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

June 20, 2015

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 constant (Pure Composition Effect). We explain why this is a relevant policy parameter although it has been neglected so far. We show that the PCE is sizeable and negative at age 7 (-1.6% for language and math) and does not vanish when children grow up to age 10. Conventional estimates are instead smaller because they are confounded by endogenous class size adjustments implemented by principals when confronted with immigrant and native inflows.

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

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.

The first contribution of this paper is to clarify the importance of this distinction. A relevant parameter for policy is the causal effect on school performance of a change in class composition due to immigrant inflows, net of two possible confounding components: first, the endogenous class size changes generated by the reactions of principals to these inflows and, second, the mechanical class size changes associated to these inflows even in the absence of reactions by principals. We call it the Pure Composition Effect (PCE).

The first and most important of these two confounding components originates from principals' expectations that non-native children are more problematic and that smaller classes 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, incurring in possibly significant costs for the necessary additional educational inputs. This intuition can be formalised building on Lazear (2001) model of the educational production function in which the absolute numbers of students in a class determines class performance, if each student has some positive probability of generating disruption. We extend this model in two ways: by allowing for two types of students, natives and immigrants, each with its own probability of disruption and by introducing the possibility that the behaviour of a student generates positive externalities, not only disruption, across or within groups.

In this extended context, class size and class composition are the results of joint decisions of the principal, constrained by the available budget and by the cost of educational inputs. Thus, class composition effects estimated without controlling for endogenous class size changes could be confounded. 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

1

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 reliable assessment of the composition effect net of principals' reactions affecting class size is therefore a necessary ingredient for a correct design of educational policies dealing with immigrants inflows.

The second contribution of our paper is to fill in this gap by adapting 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). To this end, 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 rule of class formation like the "Maimonides rule" prevailing in Israel and used by Angrist and Lavy (1999) to estimate the causal effect of class size in that context. 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. The number of classes is tentatively decided by principals in February according to pre-enrolled students following a similar "Maimonides-type 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 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 immigrants 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 0 immigrants allocated to the classes that become small just because of splitting. At the opposite side of the spectrum, classes that initially are expected to be large in the number of natives and remain large, do not have immigrants as well to avoid disruption, while the highest number of immigrants remains allocated to classes with an intermediate number of natives.

Thanks to this institutional setting we are therefore able compare classes that have different numbers of natives and immigrant only because of the interaction between rules of class formation and early/late enrolment. The population parameters that we can identify are Local Average Treatment Effects because they refer to the effect of changes in the number of immigrants and natives that are induced in a class by small variations in the school enrolment of natives around the splitting threshold. 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 and of one additional native keeping immigrants constant. The final step to obtain the PCE 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 Composition Effect on native performance is negative and statistically significant at age 7 (-1.6% for both language and math) and does not vanish when children grow up to age 10.[1] When we use instead a more conventional identification strategy that exploits within-school variation, our estimates of the effects of immigrant inflows on native performance are smaller because they are confounded by the endogenous class size adjustments implemented by principals who fear the disruption caused by such inflows.

The paper is organised as follows. We first present, in Section 2, the extension of Lazear (2001) model which allows us to clarify the importance of distinguishing between class size and class composition effects. We then review the recent literature in Section 3 to show that it gives small and imprecise negative estimates of the effect of immigrant inflows possibly

---

[1] In principle, our identification strategy could allow us to identify and estimate the PCE on immigrant performance 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.

because it does not control for the consequences of class size adjustments implemented by principals as a reaction to those inflows. After a description of our data in Section 4, we show in Section 5 that we obtain similar results when we do not control for principals' reactions as well. Therefore, we move in Section 6 to discuss an alternative approach that allows us to identify and estimate the PCE. Results are presented in Section 7 where we also show that they are robust to the possibility that test scores are manipulated in some Italian regions. Section 8 concludes.

2 Why class composition and class size are not independent

To understand why the distinction between class size and class composition is important, we start from the model of the educational production function proposed by Lazear (2001). The central idea of Lazear's model is that the time devoted by teachers to students in a classroom is a public good. A "private" use of this time (i.e. a student that asks or requires specific attention) creates negative externalities that spill over the entire class, affecting the performance of all the other pupils. Consider a class with C students. If no student asks for specific attention, all students benefit fully and equally from the time of the teacher. Let $P \in [0,1]$ be the probability that a student *does not require* specific attention by the teacher at the expenses of others. Then, the likelihood that the event of disruption does not occur is a function of the number of students in the class: P^C. Define $\bar{V}$ as the potential performance of a student (measured, for example, by a test) if the teacher could devote full attention to her. The actual performance is a fraction of potential performance:

$$V = \bar{V} P^C \leq \bar{V} \tag{1}$$

The equation above shows that the performance of students is closely linked with the size of the class C. Even small deviations from $P = 1$ (i.e. even rare episodes of disruption) can generate large performance losses when class size is large. For example, if $P = 0.98$ in a class of 25 students, $V = 0.6\bar{V}$. Therefore, a 2% individual probability of demanding specific teacher's attention decreases by 40% the potential performance of an average student in the classroom.

4

We extend the above framework in two directions. First, we generalize the model allowing for the possibility of episodes in which the behaviour of a single student has a positive externality on classmates: for example a student may be asking interesting questions that allow the teacher to clarify points that are unclear to all students. Second, we adapt the framework to the possibility that students are of two types (natives and immigrants) and, thus, class composition, in terms of these two groups, matters for performance.

In order to introduce the first kind of extension, consider the following modified educational production function:

$$V = \bar{V} P^{\phi C} \begin{cases} \geq \bar{V} & \text{if} & \phi < 0 \\ \leq \bar{V} & \text{if} & \phi > 0 \end{cases} \tag{2}$$

where $\phi \in \mathbb{R} \backslash \{0\}$. A negative $\phi < 0$, captures the situation of a constructive behaviour of students, so that the effect of class size on performance is reversed and becomes positive. If $\phi > 0$ we are back in Lazear's world in which behaviour is "mis-behaviour" and class size can only reduce performance. Therefore, under this extension, the effect of class size may change depending on ϕ for given P. In this way we can accommodate the possibility that a larger number of natives might affect positively the performance of immigrants (or viceversa).

Before introducing the possibility of heterogeneous types of students, however, it is important for the purpose of this paper to see how the optimal class size that would be chosen by a principal is affected by the introduction of the parameter ϕ in Lazear's model. The principal solves the problem:

$$\underset{C}{\text{Max}} \, \Pi = P^{\phi C} - \frac{W}{C} \tag{3}$$

where W is the wage of a teacher and the rental cost of the capital she uses. The first order condition is:

$$\phi P^{\phi C} p + \frac{W}{C^2} = 0 \tag{4}$$

where $p = Ln(P)$, which implies that the optimal class size is:

$$C^* = f(\underset{+}{P}, \underset{+}{W}, \underset{-}{\phi}) \tag{5}$$

Therefore, optimal class size increases with the good behaviour of students (higher P) and with the cost of providing the educational public good (W). As ϕ increases (for example

because the quality and usefulness of student's questions in class decreases) the optimal class size is smaller.

It is crucial to note for our purposes, as shown by Lazear, that C and P are positively related at the optimum: class size is adapted by principals to the quality of their students given input costs. This relationship hides, in observational data, the negative causal effect of C on V keeping P constant. In other words, without an exogenous variation of class size C, independent of P, it is impossible to estimate the causal effect of class size on students' performance (and viceversa).

We now relax the hypothesis of student's homogeneity within a class. Assume that students are divided in two groups with different probabilities of disruption: N natives (with $P = P_n$ and $\bar{V} = \bar{V}_n$) and I immigrants (with $P = P_i$ and $\bar{V} = \bar{V}_i$). The test scores of the average native in a class is:

$$V_n = \bar{V}_n P_n^{\phi_{nn} N} P_i^{\phi_{in} I} \tag{6}$$

where the parameters $\phi_{qh} \in \mathbb{R} \backslash \{0\}$, with $q, h \in \{n, i\}$, capture the possibility that teachers' attention asked by a pupil affects differently the performance of natives according to the ethnicity of the pupil.[2]

Using small letters for logs the performance of a native student can be written as:

$$v_n = \bar{v}_n + p_n \phi_{nn} N + p_i \phi_{in} I \tag{7}$$

which implies that the effects of class size when the number of natives (immigrants) is increased *exogenously* keeping the number of immigrants (natives) constant are, respectively:

$$\beta = \frac{\partial v_n}{\partial N} = p_n \phi_{nn} \tag{8}$$

$$\gamma = \frac{\partial v_n}{\partial I} = p_i \phi_{in}$$

From these parameters, which by definition are not confounded by the endogenous reactions of principals to immigrant and native inflows, we can derive the effect of a pure composition change:

$$\delta = \left(\frac{dv_n}{dI} \right)_{C=\bar{C}} = \gamma - \beta \tag{9}$$

---

[2]A similar equation could be defined for the performance of an average immigrant, but (see footnote 1) since the empirical analysis is restricted to the performance of natives, it would be useless to discuss it here.

which is the PCE: the effect of increasing *exogenously* the number of immigrants keeping class size constant (i.e. reducing at the same time the number of natives).

For instance, typical wisdom posits that immigrants are more in need of specific attention ($p_i < p_n < 0$) and that the effects of attention requests are more damaging for natives when they originate from immigrants ($\phi_{in} \geq \phi_{nn} > 0$). In this case, our model would predict that $\gamma < \beta < 0$ and $\delta < 0$. In words, this configuration of parameters implies that the effect of an increase of class size on native performance is negative and stronger when it occurs because of an increase in the number of immigrants and that substituting one native with an immigrant, keeping class size constant, reduces native performance.

Other configurations of the structural parameters, different from the typical wisdom, are plausible as well and only the data can say which is the relevant one. However, the crucial lesson to be taken away from the above extension of Lazear (2001) model is that to estimate the Pure Composition Effects, and specifically the two causal effects whose difference gives the PCE, it is necessary to find an identification strategy that controls for the endogenous class size and class composition adjustments that principals may implement when confronted with exogenous inflows of immigrants students in their schools.

3 The literature and the pure composition effect

Different identification strategies have been explored in the literature and here we focus on the most recent contributions. It is important to note, however, that even if many of these contributions are certainly very convincing and lead to the estimation of useful population parameters, none of them aims at estimating the PCE.

Contini (2011) and Ohinata and Van Ours (2011), building on Ammermuller and Pischke (2009), address the problem of the endogenous sorting of immigrants between schools by exploiting the variability in the share of immigrant students within schools between classes of a given grade while Hoxby (2000), Bossavie (2011), Tonello (2012) exploit the variability in ethnic composition between adjacent cohorts within the same schools. The first approach rests on the assumption that, once school fixed effects are controlled for, the allocation of immigrants between classes is as good as random. These authors find a zero or weak negative effect of immigrants concentration on average native performance, but also show that this effect becomes larger for immigrants students and for students with low family

background. The second approach argues, perhaps more convincingly, 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.

Using detailed longitudinal data on Texas students, Hanushek et al. (2009) try to address the endogeneity in the exposure to black minorities controlling for an array of individual, school, grade and year fixed effects. They find small negative effects (not significant in some cases) of black concentration on white performance and a sizeable reduction for black schoolmates.

One of the most convincing studies is Gould et al. (2009) who use 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), who 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 math performance are instead 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. Card and Rothstein (2007) overcome the endogenous sorting of students between schools by aggregating at the city level the relationship between the black/white test score gap and the degree of segregation. City fixed effects, given two ethnicities (black and white) per city, takes care of sorting across cities. They find as well a negative effect of segregation at the school and neighbourhood level on the achievement gap, with the latter being stronger than the former. Along the same line, Brunello and Rocco (2011) aggregate the data at the country level and exploit the within country variation over time in the share of immigrants in a school, finding small negative effects. 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 PCE and are not interested in controlling for the possibility that class size is endogenously adjusted by principals as a reaction to changes in class composition. What they aim for is the overall effect of a change in composition, inclusive of all its indirect consequences, among which endogenous re-equilibrating changes in class size. Such an overall effect is certainly an interesting parameter, but, if changes in class size offset the consequences of immigrant inflows, it should not come as a surprise that these studies mostly report estimated "gross" effects of immigrant concentration on native performance that are negative but often small and not significant (actually in some cases even positive).

4 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.[3]

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.

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 the Appendix Table A-1.

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:

---

[3]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.

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.

Key variables in our analysis are the numbers of natives and immigrants officially enrolled in each class at the beginning of a school year. This information is not contained in the standard files distributed by INVALSI, but it was kindly provided to us in a separate additional file. The officially enrolled natives and immigrants in each class are identified according to citizenship, as explained above in this section.

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:[4] 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 6, 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 schools in which we have at least two classes in the same grade 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.[5] 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 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

---

[4](Angrist et al., 2014, p.7) operate a similar restriction

[5]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.

fact that immigrants test scores are lower on average, has motivated the public opinion concern that immigrant inflows reduce native performance.

5 Evidence that does not control for principals' reactions to immigrant inflows

Equation (7) describes the performance of a representative native student as a function of the number of natives and immigrants in her class. An empirical counterpart of this equation is

$$v_{jskg} = \alpha + \beta N_{jskg} + \gamma I_{jskg} + \mu X_{jskg} + \eta_{sg} + \epsilon_{jskg} \tag{10}$$

where v_{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_{sg} denotes school×grade fixed effects. The term ϵ_{jskg} captures other unobservable determinants of student performance.

Using the data described in Table 1 and in Section 4, we first present results that exploit the within-school-grade variation across classes for identification, as, for example, in Contini (2011) and Ohinata and Van Ours (2011). The estimates of β, γ and δ in equations (10) based on this source of variation are reported in Table 2 for the language and math test scores of natives, pooling together the second and the fifth grades as well as separately by grade.

Note that this identification strategy cannot control for the possibility that principals, as suggested by Lazear (2001), adjust class size when they observe inflows of natives and immigrants that change class composition. Specifically, we expect the estimates of both β and γ to be confounded by potentially different factors that depend on principals beliefs about the disruption caused by more natives or immigrants in a class. For this reason, the implied estimate of $\delta = \gamma - \beta$, which in the table appears to be about -0.5%, is inconsistent for the PCE. If immigrants are expected to disrupt more than natives, principals will react more to immigrants than to natives in terms of endogenous class size adjustments, and γ

---

[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 students that report missing values in each of these variables.

should be underestimated more than β. In this scenario the estimate of δ is lower in absolute value than the true PCE.

In the next section we turn to an identification strategy that addresses this problem, suggesting that when endogenous class size adjustments due to principals' reactions are controlled for, the estimate of γ, and thus of δ, is significantly more negative.

6 An alternative identification strategy for the PCE

In the month of February of each year 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 school networks 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 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 slightly 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.

---

[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.

[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.

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 $\bar{C}_{sg}^N$ 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 $\bar{C}_{sg}^N = H$ 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 (no late enrolment), 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:

$$
I_{sg} = \begin{cases} 0 & \text{if} & \bar{C}_{sg}^N = H \\ \frac{I}{2} & \text{if} & \bar{C}_{sg}^N = M \\ \frac{I}{2} & \text{if} & \bar{C}_{sg}^N = L \end{cases} \tag{11}
$$

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,

$$
I_{sg} = \begin{cases} 0 & \text{if} & C_{sg}^N \approx H \\ \frac{I}{2} & \text{if} & C_{sg}^N \approx M \\ \frac{I}{2} \frac{1}{(1+2\pi)} & \text{if} & C_{sg}^N \approx L \end{cases} \tag{12}
$$

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 does not change in the high- and 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 immigrants in the classes that become small because of 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 final average number of natives per class. 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 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 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:

$$C_{sg}^{N} = \frac{N_{sg}}{\text{Int}\left(\frac{N_{sg}-1}{25}\right) + 1} \tag{13}$$

Figure 1 shows that the average number of natives is an increasing function of theoretical class size. This because natives are allocated to classes almost exactly according to a "Maimonides type" rule. Immigrants, instead, are under-represented in small and large classes because of the decision process and development of events described in the above discrete example.

This evidence suggests that the interaction between early/late enrolment and rules of class formation on the allocations of immigrants across schools, generates different evolutions in the number of natives and immigrants as a function of native enrolment around the

---

[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 big 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

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

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

splitting threshold. These arguably exogenous small variations in native enrolment allow us to compare classes that have different numbers of natives and immigrant only because of the interaction between rules of class formation and early/late enrolment. In the next section we exploit the non-linearities of these patterns to produce estimates of the effects of natives and immigrants in a class on the performance of natives, that are not confounded by the endogenous reactions of principals in terms of class size adjustments. This is what we need to identify and estimate the PCE.

7 New evidence

We apply the identification strategy described in the previous section to this empirical counterpart of equation (7)

$$v_{jskg} = \alpha + \beta N_{jskg} + \gamma I_{jskg} + \mu X_{jskg} + \eta_{kg} + f(N_{sg}) + u_{jskg}, \qquad (14)$$

which differs from equation (10) because fixed effects must now be defined at the institution×grade level ($\eta_{kg} \neq \eta_{sg}$) and 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.

We estimate the above equation by IV using, as in Angrist and Lang (2004), the following set of indicator variables as instruments:

$$\Psi \in \{1(1 \leq C_{sg}^N < 2), 1(2 \leq C_{sg}^N < 3), ..., 1(24 \leq C_{sg}^N < 25)\}, \qquad (15)$$

These indicators are defined for each possibile level of the theoretical number of natives in a class, C_{sg}^N, predicted by equation (13) 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.

Results are reported in Table 3. In the first column the 2nd and 5th grades are pooled together and the outcome is the average language test score. While in the correspondent column of Table 2, based on the same sample, the estimate of β obtained without controlling

---

[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 the Appendix Tables A-2 - A-5.

for endogenous class size adjustments was positive but not statistically significant, the IV estimate in Table 3 is negative and significant for this outcome: keeping constant the number of immigrants, one additional native reduces the language test score of natives by 0.2 percent.

Much larger in size than in Table 2 and similarly negative is the IV estimate of γ: keeping constant the number of natives, one additional immigrant reduces the language test score of natives by 1.8 percent. This finding is particularly remarkable given that the estimate of Table 2 is as small as -0.05 percent. As we argued in Section 2, the reason of this difference is that the IV estimates of β and γ are not confounded by the endogenous class size adjustments implemented by principals when confronted with immigrant and native inflows.

These estimates imply that the effect on native language test scores induced by *adding one immigrant to a class while taking away a native*, and thus keeping class size constant, is -1.6 percent. This is δ: the Pure Composition Effect for the language test score of natives.

When the two grades are analysed separately in columns 2 and 3 of Table 3, results are qualitatively similar although in some case less significant. Nevertheless, the estimate for γ continues to be equally negative and significant in grade 5 and even larger than in grade 2. This evidence suggests that the negative Pure Composition Effect of an immigrant inflow does not fade away when kids get older.

In the last three columns of the table the outcome is the math test score of natives. In this case the IV estimate of β is essentially null, but the estimate of γ is again negative, significant and similar in size to the one for language (-1.7 percent). Thus, also in the case of math the PCE is negative and significant (-1.6 percent) and the IV estimates are larger in absolute size than the conventional estimates because they are obtained controlling for the endogenous class adjustments operated by principals when confronted with inflows of natives and immigrants.

Table 3 reports 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 also reported for both endogenous variables (namely the number of natives and immigrants in a class): we never reject the null. First stages are reported in the Appendix Tables A-2 and A-3.

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, in Table 4 we show 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 columns 1 and 4 the specification pools together the 2nd and 5th grades and is the same as in columns 1 and 4 of Table 3, but only schools in the north and centre of the country are considered.[11] The estimates of the Pure Composition Effects are slightly smaller ($\approx$ -1.2% in both language and math instead of - 1.6%) but still statistically significant.[12]

In columns 2 and 5 of Table 4 we restrict the sample to classes in the north and centre in which, according to the "cheating indicator" proposed by Angrist et al. (2014)[13], cheating is less likely to have occurred and in this case the estimates of the PCE gain in size and significance with respect to those in columns 1 and 4 (respectively -1.5% and -1.3% in language and math).

Finally, in columns 3 and 6 of Table 4 we consider only schools belonging to northern and central institutions in which an external monitor was sent by INVALSI[14] and again the

---

[11]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.

[12]The first stage regressions for these regressions and for the remaining ones commented below, are reported in the Appendix Tables A-4 and A-5.

[13]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.

[14]In these institutions external inspectors where randomly assigned to classrooms during the INVALSI test scores, as explained in Lucifora and Tonello (2012), with the following specific tasks: i) invigilate students

estimates of the PCE are not smaller and still statistically significant (respectively -1.5% and -1.4% in language and math). 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.

8   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.

The first contribution of our paper is to clarify, using an extended version of Lazear (2001) model of the educational production function, that a useful policy parameter that should be estimated is the causal effect of a change in class composition due to immigrant inflows net of the endogenous class size changes that are typically implemented by principals when confronted with such inflows and net of the mechanical class size effects that these inflows entail. This is what we call a Pure Composition Effect and we show that the existing literature has neglected it. This is not to say that estimates of the overall effect of an immigrant inflow, inclusive of the reactions of principals who adjust class size, are not interesting. Our claim is that estimates of the PCE are necessary to principals as well if they want to calibrate correctly their reactions to immigrant inflows, specifically but not only in terms of class size adjustments, avoiding waste of resources.

We then propose an empirical strategy to identify and estimate the PCE, inspired by Angrist and Lang (2004). The enrolment of natives in conjunction with the rule imposing a cap of 25 students per class generate the exogenous sources of variation in the number of natives and immigrants in a class that we need to estimate the PCE.

Our results suggest that this effect is sizeable: adding one immigrant to a class while taking away one native, reduces native performance in both language and math by approximately 1.6% in 2nd grade and does not fade away in 5th grade. The magnitude of these

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.

estimates 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 reduction of class size implemented by principals who fear the disruption caused by immigrant inflows.

References

Ammermuller, A., Pischke, J.S., 2009. Peer Effects in European Primary Schools: Evidence from PIRLS. Journal of Labour Economics 27, 315–348.

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., 2004. Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program. American Economic Review 94, 1613–1634.

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

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

Brunello, G., Rocco, L., 2011. The Effect of Immigration on the School Performance of Natives: Cross Country Evidence Using PISA Test Scores. Discussion Papers 5479. IZA.

Card, D., Rothstein, J., 2007. Racial Segregation and the Black-White Test Score Gap. Journal of Public Economics 91, 2158–2184.

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, F281–F307.

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

Hanushek, E., Kain, J., Rivkin, S., 2009. New Evidence About Brown V. Board of Education: The Complex Effects of School Racial Composition on Achievement. Journal of Labour Economics 27.

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.

Lazear, E., 2001. Educational Production. Quarterly Journal of Economics CXVI, 777–803.

Lucifora, C., Tonello, M., 2012. Students' Cheating as a Social Interaction: Evidence from a Randomized Experiment in a National Evaluation Program. Working paper 6967. Institute for the Study of Labor (IZA).

Ohinata, A., Van Ours, J., 2011. How Immigrant Children Affect the Academic Achievement of Native Duch Children. Discussion Papers 6212. IZA.

PISA, 2009. PISA 2006 Technical Report. Technical Report. 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 , 149–171.

Tonello, M., 2012. Social Interaction Between Native and Non-Native Students: Mechanisms and Evidence. Working paper 65. Università Cattolica del Sacro Cuore.

Table 1: Descriptive statistics for the language sample

	2nd grade Mean	2nd grade S.D.	5th grade Mean	5th grade S.D.
	Panel A. Class characteristics			
Percentage (0-100) correct answers in				
test score in language (natives)	0.67	0.11	0.71	0.08
test score in mathematics (natives)	0.61	0.11	0.65	0.10
test score in language (immigrants)	0.53	0.19	0.60	0.15
test score in 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			Mathematics		
	Pooled (1)	2nd grade (2)	5th grade (3)	Pooled (4)	2nd grade (5)	5th grade (6)
Number of natives: $\hat{\beta}$	0.0001 (0.0003)	-0.0007 (0.0005)	0.0009** (0.0004)	0.0002 (0.0004)	-0.0009 (0.0006)	0.0012** (0.0005)
Number of immigrants: $\hat{\gamma}$	-0.0049*** (0.0006)	-0.0064*** (0.0010)	-0.0034*** (0.0007)	-0.0042*** (0.0007)	-0.0046*** (0.0011)	-0.0039*** (0.0009)
(Confounded) composition effect: $\hat{\delta}$	-0.0050*** (0.0006)	-0.0057*** (0.0009)	-0.0043*** (0.0006)	-0.0044*** (0.0006)	-0.0037*** (0.0010)	-0.0052*** (0.0008)
Observations	25,943	12,859	13,084	25,936	12,854	13,082
School×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 shares 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 school×grade fixed effects. Standard errors are clustered at school×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, pooling grade 2 and 5
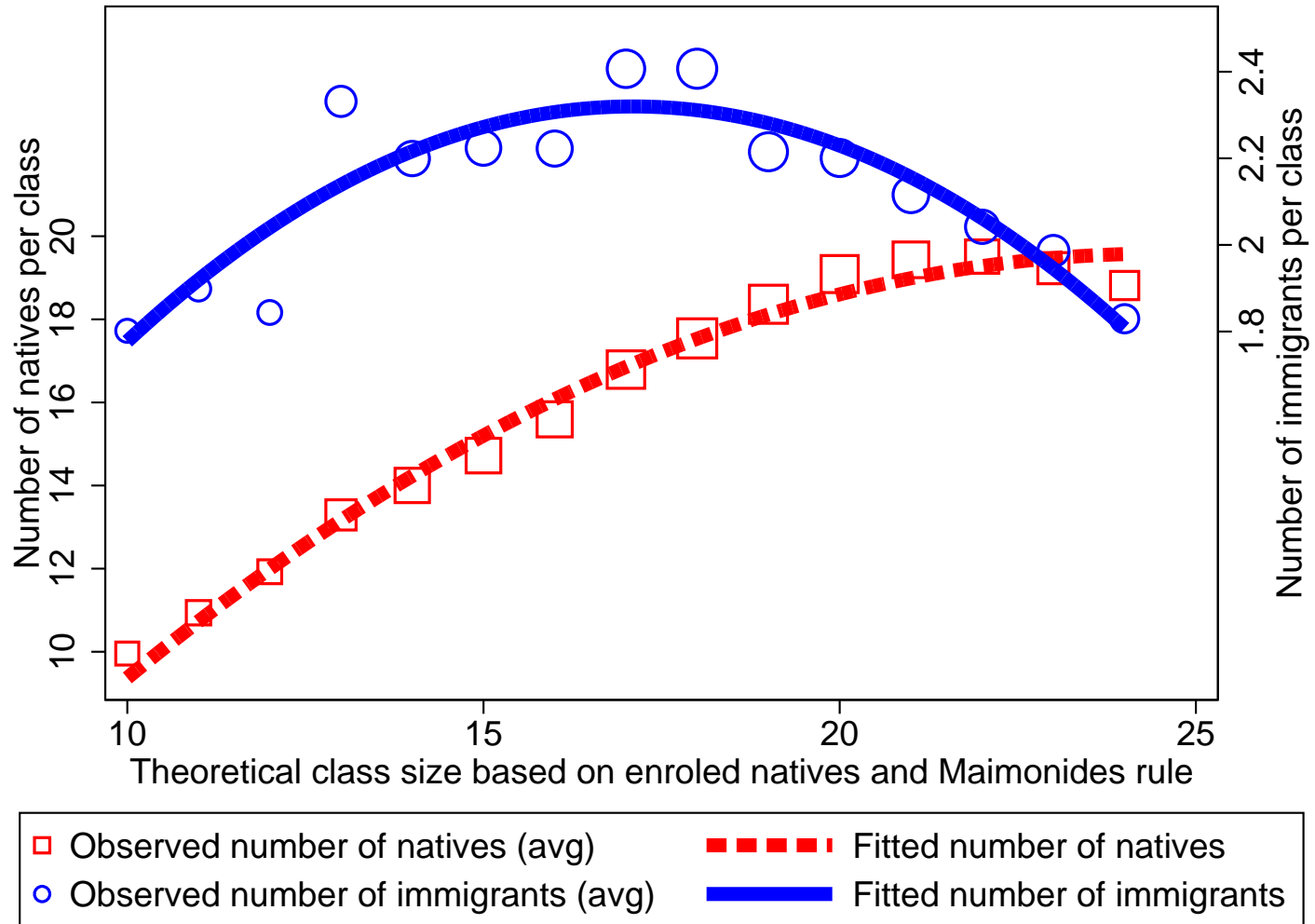
Notes: This figure is based on the language estimation sample described in Table 1, pooling together grade 2 and 5. Squares (left vertical axis) indicate the average number of natives per class in schools with the correspondent theoretical class size based on enrolled natives (horizontal axis). The dashed 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).

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

	Language			Mathematics		
	Pooled (1)	2nd grade (2)	5th grade (3)	Pooled (4)	2nd grade (5)	5th grade (6)
Number of natives: $\hat{\beta}$	-0.0019**	-0.0025**	-0.0011	-0.0010	-0.0013	-0.0005
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Number of immigrants: $\hat{\gamma}$	-0.0177***	-0.0150*	-0.0182***	-0.0168***	-0.0134	-0.0174**
	(0.005)	(0.008)	(0.006)	(0.006)	(0.008)	(0.007)
(Pure) composition effect: $\hat{\delta}$	-0.0158***	-0.0125	-0.0171**	-0.0158***	-0.0121	-0.0169**
	(0.005)	(0.008)	(0.006)	(0.005)	(0.008)	(0.007)
Observations	25,943	12,859	13,084	25,936	12,854	13,082
Institution×grade FE	✓	✓	✓	✓	✓	✓
Polynomial in natives enrolment	✓	✓	✓	✓	✓	✓
Class level controls	✓	✓	✓	✓	✓	✓
Hansen (p-value)	0.716	0.717	0.718	0.972	0.778	0.866
F test (excluded instruments)						
Natives	382.52	185.09	231.93	221.76	141.52	194.88
Immigrants	299.19	90.97	86.91	46.46	48.59	76.97

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 average test scores in language (math) for native students. 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 shares 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%.

Table 4: IV-FE estimates of the effect of the number of natives and immigrants on native test scores; schools in the north&centre

	Language			Mathematics		
	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**	-0.0008	-0.0006	-0.0019
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Number of immigrants: $\hat{\gamma}$	-0.0155***	-0.0172***	-0.0180**	-0.0129**	-0.0136**	0.0160*
	(0.005)	(0.005)	(0.008)	(0.005)	(0.006)	(0.008)
(Pure) composition effect: $\hat{\delta}$	-0.0129**	-0.0149***	-0.0147**	-0.0121**	-0.0130**	0.0141*
	(0.005)	(0.005)	(0.007)	(0.005)	(0.005)	(0.008)
Observations	19,001	18,636	4,730	19,005	18,630	4,733
Institution×grade FE	✓	✓	✓	✓	✓	✓
Polynomial in natives enrolment	✓	✓	✓	✓	✓	✓
Class level controls	✓	✓	✓	✓	✓	✓
Hansen (p-value)	0.373	0.197	0.475	0.878	0.592	0.694
F test (excluded instruments)						
Natives	243.61	230.40	92.58	165.09	156.13	103.9
Immigrants	159.90	156.49	13.17	39.65	40.14	14.35

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 (math) for natives students. Columns (1) and (4) report results for the baseline specification and sample. Columns (2) and (5) report results using only 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 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 shares 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%.

9 Appendix for online publication

Table A-1: Descriptive statistics for the maths sample

	2nd grade		5th grade	
	Mean	S.D.	Mean	S.D.
	Panel A. Class characteristics			
Percentage (0-100) correct answers in				
test score in language (natives)	0.67	0.11	0.71	0.08
test score in mathematics (natives)	0.61	0.11	0.65	0.10
test score in language (immigrants)	0.53	0.19	0.61	0.15
test score in 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 natives and immigrants; language sample

	Pooled		2nd grade		5th grade	
	Natives	Immigrants	Natives	Immigrants	Natives	Immigrants
	(1)	(2)	(3)	(4)	(5)	(6)
$1(2 \leq C_{sg}^N < 3)$	-9.7728***	4.5959***	-9.4634***	4.3086***	-	-
	(0.248)	(0.332)	(0.358)	(0.289)	(-)	(-)
$1(3 \leq C_{sg}^N < 4)$	-11.3015***	6.6418***	-12.1403***	6.4990***	-10.3415***	6.8028***
	(0.732)	(0.285)	(0.482)	(0.292)	(0.903)	(0.636)
$1(4 \leq C_{sg}^N < 5)$	-8.9129***	3.7710***	-8.8308***	3.6492***	-9.1321***	4.0337***
	(0.299)	(0.133)	(0.420)	(0.221)	(0.427)	(0.187)
$1(5 \leq C_{sg}^N < 6)$	-9.4750***	5.1826***	-8.5172***	4.3619***	-10.5157***	6.1653***
	(0.521)	(0.939)	(0.751)	(0.728)	(0.529)	(1.814)
$1(6 \leq C_{sg}^N < 7)$	-8.1358***	3.8279***	-7.6242***	3.6089***	-8.7490***	4.0409***
	(0.634)	(0.668)	(0.718)	(0.726)	(1.115)	(1.257)
$1(7 \leq C_{sg}^N < 8)$	-7.9428***	2.2046***	-7.4625***	2.0695***	-8.3415***	2.3725***
	(0.364)	(0.417)	(0.583)	(0.703)	(0.478)	(0.512)
$1(8 \leq C_{sg}^N < 9)$	-6.8997***	1.2292***	-6.7443***	0.9765***	-6.9976***	1.5179***
	(0.319)	(0.234)	(0.298)	(0.298)	(0.494)	(0.381)
$1(9 \leq C_{sg}^N < 10)$	-6.1386***	0.3754*	-5.8431***	0.0998	-6.4250***	0.6408**
	(0.296)	(0.201)	(0.414)	(0.285)	(0.428)	(0.283)
$1(10 \leq C_{sg}^N < 11)$	-5.3021***	-0.3148**	-4.9503***	-0.3994*	-5.6720***	-0.2453
	(0.281)	(0.159)	(0.388)	(0.224)	(0.410)	(0.226)
$1(11 \leq C_{sg}^N < 12)$	-4.3410***	-0.1975	-4.0613***	-0.3376	-4.6567***	-0.0452
	(0.279)	(0.156)	(0.387)	(0.225)	(0.404)	(0.213)
$1(12 \leq C_{sg}^N < 13)$	-3.2113***	-0.3530**	-2.9692***	-0.5226**	-3.4547***	-0.1731
	(0.277)	(0.156)	(0.385)	(0.228)	(0.402)	(0.211)
$1(13 \leq C_{sg}^N < 14)$	-2.0246***	-0.1174	-1.7690***	-0.2365	-2.2774***	0.0006
	(0.272)	(0.134)	(0.378)	(0.196)	(0.392)	(0.179)
$1(14 \leq C_{sg}^N < 15)$	-1.4913***	-0.1831	-1.3577***	-0.3503*	-1.6269***	-0.0117
	(0.263)	(0.129)	(0.366)	(0.189)	(0.382)	(0.173)
$1(15 \leq C_{sg}^N < 16)$	-0.7717***	-0.2249*	-0.6748*	-0.3379*	-0.8669**	-0.1137
	(0.264)	(0.132)	(0.364)	(0.194)	(0.384)	(0.178)
$1(16 \leq C_{sg}^N < 17)$	-0.0025	-0.2393*	0.0473	-0.4562**	-0.0540	-0.0224
	(0.259)	(0.128)	(0.360)	(0.190)	(0.375)	(0.171)
$1(17 \leq C_{sg}^N < 18)$	0.9159***	-0.0518	1.0483***	-0.2924	0.7807**	0.1855
	(0.259)	(0.124)	(0.359)	(0.181)	(0.375)	(0.170)
$1(18 \leq C_{sg}^N < 19)$	1.5276***	-0.0700	1.5993***	-0.2985	1.4533***	0.1608
	(0.262)	(0.126)	(0.365)	(0.185)	(0.376)	(0.171)
$1(19 \leq C_{sg}^N < 20)$	2.1222***	-0.1628	2.2547***	-0.3435*	2.0055***	0.0250
	(0.261)	(0.123)	(0.365)	(0.183)	(0.376)	(0.163)
$1(20 \leq C_{sg}^N < 21)$	2.5660***	-0.2096*	2.6524***	-0.3254*	2.4742***	-0.0842
	(0.265)	(0.124)	(0.373)	(0.181)	(0.379)	(0.167)
$1(21 \leq C_{sg}^N < 22)$	3.0235***	-0.3201**	3.0399***	-0.5385***	3.0074***	-0.1042
	(0.275)	(0.128)	(0.386)	(0.189)	(0.394)	(0.172)
$1(22 \leq C_{sg}^N < 23)$	2.8241***	-0.2892**	3.1854***	-0.5215***	2.4525***	-0.0472
	(0.286)	(0.131)	(0.395)	(0.189)	(0.415)	(0.181)
$1(23 \leq C_{sg}^N < 24)$	2.4659***	-0.3643***	2.3499***	-0.4768**	2.5863***	-0.2462
	(0.303)	(0.133)	(0.434)	(0.191)	(0.424)	(0.182)
$1(24 \leq C_{sg}^N < 25)$	1.9103***	-0.4099***	1.9594***	-0.6682***	1.8686***	-0.1432
	(0.301)	(0.135)	(0.421)	(0.197)	(0.431)	(0.185)
Observations	25,943	25,943	12,859	12,859	13,084	13,084
F test	382.5	299.2	185.1	90.97	231.9	86.91
Institution×grade FE	✓	✓	✓	✓	✓	✓
Polynomial in natives enrolment	✓	✓	✓	✓	✓	✓
Class level controls	✓	✓	✓	✓	✓	✓

Notes: The table reports in each column a different first stage regression correspondent to the IV estimates in columns (1), (2), (3) 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 (13) 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 3 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 shares 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%.

Table A-3: First Stage for the number natives and immigrants; math sample

	Pooled		2nd grade		5th grade	
	Natives	Immigrants	Natives	Immigrants	Natives	Immigrants
	(1)	(2)	(3)	(4)	(5)	(6)
$1(3 \leq C_{sg}^N < 4)$	-11.3807***	6.7143***	-12.1536***	6.5457***	-10.5265***	6.9302***
	(0.726)	(0.277)	(0.472)	(0.293)	(0.985)	(0.603)
$1(4 \leq C_{sg}^N < 5)$	-9.1258***	4.4309***	-9.4266***	4.5873***	-8.4256***	4.1289***
	(0.597)	(0.551)	(0.596)	(0.707)	(0.427)	(0.193)
$1(5 \leq C_{sg}^N < 6)$	-9.3532***	5.1456***	-8.4459***	4.3962***	-10.3433***	6.0883***
	(0.483)	(0.949)	(0.683)	(0.752)	(0.506)	(1.829)
$1(6 \leq C_{sg}^N < 7)$	-8.1968***	3.8229***	-7.8720***	3.6559***	-8.5680***	3.9841***
	(0.632)	(0.674)	(0.738)	(0.727)	(1.103)	(1.269)
$1(7 \leq C_{sg}^N < 8)$	-7.9340***	2.2149***	-7.4544***	2.0862***	-8.3256***	2.3645***
	(0.372)	(0.416)	(0.597)	(0.704)	(0.486)	(0.510)
$1(8 \leq C_{sg}^N < 9)$	-6.8945***	1.2364***	-6.7649***	1.0000***	-6.9849***	1.5197***
	(0.319)	(0.234)	(0.424)	(0.298)	(0.493)	(0.382)
$1(9 \leq C_{sg}^N < 10)$	-6.1565***	0.3921*	-5.8534***	0.1253	-6.4460***	0.6458**
	(0.296)	(0.201)	(0.412)	(0.283)	(0.428)	(0.284)
$1(10 \leq C_{sg}^N < 11)$	-5.3192***	-0.2942*	-4.9980***	-0.3529	-5.6563***	-0.2507
	(0.281)	(0.159)	(0.388)	(0.223)	(0.411)	(0.226)
$1(11 \leq C_{sg}^N < 12)$	-4.3480***	-0.1901	-4.0907***	-0.3120	-4.6428***	-0.0524
	(0.279)	(0.156)	(0.387)	(0.224)	(0.404)	(0.214)
$1(12 \leq C_{sg}^N < 13)$	-3.2119***	-0.3428**	-2.9831***	-0.4973**	-3.4468***	-0.1749
	(0.278)	(0.156)	(0.385)	(0.228)	(0.403)	(0.211)
$1(13 \leq C_{sg}^N < 14)$	-2.0155***	-0.1075	-1.7803***	-0.2197	-2.2519***	0.0055
	(0.272)	(0.134)	(0.378)	(0.196)	(0.393)	(0.180)
$1(14 \leq C_{sg}^N < 15)$	-1.5003***	-0.1669	-1.3767***	-0.3202*	-1.6309***	-0.0088
	(0.264)	(0.129)	(0.366)	(0.188)	(0.383)	(0.174)
$1(15 \leq C_{sg}^N < 16)$	-0.7625***	-0.2184*	-0.6734*	-0.3183*	-0.8541**	-0.1181
	(0.264)	(0.132)	(0.363)	(0.193)	(0.385)	(0.179)
$1(16 \leq C_{sg}^N < 17)$	-0.0010	-0.2339*	0.0360	-0.4410**	-0.0475	-0.0254
	(0.260)	(0.128)	(0.360)	(0.189)	(0.376)	(0.172)
$1(17 \leq C_{sg}^N < 18)$	0.9200***	-0.0439	1.0490***	-0.2771	0.7853**	0.1868
	(0.259)	(0.124)	(0.359)	(0.181)	(0.376)	(0.170)
$1(18 \leq C_{sg}^N < 19)$	1.5302***	-0.0663	1.5875***	-0.2831	1.4643***	0.1554
	(0.262)	(0.126)	(0.364)	(0.185)	(0.377)	(0.172)
$1(19 \leq C_{sg}^N < 20)$	2.1299***	-0.1586	2.2481***	-0.3330*	2.0224***	0.0246
	(0.261)	(0.123)	(0.365)	(0.182)	(0.377)	(0.164)
$1(20 \leq C_{sg}^N < 21)$	2.5703***	-0.2094*	2.6464***	-0.3184*	2.4827***	-0.0873
	(0.265)	(0.123)	(0.372)	(0.181)	(0.380)	(0.167)
$1(21 \leq C_{sg}^N < 22)$	3.0258***	-0.3182**	3.0381***	-0.5379***	3.0082***	-0.0994
	(0.275)	(0.128)	(0.386)	(0.188)	(0.394)	(0.172)
$1(22 \leq C_{sg}^N < 23)$	2.8402***	-0.2841**	3.2012***	-0.5098***	2.4681***	-0.0493
	(0.286)	(0.131)	(0.395)	(0.188)	(0.416)	(0.181)
$1(23 \leq C_{sg}^N < 24)$	2.4756***	-0.3614***	2.3459***	-0.4708**	2.6015***	-0.2419
	(0.303)	(0.132)	(0.434)	(0.191)	(0.424)	(0.182)
$1(24 \leq C_{sg}^N < 25)$	1.9202***	-0.4035***	1.9279***	-0.6482***	1.9152***	-0.1467
	(0.301)	(0.136)	(0.420)	(0.197)	(0.433)	(0.185)
Observations	25,936	25,936	12,854	12,854	13,082	13,082
F test	221.8	46.46	141.5	48.59	194.9	76.96
Institution×grade FE	✓	✓	✓	✓	✓	✓
Polynomial in natives enrolment	✓	✓	✓	✓	✓	✓
Class level controls	✓	✓	✓	✓	✓	✓

Notes: The table reports in each column a different first stage regression correspondent to the IV estimates in columns (4), (5), (6) 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 (13) 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 shares 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%.

Table A-4: First stage for the number of natives and immigrants; north&centre, language sample

	Baseline		Classes with cheating indicator = 0		Externally monitored institutions	
	Natives	Immigrants	Natives	Immigrants	Natives	Immigrants
	(1)	(2)	(3)	(4)	(5)	(6)
$1(2 \leq C_{sg}^N < 3)$	-9.8401***	4.6018***	-9.8554***	4.6955***		
	(0.289)	(0.357)	(0.301)	(0.383)		
$1(3 \leq C_{sg}^N < 4)$	-10.9353***	6.5424***	-10.9615***	6.5554***		
	(0.752)	(0.321)	(0.749)	(0.321)		
$1(4 \leq C_{sg}^N < 5)$	-8.6161***	3.9571***	-8.6632***	3.9563***	-10.0269***	3.6548***
	(0.348)	(0.180)	(0.352)	(0.180)	(0.660)	(0.451)
$1(5 \leq C_{sg}^N < 6)$	-9.1070***	5.2356***	-9.1793***	5.4499***	-10.6714***	4.8695**
	(0.527)	(0.961)	(0.565)	(1.040)	(0.948)	(2.403)
$1(6 \leq C_{sg}^N < 7)$	-7.5338***	4.0760***	-7.5566***	4.0981***	-9.4911***	3.8589***
	(0.685)	(0.764)	(0.700)	(0.802)	(0.628)	(1.045)
$1(7 \leq C_{sg}^N < 8)$	-7.5887***	2.2712***	-7.6561***	2.2477***	-8.4602***	2.7099***
	(0.395)	(0.454)	(0.403)	(0.464)	(0.717)	(0.883)
$1(8 \leq C_{sg}^N < 9)$	-6.5802***	1.2365***	-6.5460***	1.1803***	-7.5287***	0.3313
	(0.353)	(0.267)	(0.362)	(0.275)	(0.608)	(0.473)
$1(9 \leq C_{sg}^N < 10)$	-5.7260***	0.4008	-5.7336***	0.4149	-6.5653***	0.9445
	(0.331)	(0.248)	(0.335)	(0.254)	(0.602)	(0.586)
$1(10 \leq C_{sg}^N < 11)$	-4.8709***	-0.3461*	-4.8807***	-0.3410*	-5.7939***	-0.5055
	(0.318)	(0.202)	(0.320)	(0.204)	(0.570)	(0.413)
$1(11 \leq C_{sg}^N < 12)$	-3.9728***	-0.2173	-4.0060***	-0.2048	-4.2706***	-0.7077*
	(0.316)	(0.201)	(0.319)	(0.202)	(0.607)	(0.418)
$1(12 \leq C_{sg}^N < 13)$	-2.7097***	-0.3803*	-2.7596***	-0.3986**	-3.2830***	-0.7245**
	(0.312)	(0.196)	(0.314)	(0.197)	(0.592)	(0.364)
$1(13 \leq C_{sg}^N < 14)$	-1.7310***	-0.1541	-1.7935***	-0.1343	-2.3612***	-0.3837
	(0.301)	(0.168)	(0.302)	(0.169)	(0.568)	(0.354)
$1(14 \leq C_{sg}^N < 15)$	-1.1977***	-0.2324	-1.2306***	-0.2245	-1.8125***	-0.4806
	(0.292)	(0.164)	(0.293)	(0.164)	(0.553)	(0.364)
$1(15 \leq C_{sg}^N < 16)$	-0.3421	-0.2769	-0.3580	-0.3028*	-1.3786**	-0.6094*
	(0.296)	(0.169)	(0.297)	(0.170)	(0.558)	(0.361)
$1(16 \leq C_{sg}^N < 17)$	0.3862	-0.3320**	0.3480	-0.3291**	-0.3708	-0.2110
	(0.289)	(0.163)	(0.291)	(0.163)	(0.550)	(0.337)
$1(17 \leq C_{sg}^N < 18)$	1.2305***	-0.0964	1.1852***	-0.0845	0.6734	-0.2468
	(0.286)	(0.158)	(0.286)	(0.157)	(0.526)	(0.328)
$1(18 \leq C_{sg}^N < 19)$	1.8816***	-0.1011	1.8660***	-0.1098	1.4049***	-0.2210
	(0.292)	(0.161)	(0.293)	(0.160)	(0.532)	(0.329)
$1(19 \leq C_{sg}^N < 20)$	2.4841***	-0.2432	2.4440***	-0.2467	1.3253**	-0.2480
	(0.291)	(0.156)	(0.291)	(0.156)	(0.548)	(0.327)
$1(20 \leq C_{sg}^N < 21)$	2.8728***	-0.2558	2.8607***	-0.2754*	1.9106***	-0.3368
	(0.298)	(0.159)	(0.298)	(0.158)	(0.563)	(0.319)
$1(21 \leq C_{sg}^N < 22)$	3.3136***	-0.4454***	3.2350***	-0.4061**	3.2753***	-0.8437**
	(0.312)	(0.164)	(0.313)	(0.163)	(0.579)	(0.340)
$1(22 \leq C_{sg}^N < 23)$	2.8697***	-0.3806**	2.8202***	-0.3703**	2.3477***	-0.4726
	(0.323)	(0.170)	(0.325)	(0.169)	(0.598)	(0.349)
$1(23 \leq C_{sg}^N < 24)$	2.5866***	-0.5195***	2.5487***	-0.5206***	1.3195**	-0.5474
	(0.347)	(0.171)	(0.347)	(0.171)	(0.622)	(0.335)
$1(24 \leq C_{sg}^N < 25)$	2.0317***	-0.4925***	2.0309***	-0.5056***	1.4960**	-0.5141
	(0.344)	(0.178)	(0.346)	(0.177)	(0.622)	(0.363)
Observations	19,001	19,001	18,636	18,636	4,730	4,730
F test	243.6	159.9	230.4	156.5	92.58	13.17
Institution×grade FE	✓	✓	✓	✓	✓	✓
Polynomial in natives enrolment	✓	✓	✓	✓	✓	✓
Class level controls	✓	✓	✓	✓	✓	✓

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 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 (13) 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 shares 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%.

Table A-5: First stage for the number of natives and immigrants; north&centre, math sample

	Baseline		Classes with cheating indicator = 0		Externally monitored institutions	
	Natives	Immigrants	Natives	Immigrants	Natives	Immigrants
	(1)	(2)	(3)	(4)	(5)	(6)
$1(3 \leq C_{sg}^N < 4)$	-11.0241***	6.6141***	-11.0418***	6.6166***	-	-
	(0.753)	(0.314)	(0.752)	(0.309)	(-)	(-)
$1(4 \leq C_{sg}^N < 5)$	-8.7855***	4.5727***	-8.8284***	4.5927***	-9.9656***	3.6605***
	(0.552)	(0.548)	(0.569)	(0.564)	(0.664)	(0.439)
$1(5 \leq C_{sg}^N < 6)$	-9.0098***	5.1861***	-9.0875***	5.4091***	-10.9806***	4.9772**
	(0.489)	(0.974)	(0.521)	(1.050)	(1.062)	(2.498)
$1(6 \leq C_{sg}^N < 7)$	-7.6271***	4.0751***	-7.6551***	4.0973***	-9.6454***	3.9038***
	(0.694)	(0.771)	(0.709)	(0.809)	(0.614)	(1.027)
$1(7 \leq C_{sg}^N < 8)$	-7.5669***	2.2780***	-7.6269***	2.2471***	-8.5560***	2.7637***
	(0.403)	(0.453)	(0.412)	(0.462)	(0.729)	(0.892)
$1(8 \leq C_{sg}^N < 9)$	-6.5663***	1.2422***	-6.5255***	1.1825***	-7.5446***	0.3124
	(0.353)	(0.267)	(0.363)	(0.276)	(0.604)	(0.472)
$1(9 \leq C_{sg}^N < 10)$	-5.7234***	0.4202*	-5.7313***	0.4326*	-6.6530***	1.0029*
	(0.331)	(0.248)	(0.335)	(0.254)	(0.598)	(0.584)
$1(10 \leq C_{sg}^N < 11)$	-4.8838***	-0.3284	-4.8911***	-0.3269	-5.8730***	-0.4817
	(0.318)	(0.202)	(0.320)	(0.204)	(0.571)	(0.413)
$1(11 \leq C_{sg}^N < 12)$	-3.9752***	-0.2095	-4.0049***	-0.1993	-4.3673***	-0.6836*
	(0.316)	(0.201)	(0.319)	(0.202)	(0.604)	(0.413)
$1(12 \leq C_{sg}^N < 13)$	-2.7098***	-0.3674*	-2.7541***	-0.3887**	-3.3728***	-0.6821*
	(0.313)	(0.196)	(0.315)	(0.197)	(0.592)	(0.366)
$1(13 \leq C_{sg}^N < 14)$	-1.7159***	-0.1458	-1.7766***	-0.1279	-2.3658***	-0.3827
	(0.302)	(0.169)	(0.302)	(0.169)	(0.569)	(0.355)
$1(14 \leq C_{sg}^N < 15)$	-1.2066***	-0.2120	-1.2366***	-0.2068	-1.9017***	-0.4393
	(0.293)	(0.164)	(0.295)	(0.165)	(0.555)	(0.362)
$1(15 \leq C_{sg}^N < 16)$	-0.3364	-0.2679	-0.3491	-0.2968*	-1.4065**	-0.6045*
	(0.297)	(0.169)	(0.298)	(0.170)	(0.559)	(0.361)
$1(16 \leq C_{sg}^N < 17)$	0.3911	-0.3239**	0.3545	-0.3235**	-0.4344	-0.1805
	(0.290)	(0.163)	(0.291)	(0.163)	(0.552)	(0.337)
$1(17 \leq C_{sg}^N < 18)$	1.2342***	-0.0854	1.1895***	-0.0743	0.6357	-0.2285
	(0.287)	(0.158)	(0.287)	(0.158)	(0.527)	(0.329)
$1(18 \leq C_{sg}^N < 19)$	1.8924***	-0.0963	1.8785***	-0.1061	1.3464**	-0.2014
	(0.293)	(0.161)	(0.293)	(0.160)	(0.532)	(0.330)
$1(19 \leq C_{sg}^N < 20)$	2.4962***	-0.2395	2.4578***	-0.2423	1.2757**	-0.2215
	(0.291)	(0.156)	(0.292)	(0.156)	(0.550)	(0.328)
$1(20 \leq C_{sg}^N < 21)$	2.8703***	-0.2566	2.8590***	-0.2783*	1.8506***	-0.3131
	(0.298)	(0.159)	(0.299)	(0.158)	(0.564)	(0.319)
$1(21 \leq C_{sg}^N < 22)$	3.3185***	-0.4386***	3.2403***	-0.4022**	3.2114***	-0.8193**
	(0.313)	(0.164)	(0.314)	(0.163)	(0.582)	(0.341)
$1(22 \leq C_{sg}^N < 23)$	2.8862***	-0.3821**	2.8386***	-0.3740**	2.2879***	-0.4484
	(0.324)	(0.170)	(0.325)	(0.169)	(0.599)	(0.349)
$1(23 \leq C_{sg}^N < 24)$	2.6008***	-0.5172***	2.5641***	-0.5196***	1.2856**	-0.5412
	(0.347)	(0.171)	(0.348)	(0.171)	(0.622)	(0.335)
$1(24 \leq C_{sg}^N < 25)$	2.0357***	-0.4861***	2.0361***	-0.5012***	1.4358**	-0.4876
	(0.345)	(0.178)	(0.347)	(0.177)	(0.625)	(0.363)
Observations	19,005	19,005	18,630	18,630	4,733	4,733
F test	165.1	39.65	156.1	40.13	103.9	14.35
Institution×grade FE	✓	✓	✓	✓	✓	✓
Polynomial in natives enrolment	✓	✓	✓	✓	✓	✓
Class level controls	✓	✓	✓	✓	✓	✓

Notes: The table reports in each column a different first stage regression correspondent to the IV estimates in columns (4), (5), (6) 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 (13) 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 shares 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%.