-2.09***
	(0.19)	(0.19)	(0.19)	(0.19)	(0.39)	(0.40)
$1(12 \leq C_{sg}^N < 13)$	-1.23***	-1.22***	-1.26***	-1.28***	-1.78***	-1.79***
	(0.19)	(0.19)	(0.19)	(0.19)	(0.32)	(0.32)
$1(13 \leq C_{sg}^N < 14)$	-0.70***	-0.68***	-0.70***	-0.69***	-1.15***	-1.15***
	(0.16)	(0.16)	(0.16)	(0.16)	(0.31)	(0.31)
$1(14 \leq C_{sg}^N < 15)$	-0.61***	-0.59***	-0.61***	-0.60***	-1.06***	-1.07***
	(0.16)	(0.16)	(0.16)	(0.16)	(0.31)	(0.32)
$1(15 \leq C_{sg}^N < 16)$	-0.39**	-0.37**	-0.42***	-0.39**	-1.06***	-1.06***
	(0.16)	(0.16)	(0.16)	(0.16)	(0.32)	(0.32)
$1(16 \leq C_{sg}^N < 17)$	-0.21	-0.20	-0.22	-0.19	-0.32	-0.33
	(0.15)	(0.15)	(0.15)	(0.15)	(0.29)	(0.29)
$1(17 \leq C_{sg}^N < 18)$	0.29**	0.30**	0.29*	0.32**	-0.02	-0.03
	(0.15)	(0.15)	(0.15)	(0.15)	(0.28)	(0.28)
$1(18 \leq C_{sg}^N < 19)$	0.49***	0.50***	0.47***	0.48***	0.24	0.24
	(0.15)	(0.15)	(0.15)	(0.15)	(0.28)	(0.28)
$1(19 \leq C_{sg}^N < 20)$	0.54***	0.54***	0.52***	0.54***	0.19	0.18
	(0.15)	(0.15)	(0.15)	(0.15)	(0.28)	(0.27)
$1(20 \leq C_{sg}^N < 21)$	0.65***	0.64***	0.62***	0.64***	0.29	0.28
	(0.15)	(0.15)	(0.15)	(0.15)	(0.27)	(0.27)
$1(21 \leq C_{sg}^N < 22)$	0.59***	0.60***	0.60***	0.61***	0.22	0.22
	(0.15)	(0.15)	(0.15)	(0.15)	(0.28)	(0.28)
$1(22 \leq C_{sg}^N < 23)$	0.52***	0.52***	0.51***	0.52***	0.29	0.29
	(0.16)	(0.16)	(0.16)	(0.16)	(0.29)	(0.29)
$1(23 \leq C_{sg}^N < 24)$	0.29*	0.30*	0.28*	0.29*	-0.12	-0.12
	(0.16)	(0.16)	(0.16)	(0.16)	(0.28)	(0.28)
$1(24 \leq C_{sg}^N < 25)$	0.14	0.15	0.13	0.15	-0.02	-0.03
	(0.17)	(0.17)	(0.17)	(0.17)	(0.30)	(0.30)
Institution×grade FE	✓	✓	✓	✓	✓	✓
Polynomial in natives enrolment	✓	✓	✓	✓	✓	✓
Class level controls	✓	✓	✓	✓	✓	✓
Observations	19,001	19,005	18,636	18,697	4,730	4,733
F stat	46.76	32.34	47.33	39.94	12.41	11.83

Notes: The table reports in each column a different first stage regression correspondent to the IV estimates in columns (4), (5), (6) of Table 7 and 8 . The unit of observation is a class. The dependent variable is the number of natives (immigrants) in a class. The instruments are a set of 25 dummies, one for each level of the theoretical number of natives in a class predicted by equation (18) according to the rules of class formation as a function of native enrolment at the school×grade level. The omitted category corresponds to a number of natives in a class equal to 25. There are no schools in which the number of natives in a class is less than 3 in both grades and less than 4 in grade 5 All regressions include a 2nd order polynomial of natives enrolment at the school×grade level. The controls are aggregated at the class 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/or nursery) 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. In columns (3) and (4) the analysis is restricted to classes in which the cheating indicator constructed by Angrist and Lang (2004) is equal to zero. In columns (5) and (6) it is instead restricted to classes in institutions in which INVALSI sent an external monitor. Standard errors are clustered at the institution×grade level. A * denotes significance at 10%; a ** denotes significance at 5%; a *** denotes significance at 1%. The table reports also the value of the F test of null hypothesis that the coefficients of the instruments are all zero in each first stage equation.