

IZA DP No. 8732

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

Rosario Maria Ballatore
Margherita Fort
Andrea Ichino

December 2014

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

Rosario Maria Ballatore

Bank of Italy

Margherita Fort

*University of Bologna,
CESifo and IZA*

Andrea Ichino

*University of Bologna, EUI,
CEPR, CESifo and IZA*

Discussion Paper No. 8732
December 2014

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

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

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

Corresponding author:

Andrea Ichino
Department of Economics
European University Institute
Via Piazzola 43
50133 Firenze
Italy
E-mail: andrea.ichino@eui.eu

* 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 useful comments from seminar participants at the European University Institute, the University of Bologna, the II International Workshop on Applied Economics of Education and 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. 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. 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 fails to distinguish between class composition and class size effects.

The first contribution of this paper is to clarify the importance of this distinction: the 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. First, by allowing for two types of students, natives and immigrants, each with its own probability of disruption. Second, 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

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. However, the importance of controlling for these principals' reactions seems to have been disregarded in the existing literature.

The second contribution of our paper is to fill in this gap by adapting to our context 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 and classes are formed according to pre-enrolled students with a similar cap at 25. On this basis, requests for teachers are presented by each school to local educational authorities. Due to higher geographic mobility and lack of information on Italian institutions, immigrants typically do not pre-enrol and are allocated to classes in September when they show up for school. Therefore the pre-enrolment of natives generates, at each 25-or-multiples threshold, a splitting of classes that at the same time makes room for the admission of "last minute" immigrants. To apply the identification strategy of Angrist and Lang (2004), we restrict the empirical analysis to "schools" defined as units in which students can pre-enrol, but that are grouped in educational "institutions" managed by one principal. As we will see, the Italian Ministry of Education instructs principals of these educational institutions to use the information on the predicted size of classes in the schools that they manage, in order to allocate immigrant students where there is more space because of the exogenous interaction of pre-enrolment and rules of class

formation. The explicit assumption behind these instructions is that the re-allocation would help reducing the disruption that immigrants may cause.

As a result, when the expected class size in a school, based on native pre-enrolment, is large (e.g. in the 23-25 range), principals try to avoid filling the last available places with immigrants and exploit the other schools, in their educational institution, that have a smaller predicted class size, if they exist. These are schools that have a pre-enrolment level significantly below 25 or that, as a result of splitting induced by native pre-enrolment, have a class size in the order of 15-20 students. Therefore the pre-enrolment of natives in conjunction with the rule imposing a cap of 25 students per class, generates an exogenous variation in both the numbers of natives and immigrants, *without leaving room for endogenous reactions by principals in terms of class size adjustments*. This is 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 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

¹In 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.

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 maximum 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 maximum performance:

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

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.

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:

$$V = \bar{V}P^{\phi C} < \bar{V} \quad (2)$$

where $\phi \in \mathbb{R} \setminus \{0\}$. 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:

$$\text{Max}_C \Pi = P^{\phi C} - \frac{W}{C} \quad (3)$$

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

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

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

$$C^* = f(P, W, \phi) \quad (5)$$

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:

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

where the parameters $\phi_{qh} \in \mathbb{R} \setminus \{0\}$, 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:

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

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:

$$\begin{aligned} \beta &= \frac{\partial v_n}{\partial N} = p_n \phi_{nn} \\ \gamma &= \frac{\partial v_n}{\partial I} = p_i \phi_{in} \end{aligned} \quad (8)$$

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:

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

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

²A 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.

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 none of them aims at identifying 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: gender, previous attendance of nursery or kindergarten, highest educational level achieved by parents and parental occupational status. We aggregate this information 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

³In 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.

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 follow (Angrist et al., 2014, p.7) in dropping from the analysis the schools in which less than 10 and more than 160 students are enrolled in the 2nd or in the 5th grade: this leaves us with 15,398 schools. We then drop the “stand-alone” schools that are not grouped together with other schools in educational institutions managed by a single principal; as 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.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 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.

⁴Note 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.

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

$$v_{jsg} = \alpha + \beta N_{jsg} + \gamma I_{jsg} + \mu X_{jsg} + \eta_{sg} + \epsilon_{jsg} \quad (10)$$

where v_{jsg} is the (log) of the average test score of natives in class j belonging to school s of institution k in grade g ; N_{jsg} (I_{jsg}) is the number of natives (immigrants) in class j ; X_{jsg} is a set of predetermined control variables defined at the class level⁵, while η_{sg} denotes school \times grade fixed effects. The term ϵ_{jsg} 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 should be underestimated more than β . In this scenario the estimate of δ is lower than the true PCE.

⁵Specifically, 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.

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 Space for immigrants and an alternative identification strategy for the PCE

Although they do not focus on the effects of immigrants in a classroom, Angrist and Lang (2004) study a similar problem with an identification strategy that we can adapt to the Italian case in order to obtain a more convincing estimate of the PCE. They are interested in identifying the effect of an increase in the number of disadvantaged students that are offered the possibility to attend better schools in the Boston area, within the de-segregation program known as Metropolitan Council for Educational Opportunity (Metco). In the receiving schools, classes are formed according to a "Maimonides-type rule" (Angrist and Lavy, 1999) such that principals are required to cap class size at 25 students per class, splitting classes whenever enrolment of "Non-Metco" students goes above the 25 threshold or its multiples. As a result of this rule, the enrolment of "Non-Metco" students generates, at each threshold, an exogenous source of variation for both the number of "Non-Metco" students in a class and the remaining available space to admit "Metco" students. These students are in fact admitted only in those classes in which, as a result of the rule of class formation in combination with the enrolment of "Non-Metco" students, class size is small enough to generate sufficient space.

In Italy, the interaction between a similar rule of class formation and the enrolment of natives offers in the same way a source of exogenous variation in the number of natives and immigrants in a class, that can be used to construct the instruments that we need.

In the month of February of each year Italian families are invited to pre-enrol their children, for the following academic year, in the schools near where they live.⁶ On the basis of this pre-enrolment information, principals decide the number of classes being constrained by a "Maimonides-type rule" like the one prevailing in the schools of the Boston area: also in Italy, no class should have more than 25 students (and less than 10), with a 10% margin of flexibility around these thresholds. This margin is used, if necessary, when total effective enrolment becomes known in September, at the beginning of the academic year.

⁶The 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.

Ideally we would like to observe, for each school, the number of pre-enrolled students in February, which is effectively the indicator on which classes are formed by principals, but information on this variable is not available. Nevertheless, native enrolment is a good proxy of February pre-enrolment because foreign parents are less likely to pre-enrol their children, given their higher geographic mobility and limited information on Italian institutional requirements. For example, according to the Ministry⁷, in the school year 2013-2014 approximately 35000 students enrolled later than February, corresponding to approximately 6% of total enrolment. Of these late enrolled students, almost 29% are immigrants, while the fraction of non-Italian students out of total enrolment is less than 13% in our data (see Table 1).

As a result of this process of class formation, the theoretical number of natives in a generic class of grade g in school s , that we denote as C_{sg}^N , is determined by native enrolment N_{sg} in that school and grade according to the following “Maimonides-type” rule:

$$C_{sg}^N = \frac{N_{sg}}{\text{Int}\left(\frac{N_{sg}-1}{25}\right) + 1} \quad (11)$$

Figure 1 shows that the effective number of natives in a class tracks fairly closely the theoretical number of natives as predicted by equation (11).⁸ The dashed line represents C_{sg}^N , while the dotted line describes the number of natives effectively observed in a class for each level of native enrolment N_{sg} . At the thresholds corresponding to 25 enrolled natives (or multiples), both the dashed and the dotted lines decrease sharply and gradually increase with native enrolment immediately after, generating discontinuities in the relationship between native enrolment and the number of natives in a class that we can exploit for identification.

At the same time, the interaction between the rules of class formation and native enrolment generates also an exogenous source of variation for the number of immigrants in a class. Because of late enrolment in September, in which immigrants are over-represented, principals are forced to adapt the final number of students in at least some classes. These late adjustments in September typically cannot concern the number of classes in a school-grade (no more time for such changes) but can affect the number of students in those classes in which,

⁷We are grateful to Dr. Gianna Barbieri who gave us this aggregate information that concerns the enrolment in the first year of primary school.

⁸In this figure, which is based on the language sample described in Table 1, grades 2 and 5 are pooled together. The Appendix Figure A-1 shows the same evidence separately for grades 2 and 5.

because of how they have been formed in February, there is room for more students. For example, in a school-grade in which 30 students pre-enrol in February, two classes are formed with 15 students. In each of these classes the space for admitting late enrolled students in September amounts to 10 slots. In a school-grade in which 25 students enrol in February, instead, only one class is formed with no space for late enrolled students in September. Or in a school-grade with 23 enrolled students in February, only one class is formed as well and space for only two late students is available in September.

The availability of this space is crucially relevant for the determination of the number of immigrants in Italian classes. Principals are instructed by the Ministry to direct foreign students towards schools where, because of how classes are formed in February, the predicted number of natives per class makes enough space for immigrants to reduce any disruption (whether correctly anticipated or not). 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 form networks of schools in which a rational territorial distribution of these students can be achieved. [...] 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.”

Figure 2 shows how, because of the rules of class formation, the theoretically available and the actually used space for immigrants are a non linear function of native enrolment. The dashed line describes the available space for immigrants that is approximated by the maximum number of students in a class minus the predicted number of natives as defined by equation (11). For example, as shown in the left panel of Figure 2, if 10 natives enrol before February in a school, there will be one class in which the available space for immigrants, who might enrol later, is 15 units. If native enrolment increases up to 25, the available space for immigrants declines down to zero. When native enrolment goes above the 25 threshold, there will be two classes in which the available space for immigrants is high again (starting from approximately 12-13 units), and from this levels it goes once more down to zero if native enrolment grows to 50. This pattern replicates similarly over the entire range of native enrolment.

The solid line describes instead the space actually used for immigrants in a class as a function of native enrolment. More precisely, this is the average number of immigrants

actually observed in all the classes corresponding to each level of natives enrolment, net of institution fixed effects.⁹ The reason to partial out the institution fixed effects is that principals have leeway to allocate immigrants according to the available space but only between the schools that they control, i.e. the ones belonging to the institution that they manage. This is the reason why, as anticipated in Section 4, we restrict the analysis to institutions comprising more than one school under the same principal. In the left panel of Figure 2, the average space that principals actually use for immigrants in the schools that they manage (the solid line) is correlated with the available space predicted according to native enrolment in each school (the dashed line), although the latter is clearly not the only determinant of the former: specifically, both the available and the used space decrease when the predicted number of natives in a class based on native enrolment grows near 25 units, and spike up immediately after each class-splitting threshold. The right panel of Figure 2 shows however that this pattern weakens at higher enrolment levels, similarly to what happens in Figure 1 for the predicted and actual number of natives in a class.

It should be clear, from what we said so far, that one single underlying driving force, represented by native enrolment in each school-grade combined with the rules of class formation, is non-linearly generating two sources of variation for the number of immigrants and natives. This feature of the identification strategy may potentially jeopardise the possibility to obtain precise IV estimates. However, in our specific application, likewise in Angrist and Lang (2004), the non-linearities of the relationship between the instruments and the numbers of natives and immigrants are sufficient to produce precise estimates that are also not confounded by the endogenous reactions of principals in terms of class size adjustments. This is exactly what we need to identify and estimate the PCE, as the next section shows.

7 New evidence

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

$$v_{jskg} = \alpha + \beta N_{jskg} + \gamma I_{jskg} + \mu X_{jskg} + \eta_{kg} + f(N_{sg}) + u_{jskg}, \quad (12)$$

⁹To compute this average number, we have regressed the number of immigrants in a class on a set of dummies for each level of native enrolment, absorbing the institutions fixed effect. The solid line plots the coefficients of the native enrolment dummies in these regressions, rescaled by adding the constant.

which differs from equation (10) because fixed effects must now be defined at the institution \times grade level ($\eta_{kg} \neq \eta_{sg}$) and because a polynomial in native enrolment at the school \times 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:

$$\Psi \in \{1(1 \leq C_{sg}^N < 2), 1(2 \leq C_{sg}^N < 3), \dots, 1(24 \leq C_{sg}^N < 25)\}, \quad (13)$$

These indicators are defined for each possible level of the theoretical number of natives in a class, C_{sg}^N , predicted by equation (11) according to the rules of class formation as a function of native enrolment at the school \times 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 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.

¹⁰Note 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.

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. As a result the 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). Once again, also for this outcome the IV estimates are more negative and significant 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 INVALSI¹⁴ and again the 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 this issue. An example is the rule introduced by the Italian Ministry of Education, according

¹¹North 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.

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

¹³This 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, p.47-48) for more details. We thank these authors for having shared with us the information that they constructed.

¹⁴In these institutions external inspectors were randomly assigned to classrooms during the INVALSI test scores (see Lucifora and Tonello (2012)) 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.

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 the relevant policy parameter to be estimated should be 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 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, borrowed from Angrist and Lang (2004). The pre-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 in the space for 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 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 Skills. *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		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	0.60	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	

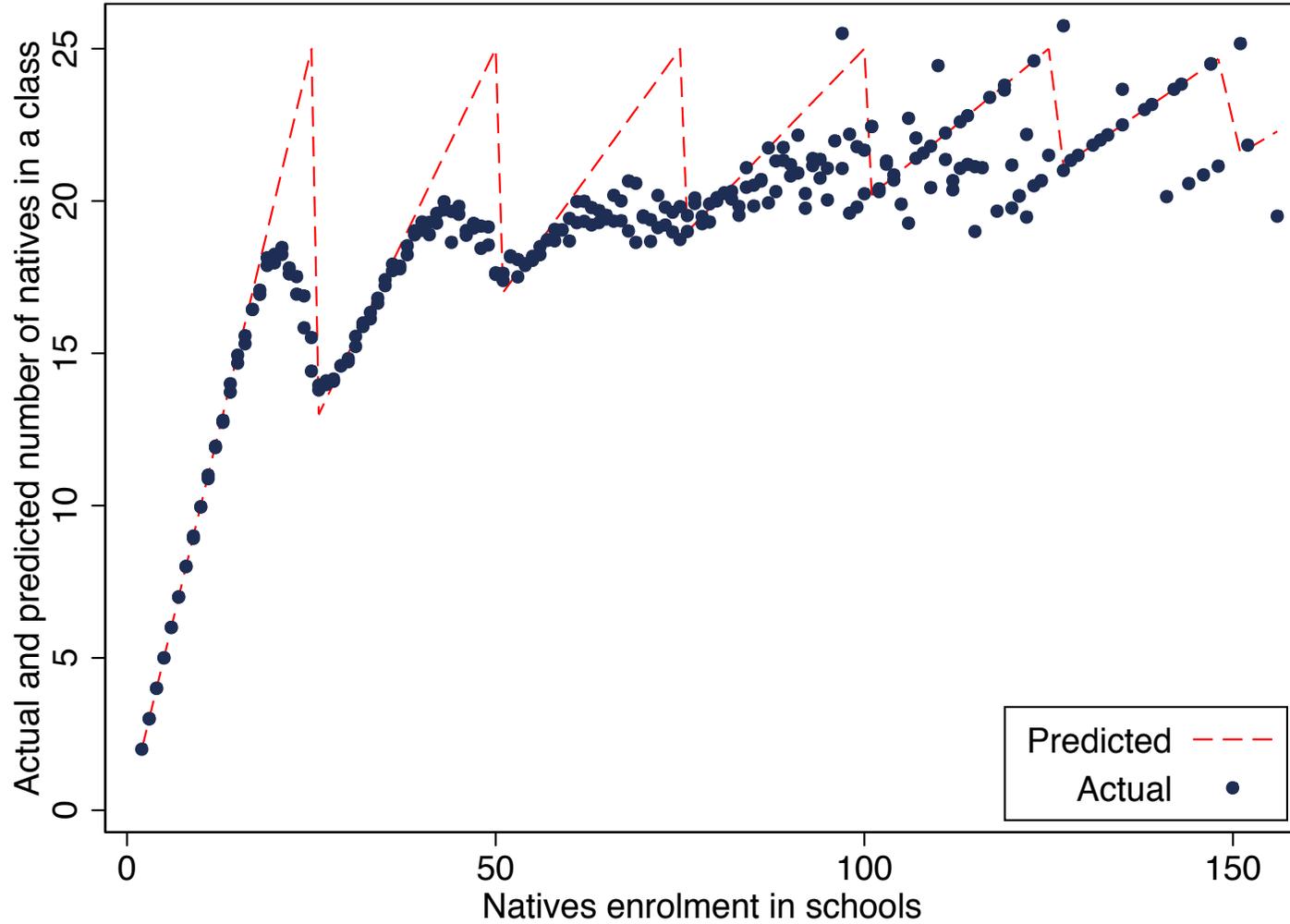
Notes: The unit of observation is a class in Panel A, a school in Panel B and an institution in Panel C. Institutions are groups of schools managed by the same principal. The family and individual background characteristics in Panel A are the shares of natives in a class who have that specific characteristic 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 \times 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 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 \times grade fixed effects. Standard errors are clustered at school \times grade level. A * denotes significance at 10%; a ** denotes significance at 5%; a *** denotes significance at 1%.

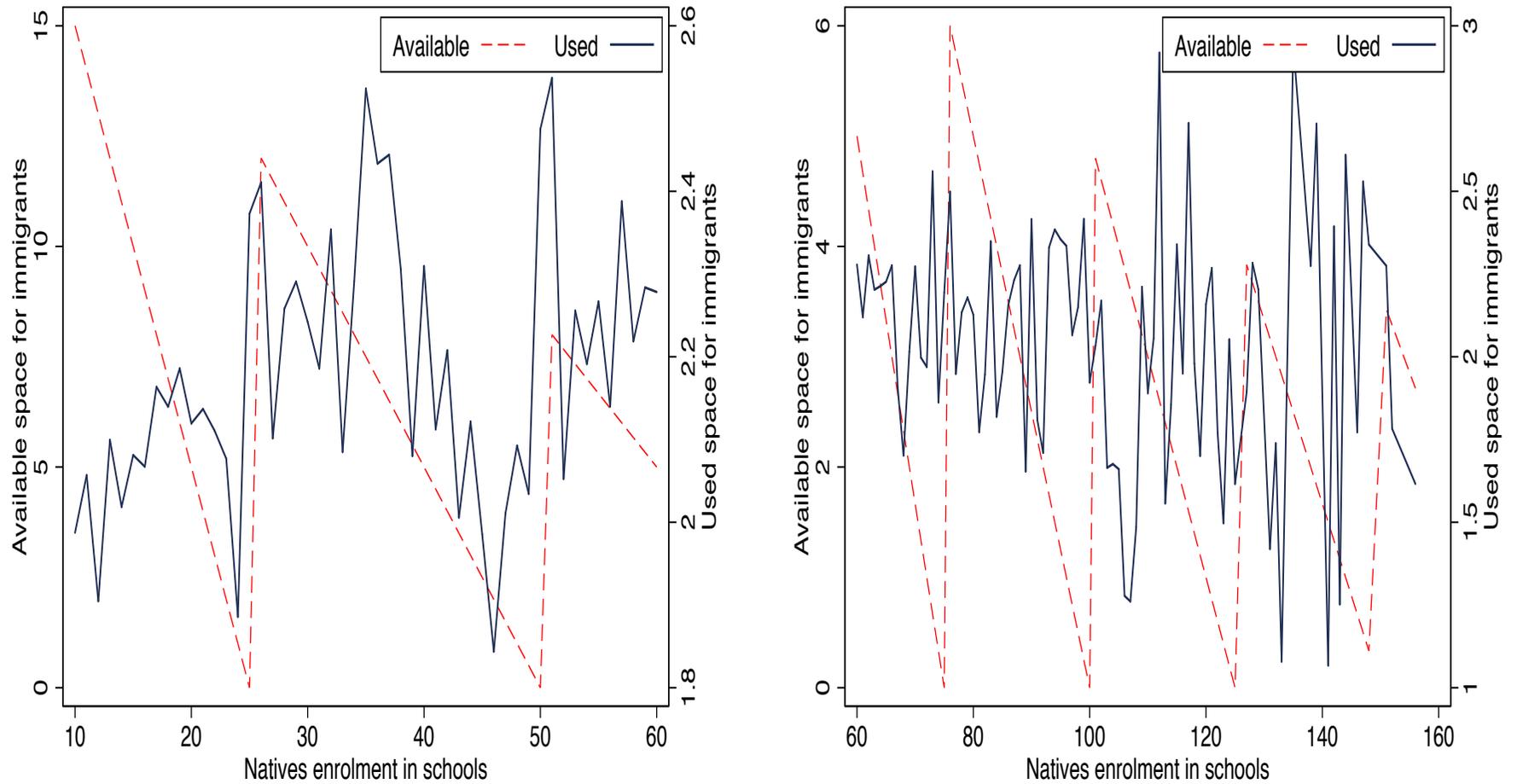
Figure 1: Actual and predicted number of natives in a class based on native enrolment, pooling grade 2 and 5



Notes: This figure is based on the language estimation sample described in Table 1, pooling together grade 2 and 5. The dashed line plots the theoretical number of natives in a class predicted by equation (11) according to the rules of class formation as a function of native enrolment at the school \times grade level. The dotted line describes instead the actual average number of natives for each level of native enrolment.

Figure 2: Available and used space for immigrants according to natives enrolment in schools, pooling grade 2 and grade 5

25



Notes: This figure is based on the language estimation sample described in Table 1, pooling together grade 2 and grade 5. The available space for immigrants, described by the dashed line, is defined as the maximum number of students in a class (25) minus the theoretical number of natives in a class predicted by equation (11) according to the rules of class formation as a function of native enrolment at the school \times grade level. The used space for immigrants (the solid line) is the average number of immigrants in a class by level of native enrolment net of institution fixed effects. More precisely, we regress the number of immigrants in a class on a set of dummies for each level of native enrolment, absorbing the institutions fixed effects. The solid line plots the coefficients of the native enrolment dummies in these regressions, rescaled by adding the constant. The left scale reports the range of variation of the available space for immigrants. The right scale refers to the used space for immigrants.

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** (0.001)	-0.0025** (0.001)	-0.0011 (0.001)	-0.0010 (0.001)	-0.0013 (0.001)	-0.0005 (0.001)
Number of immigrants: $\hat{\gamma}$	-0.0177*** (0.005)	-0.0150* (0.008)	-0.0182*** (0.006)	-0.0168*** (0.006)	-0.0134 (0.008)	-0.0174** (0.007)
(Pure) composition effect: $\hat{\delta}$	-0.0158*** (0.005)	-0.0125 (0.008)	-0.0171** (0.006)	-0.0158*** (0.005)	-0.0121 (0.008)	-0.0169** (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 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¢re

	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*** (0.001)	-0.0022*** (0.001)	-0.0033** (0.001)	-0.0008 (0.001)	-0.0006 (0.001)	-0.0019 (0.002)
Number of immigrants: $\hat{\gamma}$	-0.0155*** (0.005)	-0.0172*** (0.005)	-0.0180** (0.008)	-0.0129** (0.005)	-0.0136** (0.006)	0.0160* (0.008)
(Pure) composition effect: $\hat{\delta}$	-0.0129** (0.005)	-0.0149*** (0.005)	-0.0147** (0.007)	-0.0121** (0.005)	-0.0130** (0.005)	0.0141* (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 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

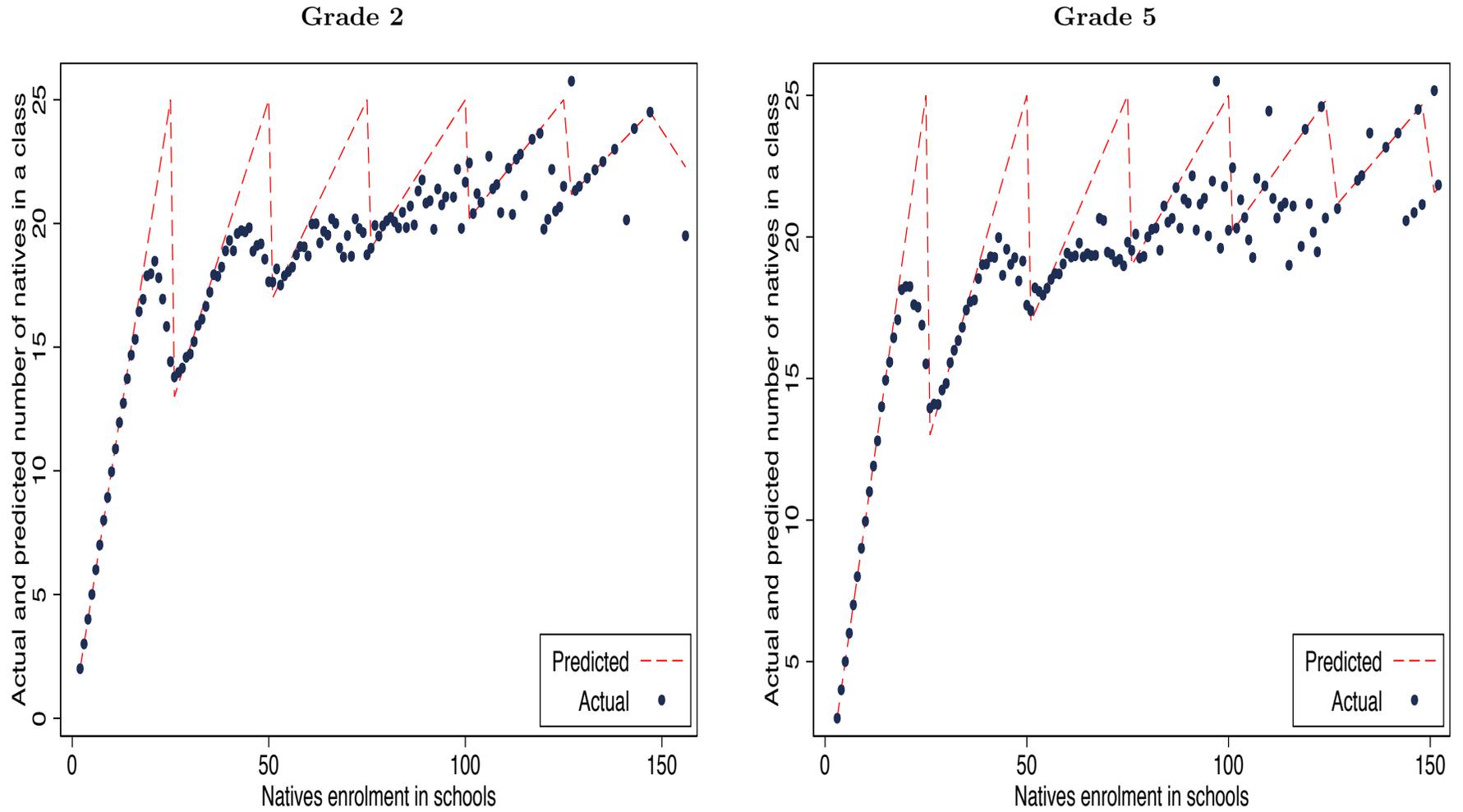
Table A-1: Descriptive statistics for the maths sample

	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.

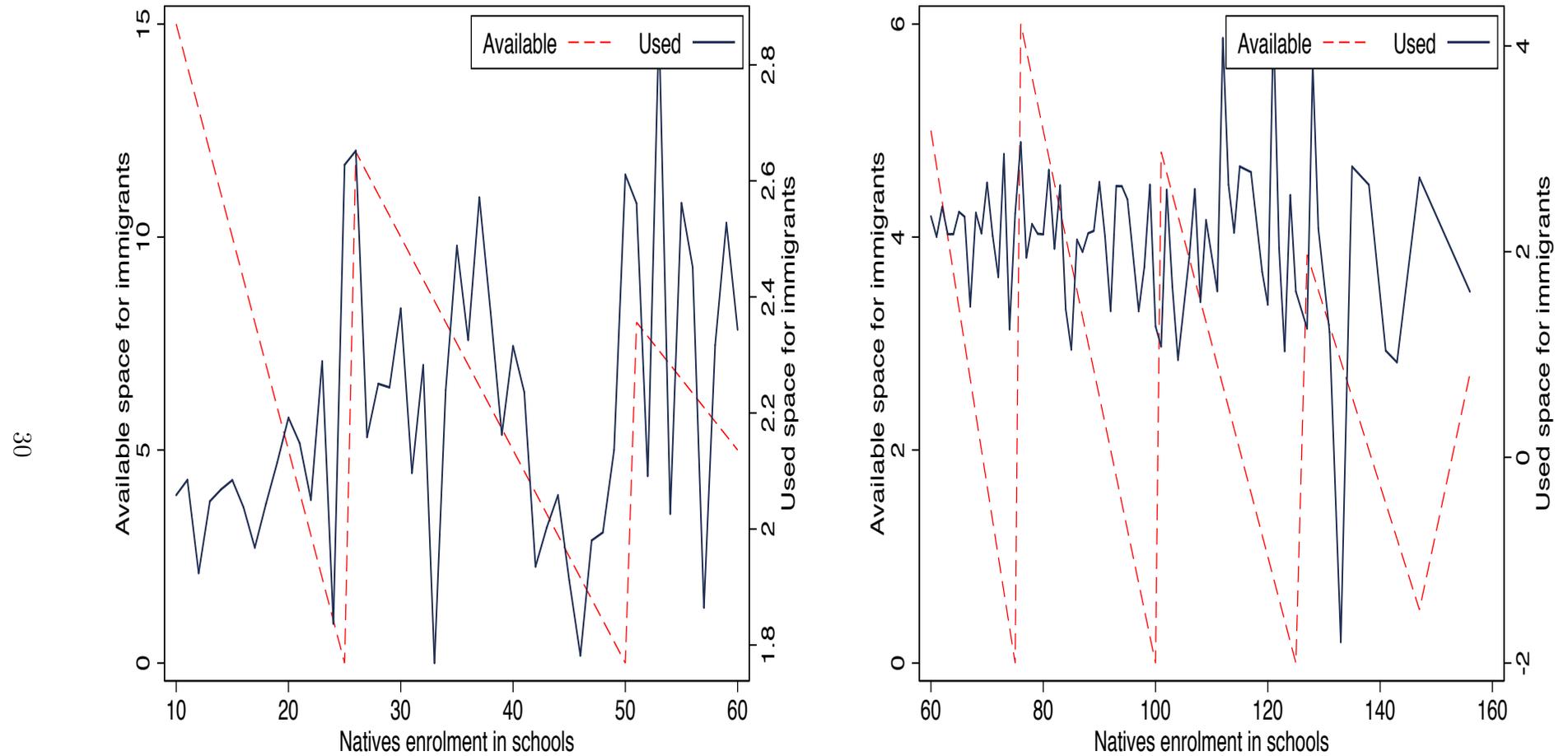
Figure A-1: Actual and predicted number of natives in a class based on native enrolment, separately by grade; language sample.

29



Notes: These figures are based on the language estimation sample described in Table 1, for grade 2 and 5. The dashed line plots the theoretical number of natives in a class predicted by equation (11) according to the rules of class formation as a function of native enrolment at the school \times grade level. The dotted line describes instead the actual average number of natives in a class for each level of native enrolment.

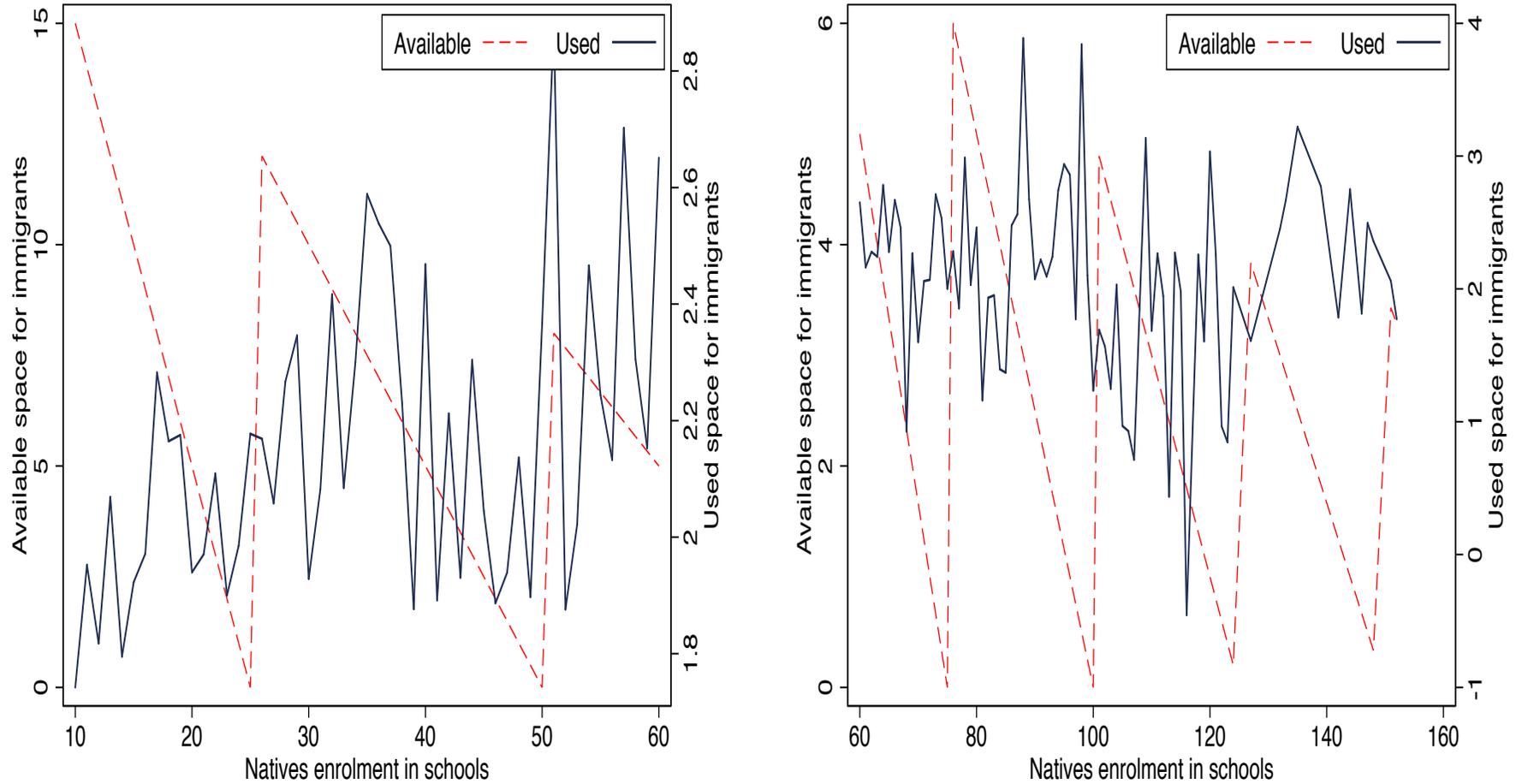
Figure A-2: Available and used space for immigrants according to natives enrolment in schools of grade 2; language sample



Notes: This figure is based on the language estimation sample described in Table 1, for grade 2. The available space for immigrants, described by the dashed line, is defined as the maximum number of students in a class (25) minus the theoretical number of natives in a class predicted by equation (11) according to the rules of class formation as a function of native enrolment at the school \times grade level. The used space for immigrants (the solid line) is the average number of immigrants in a class by level of native enrolment net of institution fixed effects. More precisely, we regress the number of immigrants in a class on a set of dummies for each level of native enrolment, absorbing the institutions fixed effects. The solid line plots the coefficients of the native enrolment dummies in these regressions, rescaled by adding the constant. The left scale reports the range of variation of the available space for immigrants. The right scale refers to the used space for immigrants.

Figure A-3: Available and used space for immigrants according to natives enrolment in schools of grade 5; language sample

13



Notes: This figure is based on the language estimation sample described in Table 1, for grade 5. The available space for immigrants, described by the dashed line, is defined as the maximum number of students in a class (25) minus the theoretical number of natives in a class predicted by equation (11) according to the rules of class formation as a function of native enrolment at the school \times grade level. The used space for immigrants (the solid line) is the average number of immigrants in a class by level of native enrolment net of institution fixed effects. More precisely, we regress the number of immigrants in a class on a set of dummies for each level of native enrolment, absorbing the institutions fixed effects. The solid line plots the coefficients of the native enrolment dummies in these regressions, rescaled by adding the constant. The left scale reports the range of variation of the available space for immigrants. The right scale refers to the used space for immigrants.

Table A-2: First Stage for the number natives and immigrants; language sample

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

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

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

	Baseline		Classes with cheating indicator = 0		Externally monitored institutions	
	Natives (1)	Immigrants (2)	Natives (3)	Immigrants (4)	Natives (5)	Immigrants (6)
1($2 \leq C_{sg}^N < 3$)	-9.8401*** (0.289)	4.6018*** (0.357)	-9.8554*** (0.301)	4.6955*** (0.383)		
1($3 \leq C_{sg}^N < 4$)	-10.9353*** (0.752)	6.5424*** (0.321)	-10.9615*** (0.749)	6.5554*** (0.321)		
1($4 \leq C_{sg}^N < 5$)	-8.6161*** (0.348)	3.9571*** (0.180)	-8.6632*** (0.352)	3.9563*** (0.180)	-10.0269*** (0.660)	3.6548*** (0.451)
1($5 \leq C_{sg}^N < 6$)	-9.1070*** (0.527)	5.2356*** (0.961)	-9.1793*** (0.565)	5.4499*** (1.040)	-10.6714*** (0.948)	4.8695** (2.403)
1($6 \leq C_{sg}^N < 7$)	-7.5338*** (0.685)	4.0760*** (0.764)	-7.5566*** (0.700)	4.0981*** (0.802)	-9.4911*** (0.628)	3.8589*** (1.045)
1($7 \leq C_{sg}^N < 8$)	-7.5887*** (0.395)	2.2712*** (0.454)	-7.6561*** (0.403)	2.2477*** (0.464)	-8.4602*** (0.717)	2.7099*** (0.883)
1($8 \leq C_{sg}^N < 9$)	-6.5802*** (0.353)	1.2365*** (0.267)	-6.5460*** (0.362)	1.1803*** (0.275)	-7.5287*** (0.608)	0.3313 (0.473)
1($9 \leq C_{sg}^N < 10$)	-5.7260*** (0.331)	0.4008 (0.248)	-5.7336*** (0.335)	0.4149 (0.254)	-6.5653*** (0.602)	0.9445 (0.586)
1($10 \leq C_{sg}^N < 11$)	-4.8709*** (0.318)	-0.3461* (0.202)	-4.8807*** (0.320)	-0.3410* (0.204)	-5.7939*** (0.570)	-0.5055 (0.413)
1($11 \leq C_{sg}^N < 12$)	-3.9728*** (0.316)	-0.2173 (0.201)	-4.0060*** (0.319)	-0.2048 (0.202)	-4.2706*** (0.607)	-0.7077* (0.418)
1($12 \leq C_{sg}^N < 13$)	-2.7097*** (0.312)	-0.3803* (0.196)	-2.7596*** (0.314)	-0.3986** (0.197)	-3.2830*** (0.592)	-0.7245** (0.364)
1($13 \leq C_{sg}^N < 14$)	-1.7310*** (0.301)	-0.1541 (0.168)	-1.7935*** (0.302)	-0.1343 (0.169)	-2.3612*** (0.568)	-0.3837 (0.354)
1($14 \leq C_{sg}^N < 15$)	-1.1977*** (0.292)	-0.2324 (0.164)	-1.2306*** (0.293)	-0.2245 (0.164)	-1.8125*** (0.553)	-0.4806 (0.364)
1($15 \leq C_{sg}^N < 16$)	-0.3421 (0.296)	-0.2769 (0.169)	-0.3580 (0.297)	-0.3028* (0.170)	-1.3786** (0.558)	-0.6094* (0.361)
1($16 \leq C_{sg}^N < 17$)	0.3862 (0.289)	-0.3320** (0.163)	0.3480 (0.291)	-0.3291** (0.163)	-0.3708 (0.550)	-0.2110 (0.337)
1($17 \leq C_{sg}^N < 18$)	1.2305*** (0.286)	-0.0964 (0.158)	1.1852*** (0.286)	-0.0845 (0.157)	0.6734 (0.526)	-0.2468 (0.328)
1($18 \leq C_{sg}^N < 19$)	1.8816*** (0.292)	-0.1011 (0.161)	1.8660*** (0.293)	-0.1098 (0.160)	1.4049*** (0.532)	-0.2210 (0.329)
1($19 \leq C_{sg}^N < 20$)	2.4841*** (0.291)	-0.2432 (0.156)	2.4440*** (0.291)	-0.2467 (0.156)	1.3253** (0.548)	-0.2480 (0.327)
1($20 \leq C_{sg}^N < 21$)	2.8728*** (0.298)	-0.2558 (0.159)	2.8607*** (0.298)	-0.2754* (0.158)	1.9106*** (0.563)	-0.3368 (0.319)
1($21 \leq C_{sg}^N < 22$)	3.3136*** (0.312)	-0.4454*** (0.164)	3.2350*** (0.313)	-0.4061** (0.163)	3.2753*** (0.579)	-0.8437** (0.340)
1($22 \leq C_{sg}^N < 23$)	2.8697*** (0.323)	-0.3806** (0.170)	2.8202*** (0.325)	-0.3703** (0.169)	2.3477*** (0.598)	-0.4726 (0.349)
1($23 \leq C_{sg}^N < 24$)	2.5866*** (0.347)	-0.5195*** (0.171)	2.5487*** (0.347)	-0.5206*** (0.171)	1.3195** (0.622)	-0.5474 (0.335)
1($24 \leq C_{sg}^N < 25$)	2.0317*** (0.344)	-0.4925*** (0.178)	2.0309*** (0.346)	-0.5056*** (0.177)	1.4960** (0.622)	-0.5141 (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 (11) 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 (1)	Immigrants (2)	Natives (3)	Immigrants (4)	Natives (5)	Immigrants (6)
1($3 \leq C_{sg}^N < 4$)	-11.0241*** (0.753)	6.6141*** (0.314)	-11.0418*** (0.752)	6.6166*** (0.309)	- (-)	- (-)
1($4 \leq C_{sg}^N < 5$)	-8.7855*** (0.552)	4.5727*** (0.548)	-8.8284*** (0.569)	4.5927*** (0.564)	-9.9656*** (0.664)	3.6605*** (0.439)
1($5 \leq C_{sg}^N < 6$)	-9.0098*** (0.489)	5.1861*** (0.974)	-9.0875*** (0.521)	5.4091*** (1.050)	-10.9806*** (1.062)	4.9772** (2.498)
1($6 \leq C_{sg}^N < 7$)	-7.6271*** (0.694)	4.0751*** (0.771)	-7.6551*** (0.709)	4.0973*** (0.809)	-9.6454*** (0.614)	3.9038*** (1.027)
1($7 \leq C_{sg}^N < 8$)	-7.5669*** (0.403)	2.2780*** (0.453)	-7.6269*** (0.412)	2.2471*** (0.462)	-8.5560*** (0.729)	2.7637*** (0.892)
1($8 \leq C_{sg}^N < 9$)	-6.5663*** (0.353)	1.2422*** (0.267)	-6.5255*** (0.363)	1.1825*** (0.276)	-7.5446*** (0.604)	0.3124 (0.472)
1($9 \leq C_{sg}^N < 10$)	-5.7234*** (0.331)	0.4202* (0.248)	-5.7313*** (0.335)	0.4326* (0.254)	-6.6530*** (0.598)	1.0029* (0.584)
1($10 \leq C_{sg}^N < 11$)	-4.8838*** (0.318)	-0.3284 (0.202)	-4.8911*** (0.320)	-0.3269 (0.204)	-5.8730*** (0.571)	-0.4817 (0.413)
1($11 \leq C_{sg}^N < 12$)	-3.9752*** (0.316)	-0.2095 (0.201)	-4.0049*** (0.319)	-0.1993 (0.202)	-4.3673*** (0.604)	-0.6836* (0.413)
1($12 \leq C_{sg}^N < 13$)	-2.7098*** (0.313)	-0.3674* (0.196)	-2.7541*** (0.315)	-0.3887** (0.197)	-3.3728*** (0.592)	-0.6821* (0.366)
1($13 \leq C_{sg}^N < 14$)	-1.7159*** (0.302)	-0.1458 (0.169)	-1.7766*** (0.302)	-0.1279 (0.169)	-2.3658*** (0.569)	-0.3827 (0.355)
1($14 \leq C_{sg}^N < 15$)	-1.2066*** (0.293)	-0.2120 (0.164)	-1.2366*** (0.295)	-0.2068 (0.165)	-1.9017*** (0.555)	-0.4393 (0.362)
1($15 \leq C_{sg}^N < 16$)	-0.3364 (0.297)	-0.2679 (0.169)	-0.3491 (0.298)	-0.2968* (0.170)	-1.4065** (0.559)	-0.6045* (0.361)
1($16 \leq C_{sg}^N < 17$)	0.3911 (0.290)	-0.3239** (0.163)	0.3545 (0.291)	-0.3235** (0.163)	-0.4344 (0.552)	-0.1805 (0.337)
1($17 \leq C_{sg}^N < 18$)	1.2342*** (0.287)	-0.0854 (0.158)	1.1895*** (0.287)	-0.0743 (0.158)	0.6357 (0.527)	-0.2285 (0.329)
1($18 \leq C_{sg}^N < 19$)	1.8924*** (0.293)	-0.0963 (0.161)	1.8785*** (0.293)	-0.1061 (0.160)	1.3464** (0.532)	-0.2014 (0.330)
1($19 \leq C_{sg}^N < 20$)	2.4962*** (0.291)	-0.2395 (0.156)	2.4578*** (0.292)	-0.2423 (0.156)	1.2757** (0.550)	-0.2215 (0.328)
1($20 \leq C_{sg}^N < 21$)	2.8703*** (0.298)	-0.2566 (0.159)	2.8590*** (0.299)	-0.2783* (0.158)	1.8506*** (0.564)	-0.3131 (0.319)
1($21 \leq C_{sg}^N < 22$)	3.3185*** (0.313)	-0.4386*** (0.164)	3.2403*** (0.314)	-0.4022** (0.163)	3.2114*** (0.582)	-0.8193** (0.341)
1($22 \leq C_{sg}^N < 23$)	2.8862*** (0.324)	-0.3821** (0.170)	2.8386*** (0.325)	-0.3740** (0.169)	2.2879*** (0.599)	-0.4484 (0.349)
1($23 \leq C_{sg}^N < 24$)	2.6008*** (0.347)	-0.5172*** (0.171)	2.5641*** (0.348)	-0.5196*** (0.171)	1.2856** (0.622)	-0.5412 (0.335)
1($24 \leq C_{sg}^N < 25$)	2.0357*** (0.345)	-0.4861*** (0.178)	2.0361*** (0.347)	-0.5012*** (0.177)	1.4358** (0.625)	-0.4876 (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 (11) 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%.