PRELIMINARY AND INCOMPLETE

In a Small Moment:
Cheating and Class Size in Italian Primary Schools*

Joshua D. Angrist†, Erich Battistin‡and Daniela Vuri§

September 2013

Abstract

Using a Maimonides-Rule identification strategy based on class-size cutoffs around 25, we document a payoff to smaller classes in Italian primary schools. These gains are driven mainly by schools in Southern Italy, suggesting a substantial return to class size for relatively poor and less educated residents of the Mezzogiorno. In addition to low SES, however, the Mezzogiorno is distinguished by pervasive teacher cheating on standardized tests, a fact established by an experiment randomly assigning school monitors. We use Italy's Maimonides Rule to show that small classes facilitate teacher cheating, providing an alternative explanation for the causal effects of class size on test scores in Southern Italy. This motivates a causal model for achievement with two endogenous variables, class size and proportion cheating. The model is identified by a combination of class size cutoffs and randomly assigned classroom monitors. The resulting estimates suggest that the effects of class size on measured achievement in Italian primary schools are driven entirely by the relationship between class size and teacher cheating. Models that estimate class size effects jointly with cheating effects generate precise zeros for the former.

*Special thanks go to Patrizia Falzetti, Roberto Ricci and Paolo Sestito at INVALSI for providing the achievement data used here and to INVALSI staffers Paola Giangiacomo and Valeria Tortora for advice and guidance in our work with the data. Grateful thanks also go to Gianna Barbieri, Angela Iadecola and Daniela Di Ascenzo at the Ministry of Education (MIUR) for access to and assistance with administrative data. Chiara Perricone provided expert research assistance. This research is supported by the Einaudi Institute of Economics and Finance (EIEF) - Research Grant 2011 and by the Fondazione per la Scuola della Compagnia di San Paolo in Turin. The views expressed here are those of the authors alone.

†MIT Department of Economics and NBER

‡Department of Economics, University of Padua, IRVAPP and IZA

§Department of Economics and Finance, University of Rome Tor Vergata

1 Introduction

Efforts at school improvement often focus on inputs to education production, the most important of which are staffing ratios and class size. Parents, teachers, and policy makers look to small classes to boost achievement and human capital. The question of whether changes in class size have a causal effect on achievement remains controversial, however. Regression estimates often show little gain to class size reductions, with students in larger classes sometimes appearing to do better as a result (Hanushek, 1995). At the same time, a large randomized study, the Tennessee STAR experiment, generated evidence of substantial learning gains in smaller classes (Krueger, 1999). More recently, an investigation of longer-term effects of random assignment to small elementary school classrooms in the STAR experiment shows increases in college attendance for treated students (Chetty et al., 2011).

Randomized evaluations of changes in class size are exceedingly rare. Researchers have therefore turned to quasi-experimental designs in the hope of capturing class size causal effects without the benefit of random assignment. An influential design pioneered by Angrist and Lavy (1999) and Hoxby (2000) exploits the variation in class size generated by rules for classroom assignment in a regime with class size caps. Angrist and Lavy call this source of variation "Maimonides' Rule," after the 12th century Rabbinic sage who proposed a class size cap for bible study. In contemporary Italy, the setting for our study, Maimonides' Rule applies with a cap of 25 (recently increased to 27). Division of classes at integer multiples of such a cutoff generates discontinuities and changes in slope in the relationship between class size and grade enrollment, variation that potentially identify causal effects. Using Maimonides' Rule to estimate the effect of class size on achievement in Israeli elementary schools, Angrist and Lavy (1999) found returns to class on the order of those generated by the STAR experiment, though Hoxby (2000)'s estimates for Connecticut are not significantly different from zero.

Our investigation of class size effects in Italy begins with Maimonides-style estimates for the population of second and fifth graders, most of whom attend much smaller classes than those seen in Israel, where classes as large as 35 are common. The estimates for Italy suggest a modest return to small classes, with a 10 student reduction in class size estimated to boost scores by about $.05\sigma$. Importantly, however, our estimated returns to class reductions are at least 3 times larger in Southern Italy than for the rest of the country; estimates for Northern and Central Italy are small and only marginally significantly different from zero, at best. This paper develops and then substantiate a simple hypothesis for these large regional differences in Italian class size effects.

Consistent with unified Italy's emergence from a diverse collection of small states in the 19th century, modern Italy is marked by wide-ranging regional differences. The South, known as the Mezzogiorno, is distinguished by persistently higher unemployment, lower per-capita income, higher crime rates, and lower educational attainment than is characteristic of North and Central Italy.[1] The Mezzogiorno also lags the rest of Italy in financial development (Guiso et al., 2004), political accountability (Nannicini et al., 2013), and workplace productivity (Ichino and Maggi, 2000). Italy's North-South divide, which is larger and more persistent than that seen across America's Mason-Dixon line, has been linked to cultural differences and differences in residents' view of the role of government (Putnam et al., 1993).

Against a backdrop of stubborn under-development, the Mezzogiorno is also distinguished by widespread cheating on the standardized tests given in Italian primary school.s This can be seen in Figure 1, which reproduces regional estimates of cheating prevalence among elementary school students. These estimates come from the Italian Instituto Nazionale per la Valutazione del Sistema Dell'istruzione (INVALSI), a government agency charged with educational assessment. Cheating classes are identified by screening for surprisingly large average scores, low within-class variability, and suspect patterns of missing values.[2] Measured in this way, the proportion of compromised scores is about 5 percent, much as reported for Chicago elementary schools by Jacob and Levitt (2003). In Southern Italy, however, the proportion of compromised exams averages about 13 percent and reaches 25 percent in some provinces. Further evidence on cheating comes from Bertoni et al. (2013), who analyze data generated by the random assignment of classroom monitors. This experimental analysis also uncovers evidence of a substantial regional gradient in cheating.

We argue below that cheating in Italian primary schools reflects teacher behavior and, for institutional reasons, that the cost of teacher cheating is mostly a time cost that increases linearly with class size. We then use Maimonides' Rule to show that in regions where cheating is common, larger classes reduce the number of test scores that appear to have been adulterated. A juxtaposition of regional patterns in the causal effects of class size on achievement and the causal effects of class size on cheating uncovers striking parallels in the two relationships. This parallelism leads us to model achievement as a function of two endogenous variables, class size and cheating. The model is identified by a combination of Maimonides' Rule and randomly assigned classroom monitors. The resulting estimates suggest that the relationship between class size and INVALSI assessments

[1]The Mezzogiorno consists of the administrative regions of Basilicata, Campania, Calabria, Puglia, Abruzzo, Molise, and the islands of Sicily and Sardinia.

[2]The INVALSI testing program is described below and in (INVALSI, 2010). The INVALSI cheating variable identifies classes with substantially anomalous score distributions, coding a probability of cheating for each (see Quintano et al. 2009). Figure 1 uses this variable for the 2009-11 scores of second and fifth graders.

is explained entirely by cheating. Double-endogenous variable models produce precise zeros for non-cheating-related class size effects.

Why is the fact that cheating explains putative returns to class size important? The Maimonides Rule research design is motivated by an effort to quantify causal effects of class size in observational studies, without resorting to costly and time consuming random assignment. The design is not guaranteed to work; Urquiola and Verhoogen (2009) show how endogenous sorting induces selection bias in comparisons across class size caps in Chile. By contrast, our analysis uncovers a substantive problem inherent in analyses of the causal effects of class size, regardless of design. We show that even where class size boosts measured achievement, this need not signal increased human capital. Our findings also provide evidence of a surprisingly large behavioral response to what would seem to be weak incentives. Italian teachers work in a highly regulated public sector, with virtually no risk of termination, and are subject to a pay and promotion structure that's largely independent of performance. Even so, and in spite of the fact that cheating is probably not effortless, large numbers of Italian teachers seem to find cheating worthwhile. The question of why this is, and why the effort to cheat is especially likely to be deemed worthwhile by those who teach in the South, is an intriguing topic for further study.

The rest of the paper is organized as follows. The next section present institutional background on Italian schools and achievement tests, and explains how the tests are scored. Section 3 describes our data and documents the Maimonides first stage. Following a brief graphical analysis, Section 4 reports Maimonides-style estimates of effects of class size on achievement and cheating, while Section 5 summarizes results from the monitoring experiment. Section 6 reports jointly estimated class size and cheating effects. This section also considers the econometric implications of mismeasured and underestimated cheating rates for our empirical strategy. Section 7 concludes.

2 Background

Italian Schools and Tests

Primary and Secondary schooling in Italy are compulsory from 6 to 16 years of age, with three stages: 5 years of elementary school (*scuola elementare*), lower secondary school covering grades 6-8 (*scuola media*), and high school (*scuola superiore*), which runs from 3-5 years. Schools are organized into single- or multi-unit "institutions," much as a single campus might house more than one school in American public systems (a local example is the South Boston Educational Complex).

Italian schools have long used matriculation exams for tracking and placement in the transition

from elementary to middle school and throughout high school, but standardized testing for evaluation purposes is a recent development. In 2008, INVALSI piloted voluntary assessments in elementary school; in 2009 these became compulsory for all schools and students. INVALSI assessments cover mathematics and Italian language skills in a national administration lasting two days in the Spring.[3] Tests are proctored by local administrators and teachers. Proctors and other teachers are expected to copy students' original responses onto machine-readable answer sheets (*scheda risposta*), which are then sent to INVALSI. Copying is not passive; some of this processing requires teachers to interpret a student's original response as being correct, incorrect, or missing. We give examples in a brief appendix with sample test items and score sheets. This procedure opens the door to score manipulation, as does the fact that proctors are local staff, not necessarily pledged to uphold the integrity of their students' results.

Related work

Maimonides'-Rule type empirical strategies have been used to identify class size effects in the US (Hoxby 2000), France (Piketty 2004 and Gary-Bobo and Mahjoub 2006), Norway (Bonesronning 2003 and Leuven et al. 2008)) and the Netherlands (Dobbelsteen et al. 2002). On balance these results point to modest returns, mostly lower than those reported in Angrist and Lavy (1999) for Israel. A natural explanation for the relatively large Israeli findings is the unusually large classes characteristic of Israeli elementary schools. In line with this view, Woessmann (2005)'s finds a mostly weak association between class size and achievement in a cross-country panel covering 17 Western European school systems.

The returns to class size in Italy have received little attention from researchers to date, partly because large representative samples with achievement data have only recently become available. One of the few existing studies, Bratti et al. (2007), reports regression estimates showing an insignificant class size effect. In an aggregate analysis, Brunello and Checchi (2005) look at the relationship between staffing ratios and educational attainment for cohorts born before 1970; they find that higher pupil-teacher ratio at the regional level are associated with higher average schooling attainment.

The monitoring experiment used here to identify the effects of cheating and class size jointly was first analyzed by Bertoni et al. (2013). We start by replicating some of the earlier findings, while also adjusting for potentially confounding features of the monitoring experiment's sample design not fully accounted for in the earlier analysis. Failure to make these adjustments leads to covariate imbalance across treatment and control groups; our analysis corrects this problem. The resulting estimates

[3]For results and additional background, see http://www.invalsi.it)

suggest that classroom monitors sharply reduce cheating, and that cheating boosts measured scores dramatically. Both of these effects are markedly larger in Southern Italy. Cheating is a marginal phenomenon elsewhere.

3 Data and First Stage

Data and descriptive statistics

The asessments used in this study come from INVALSI's testing program in Italian elementary schools, with score data from 2009-11. Test scores are measured as the number of correct answers and, for the purpose of regression and two-stage least squares (2SLS) estimation, standardized by subject, year of survey and grade to have zero mean and unit variance. The INVALSI data on test scores are complemented with administrative and survey information describing institutions, schools, classes, and students. Class size can be measured by administrative enrollment counts as well as the number of test-takers (we use the former). Student data include gender, citizenship, and information on parents' employment status and educational background. Students attending private schools (less than 10 percent of the national student population) are omitted from our study.

The unit of observation in our statistical analysis is the class since this is the level where the regressor of interest varies. The empirical analysis is restricted to classes with more than the minimum number of students set by law (10 before 2010 and 15 from 2011, allowing a 10 percent tolerance). We also drop institutions with fewer than 10 students and more than 160 in a grade, as these are outside the range where Maimonides' Rule is likely to matter (these size cutoffs fall at the 1st and 99th percentiles of the enrollment-weighted class size distribution).

The resulting dataset includes almost 136,000 classes (for a total of almost 1,300,000 students in each grade). Table 1 reports descriptive statistics separately by grade, at the class level in Panels A and B, at the school level in Panel C, and at the institution level in Panel D. For both grades, class size averages about 20, and is slightly lower in the South. The distribution of test scores, shown in Panel B, refers to the class average percent correct. Scores are higher in language than in math and higher in 5th grade than in 2nd grade. The table reports descriptive statistics for INVALSI's class-level probability of cheating and for a similar though slightly more comprehensive cheating variable that we've constructed (Section 3, below, describes our procedure). Descriptive statistics for the two measures are similar, and indeed, the two cheating variables are highly correlated.

Maimonides in Italy

The variation exploited by our identification strategy comes from an enforced national minimum-maximum class-size rule applied in Italy (a consequence of a regulation known as *Decreto Ministeriale 331/98*). Until the 2008-09 school year, primary schools were subject to a minimum class size of 10 students and a maximum of 25 students. If grade enrollment at school exceeds 25 pupils (or multiples of 25), a new class must be added. The rule allows exceptions: principals can reduce the size of classes with one or more disabled students, and schools in mountain or remote areas are allowed to retain classes with fewer than 10 students. Finally, the law allows a 10% tolerance from the maximum in either direction. In 2009/10, a reform of the Italian school system allowed larger classes, now capped at 27, with a minimum size of 15 (*Decreto del Presidente della Repubblica 81/2009*). This reform was rolled in, first grade in 2009-10, first and second grade in 2010-11 and first through third grade in 2011-12. In our data, second graders in 2009-10 and fifth graders in any year are subject to the old rule, while second graders in 2010-11 and 2011-12 follow the new rule.

Maimonides' Rule for Italy is the following non-linear and discontinuous function of enrollment:

$$f_{sc} = {}^{e_s}\!/_{[int((e_s-1)/25)+1]} \tag{1}$$

where e_s is the total enrollment in school s in a given grade at the beginning of the year, f_{sc} is the class size assigned to class c in school s, for that grade, and $int(x)$ is the largest integer smaller or equal to x. This equation derives the theoretical class size as a function of school enrollment and captures discontinuities at *25* (or *27* under the new regime) as derived from the Maimonides' rule. Figures 2 and 3 plot the average class size by enrollment e_s at school level for second and fifth graders, separately before and after the reform, along with the predicted class size function f_{sc}. The figures show that class size increases linearly with enrollment size up to the corners of the class size functions and whenever enrollment reaches a multiple integer of the maximum class size, the average class size drops sharply. Maimonides' Rule does a better job of explaining class size at low levels of enrollment. Moreover, in practice, many classes are split before reaching the theoretical maximum.

In analysis of Chilean urban private voucher schools, Urquiola and Verhoogen (2009) document endogenous enerollment sorting in a regime with class size caps. Schools in their sample can act to move one way or another around the cutoff, and many schools seem to aspire to avoid the opening of a an additional classroom where possible. If, as seems likely in Chile, higher-income households sort into higher-productivity schools, identification strategies based on Maimonides' Rule are compromised. Such sorting seems unlikely in Italy, where public schools students are required to attend local schools and mobility across school districts is rare. Schools appear to react mechanically

to enrollment, and enrollment is observed by families when they register their children for school.[4] We provide empirical evidence supporting this contention below, showing that, in our data, student covariates seem to be largely unrelated to Maimonides' Rule.

Cheating Data and Cheating Behavior

We use a variation on INVALSI's procedure to construct a class-level measure of cheating. Neither cheating measure is perfect, of course. Section 6 discusses the likely impact of cheating misclassification on our empirical strategy.

Our cheating variable is constructed as a function of the within-class average and standard deviation of test scores, the percentage of missing items, and a Herfindahl index of the share of students with similar response patterns at the test. Using these indicators, we flag as suspicious those classes with abnormally high performance, small dispersion of test scores, unusually low proportion missing, and highly concentrated in response patterns. Our procedure is similar to that used by Quintano et al. 2009(DETAILS AND COMP TO JACOB-LEVITT TBD)

Why does class size affect cheating? First, cheating is facilitated by the need for local proctors and teachers to copy students' original answer sheets onto the machine readable *scheda risposta*. This task, which requires those doing the copying to make judgements about which answers are correct or missing, must be completed by the end of the day on which tests are taken. The extent of the necessary interpretation is documented with examples and a copy of the *scheda risposta* in the Appendix. Additionally, In larger classes, teachers may not have time to make as many adjustments to scores.[5] Second, exams are locally proctored. Although it seems unlikely that teachers or proctors announce answers to an entire classroom at once, they may offer individual students assistance. Finally, a large class increases the risk that cheating behavior will be exposed. The fact that monitoring reduces cheating suggests those doing it see exposure as undesireable.

4 Effects of Class size on Achievement and Cheating

Graphical Analysis

Our investigation of class size effects begins by framing the problem in a standard setting, thus ignoring the possibility of confounding effects of cheating on test scores. By using the class as

[4] This is similar to the institutional setting described by Angrist and Lavy (1999).

[5] The test, equal for everybody, is made of multiple choice, true/false or open questions with a univocally correct answer. The presence of open questions in the test implies some form of correction by proctors or teachers and not simply a copying mechanism.

statistical unit of analysis, we start by deriving descriptive evidence that is suggestive of sizable effects in the Mezzogiorno. The graphical analysis presented in Figure 4 is employed to document the informational content of the data, and is replicated for the various outcomes considered. Dots in the figure are obtained according to the following procedure. After taking deviations of the running variable - grade enrollment - from class size cutoffs, we keep only observations in the [-12,12] window. For each point in this bounded support, we compute averages of the variable considered on the vertical axis – e.g. class size, in Figure 4 – using only the first four cutoffs in Figure 2 and Figure 3. Finally, we apply a 3-point rolling average to smooth profiles by grade enrollment. To emphasize the non-parametric nature of our exercise, the analysis is carried out separately for second and fifth graders. The additional stratification that results from considering years before and after the increase of the class size cap from 25 to 27 proved redundant, and thus is omitted to ensure compactness in the presentation of results.

Superimposed are local linear regression estimates of outcome profiles, which we compute separately over the [-12,0] and [1,12] intervals using the Epanechnikov kernel and setting the bandwidth to its optimal value according to the criterion employed by Imbens and Kalyanaraman (2012). As explained in Section 3, the administrative rule for the class size formation has one important exception, in that departures from the maximum in either direction (i.e. towards smaller or larger classes) are allowed within a 10 percent limit. These add fuzziness to the Maimonides' rule, which we quantify as plus or minus two students. Because of this, local linear regressions are fitted excluding two observations either side of the cutoff.

The graphical analysis is carried out on class size (Figure 4) and test scores (Figure 5). The discontinuity depicted by the former figure results from the average drop in class size across cutoffs in Figure 2 and Figure 3. Although the discontinuity is less pronounced in the Mezzogiorno and in fifth grade, our results point to a sizable difference in school inputs around the cutoff. Figure 5 is suggestive of how differences in the learning environment affect performance, the top and the bottom panels referring to Math and language, respectively. Figure 7 rules out the possibility that the sharp effects on test scores solely reflect compositional differences that may result from sorting around the class size cap.

Empirical Framework

Moving from the non-parametric analysis above, we investigate the statistical relevance of the effects documented by framing the problem in a regression setting. Let y_i be the outcome of interest (e.g. test scores), r_i the running variable and s_i class size, all quantities varying across classes indexed

by i. We start by considering the following specification that pools observations across grades and years:

$$y_i = \beta_0 + \beta_1 s_i + \beta_2 r_i + \beta_3 r_i^2 + u_i \tag{2}$$

and where the running variable enters only through a quadratic polynomial (alternatives to this choice for the order of the polynomial convey the same result). The model specification also includes a full set of grade, year and area dummies, as well as the array of demographics presented in Table 2. Throughout the analysis standard errors are clustered by institution, which we reckon the most cautionary choice for the context at hand. The equation is first estimated using OLS. Endogeneity of class size is addressed using 2SLS. We employ as instrument class size predicted from the Maimonides' rule, which we denote by $\hat{s_i}$ and builds in the policy reform that changed the cap from 25 to 27. We also present the reduced form counterpart of equation 2, which is obtained by replacing s_i with $\hat{s_i}$. The relevant first stage regressions for class size are straightforwardly obtained using the same model specification.

We investigated the sensitivity of our inferential conclusions to variants of this specification that allow for a more flexible running variable control. In particular, we define non-overlapping segments centered at each cutoff and spanning the support of the running variable up to the 99^{th}percentile. As before, the definitions employed take into account exogenous changes to the class size cap. We then saturate the model by adding a full set of segment dummies, and allowing the running variable polynomial to vary across segments. The OLS and 2SLS analyses are replicated using this alternative specification.

Results

Estimates of class size effects are reported in Table 2, and are robust with respect to the specification adopted. The same analysis conducted in test scores is replicated using cheating as outcome variable. Results are presented in Figure 6 and Table 3.

5 The Monitoring Experiment

In an effort to increase test reliability, INVALSI randomly selects institutions to be monitored by an external examiner. Classes to be monitored are also selected randomly within institution. Examiners are selected by the regional education offices from a pool of mainly retired teachers and principals, who did not work in their assigned school or municipality for at least two years preceding the test. Monitors are assigned randomly through a multi-stage sampling scheme, stratified by region.

Institutions are first selected with probability proportional to grade enrollment; depending on the size of the sampled institution, observers are randomly assigned to one or two classes in grade. Typically, within the same institution, there are classes with and without external monitoring.

Monitors observe both testing and acore sheet transcription. The presence of an external examiner should discourage cheating both during the test, e.g. students copying from one another or teachers communicating the correct answers, or during the transcript, e.g. teachers changing students' answers or filling in missing answers. In classes without the external examiner, the test is administered by teachers of the same school but not of the same class. Therefore it is reasonable to argue that test score results in classes with external observer are not biased by cheating and can be used as benchmark to assess test outcomes in classes without external monitoring (see also INVALSI (2010) which shows that there is no evidence of cheating in classes with monitors).

To assess the validity of the monitoring experiment, we check whether the main observable characteristics are balanced between monitored (treated) and non monitored (control) institutions. We distinguish between administrative data (Panel A) and survey data (Panel B), separately by geographic area. As explained in the Introduction, institutions with external monitoring are selected using a multi-stage sampling that ensures stratification by region and, within regions, is proportional to grade enrollment. Differently from Bertoni et al. (2013), who exploit the assignment of external observers to classes, we prefer to employ a coarser level of analysis that relies on institutions. Although by design classes with external monitoring should result from a random sampling within institutions, the evidence in our data is strongly against this hypothesis and in line with the descriptives in Bertoni et al. (2013). On the contrary, the evidence reported next shows compliance with the assignment mechanism of observers to institutions.

For each variable in Table 4 we run a regression on a dummy for classes in monitored institutions adding sampling strata controls to adjust for the sampling methodology. In our preferred specification, these are represented by a full set of region dummies, and a quadratic polynomial in grade enrollment at institution which we allow to vary across regions. Standard errors are clustered by institution, which is the level of variability defining treatment and control classes. Table 4 presents corroborative evidence in favor of random assignment of observers to institutions. Administrative variables are balanced across groups (as evidenced by the effects in the "Treatment Difference" columns). In particular, the presence of an external examiner at institution does not affect the fraction of test takers either in class, or in school or in institution. The percent of female students is also well balanced between treated and controls. Balancing is also attained for the number of students enrolled in class and grade enrollment at school.

Turning to individual variables, we notice that some covariates are not balanced especially for father education and mother occupation. It's also worth noting that data are less likely to be missing in monitored institutions, likely due to more careful reporting by administrative secretaries.

6 Effects of Class Size and Cheating on Achievement

2SLS results that take into account endogeneity of cheating and class size are reported in Table 7.

Below we consider the implications of possible measurement error in cheating.

Cheating with Misclassification

Consider the following regression:

$$y = s\beta_1 + c^*\beta_2 + \varepsilon, \tag{3}$$

where y is test score, s is class size and c^* is individually-defined true cheating. Let $P_Z \equiv Z(Z'Z)^{-1}Z$ be the $n \times n$ projection matrix defined from the $n \times k$ matrix of instruments Z. Write the 2SLS estimator of β_1 as:

$$(s'P_Z s)^{-1}s'P_Z y = \beta_1 + (s'P_Z s)^{-1}s'P_Z c^*\beta_2, \tag{4}$$

which implies:

$$\hat{\beta}_1(\beta_2) = (s'P_Z s)^{-1}s'P_Z y - (s'P_Z s)^{-1}s'P_Z c^*\beta_2.$$

This is equation (6) in AK (1992). Now allow for misclassification in the cheating indicator c, so that there is $c \neq c^*$ and:

$$c = (1 - \pi_0) + (\pi_0 + \pi_1 - 1)c^* + v,$$

where $\pi_i \equiv P[c = i|c^* = i]$. Throughout we maintain the assumption that π_i are at least equal to *0.5*, which amounts to saying that the cheating indicator is more informative that coin tossing. The measurement error can be written

$$u \equiv c - c^* = (1 - \pi_0) + (\pi_0 + \pi_1 - 2)c^* + v.$$

Consider the feasible counterpart of equation (3), with measured cheating instead of true. The r.h.s. of equation (4) becomes:

$$\beta_1 + (s'P_Z s)^{-1}s'P_Z c\beta_2 - (s'P_Z s)^{-1}s'P_Z u\beta_2.$$

The expressions given above for c and c^* yield:

$$\begin{aligned}\hat{\beta}_1(\beta_2) &= (s'P_Z s)^{-1}s'P_Z y - (s'P_Z s)^{-1}s'P_Z c^*(\pi_0+\pi_1-1)\beta_2 \\ &+ (s'P_Z s)^{-1}s'P_Z c^*(\pi_0+\pi_1-2)\beta_2, \\ &= (s'P_Z s)^{-1}s'P_Z y - (s'P_Z s)^{-1}s'P_Z c^*\beta_2.\end{aligned}$$

The Cheating Coefficient

Let y denote the $n \times 1$ outcome vector, X the $n \times 2$ matrix of endogenous variables and Z the $n \times k$ matrix of instruments. Write (3) and its feasible counterpart in matrix notation:

$$\begin{aligned}y &= X^*\beta + \varepsilon, \\ y &= X\beta - \beta_2 u + \varepsilon,\end{aligned}$$

where:

$$\underset{n\times 2}{X^*} = [s, c^*], \qquad \underset{n\times 2}{X} = [s, c].$$

The bias on the 2SLS estimator from the feasible equation can be written as:

$$\Delta[\hat{\beta}] = -(Z'X)^{-1}Z'u\beta_2.$$

Calculations yield:

$$\Delta[\hat{\beta}] = \begin{bmatrix} 0 \\ -\beta_2 \frac{\pi_0+\pi_1-2}{\pi_0+\pi_1-1} \end{bmatrix},$$

implying that the 2SLS estimator will converge to:

$$\begin{bmatrix} \beta_1 \\ \beta_2 \frac{1}{\pi_0+\pi_1-1} \end{bmatrix},$$

so that the cheating coefficient is upward biased if $\pi_0 + \pi_1 \geq 1$, while the class size coefficient is unaffected.

7 Summary and Directions for Further Work

Table 1: Descriptive statistics

	2 grade			5 grade		
	Italy	**North/Centre**	**South**	**Italy**	**North/Centre**	**South**
	(1)	(2)	(3)	(4)	(5)	(6)
			A. Student characteristics			
female	0.47	0.48	0.47	0.49	0.49	0.49
	(0.13)	(0.12)	(0.14)	(0.11)	(0.11)	(0.11)
behind grade level	0.02	0.02	0.02	0.03	0.04	0.02
	(0.04)	(0.04)	(0.03)	(0.05)	(0.05)	(0.04)
ahead grade level	0.01	0.00	0.02	0.01	0.01	0.03
	(0.03)	(0.02)	(0.04)	(0.04)	(0.02)	(0.05)
immigrant 1st generation	0.03	0.04	0.01	0.05	0.07	0.02
	(0.05)	(0.06)	(0.03)	(0.07)	(0.08)	(0.04)
immigrant 2nd generation	0.07	0.10	0.02	0.05	0.07	0.02
	(0.09)	(0.11)	(0.04)	(0.07)	(0.08)	(0.03)
father HS diploma	0.26	0.27	0.25	0.25	0.26	0.24
	(0.17)	(0.16)	(0.18)	(0.17)	(0.16)	(0.17)
mother employed	0.45	0.54	0.31	0.45	0.53	0.31
	(0.27)	(0.26)	(0.22)	(0.26)	(0.26)	(0.21)
N	1,294,423	822,632	471,791	1,279,466	793,380	486,086
			B. Class characteristics			
pct correct answers in math	47.35	45.59	50.37	64.61	63.65	66.17
	(14.39)	(12.71)	(16.46)	(11.43)	(8.58)	(14.80)
pct correct answers in language	69.86	69.28	70.86	69.75	69.95	69.44
	(10.88)	(9.15)	(13.30)	(9.37)	(7.30)	(11.97)
class size	20.11	20.24	19.88	20.13	20.34	19.79
	(3.39)	(3.34)	(3.46)	(3.36)	(3.33)	(3.39)
instruction time <= 30 hours p/w	0.61	0.49	0.81	0.63	0.49	0.85
	(0.49)	(0.50)	(0.39)	(0.48)	(0.50)	(0.36)
instruction time >30 hours p/w	0.36	0.48	0.14	0.34	0.48	0.10
	(0.48)	(0.50)	(0.34)	(0.47)	(0.50)	(0.30)
cheating in math	0.06	0.02	0.14	0.06	0.02	0.13
	(0.24)	(0.13)	(0.35)	(0.24)	(0.14)	(0.34)
cheating in language	0.05	0.02	0.11	0.05	0.02	0.11
	(0.23)	(0.15)	(0.31)	(0.23)	(0.15)	(0.31)
cheating (INVALSI) in math	0.07	0.02	0.14	0.06	0.03	0.12
	(0.20)	(0.10)	(0.28)	(0.18)	(0.10)	(0.26)
cheating (INVALSI) in language	0.06	0.03	0.11	0.06	0.03	0.10
	(0.17)	(0.10)	(0.24)	(0.17)	(0.11)	(0.24)
N	67,942	42,994	24,948	67,854	41,988	25,866
			C. School characteristics			
number of classes	1.95	1.87	2.1	1.96	1.86	2.14
	(1.11)	(1.00)	(1.26)	(1.12)	(0.99)	(1.29)
enrolment	40.41	38.77	43.59	40.88	39.02	44.35
	(25.08)	(22.92)	(28.54)	(25.16)	(22.68)	(28.90)
N	34889	23008	11881	34600	22531	12069

Table I. Descriptive statistics (cont.)

	2 grade			5 grade		
	Italy (1)	**North/Centre** (2)	**South** (3)	**Italy** (4)	**North/Centre** (5)	**South** (6)
			D. Institution characteristics			
number of schools						
number of classes						
enrolment						
% with external observer in 2009/10						
% with external observer in 2010/11						
% with external observer in 2011/12						
% classes in 2009/10						
% classes in 2010/11						
% classes in 2011/12						
N						

"Mean" and "s.d." for student characteristics are computed using class averages; "Mean" and "s.d." for class characteristics are computed using one observation per class; "Mean" and "s.d." for school characteristics are computed using one observation per school; "Mean" and "s.d." for institutions are computed using one observation per institution. Categories for educational qualification of father and occupation of mother do not add up to one because of missing values.

Table 2: OLS and IV/2SLS estimates of the effect of class size on scores

	OLS			IV/2SLS					
	Italy	**North/Centre**	**South**	**Italy**	**North/Centre**	**South**	**Italy**	**North/Centre**	**South**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
					A. math				
Class size	-0.0012	-0.0024***	0.0000	-0.0053***	-0.0042***	-0.0114***	-0.0062***	-0.0037**	-0.0144***
	(0.0007)	(0.0007)	(0.0015)	(0.0014)	(0.0012)	(0.0039)	(0.0020)	(0.0018)	(0.0051)
enrolment	✗	✗	✗	✗	✗	✗	✗	✗	✗
enrolment squared	✗	✗	✗	✗	✗	✗	✗	✗	✗
interactions							✗	✗	✗
N	135,794	84,980	50,814	134,824	84,427	50,397	134,824	84,427	50,397
					B. language				
Class size	0.0003	-0.0016***	0.0025**	-0.0037***	-0.0024**	-0.0080***	-0.0039**	-0.0011	-0.0109***
	(0.0006)	(0.0006)	(0.0012)	(0.0011)	(0.0010)	(0.0031)	(0.0016)	(0.0014)	(0.0040)
enrolment	✗	✗	✗	✗	✗	✗	✗	✗	✗
enrolment squared	✗	✗	✗	✗	✗	✗	✗	✗	✗
interactions							✗	✗	✗
N	135,796	84,982	50,814	134,826	84,429	50,397	134,826	84,429	50,397

Notes: The left hand side panel presents OLS estimates of the effect of class size on score. The right hand side panel presents IV/2SLS estimates of the effect of class size on score using Maimonides' rule as instrument. The unit of observation is the class in all regressions. Robust standard errors, clustered on school and grade, are shown in parentheses. Control variables include: % female students, % students behind and ahead grade level, % immigrants, % fathers at least high school graduate, % unemployed mothers, % mother NILF, grade and year dummies, and dummies for missing values in these variables. All regressions have sampling strata controls (grade enrolment at institution, region dummies and their interactions). * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 3: OLS and IV/2SLS estimates of the effect of class size on cheating

	OLS			IV/2SLS					
	Italy	**North/Centre**	**South**	**Italy**	**North/Centre**	**South**	**Italy**	**North/Centre**	**South**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
					A. math				
Class size	-0.0015***	-0.0008***	-0.0028***	-0.0019***	-0.0004	-0.0059***	-0.0021***	-0.0007	-0.0054***
	(0.0003)	(0.0002)	(0.0006)	(0.0005)	(0.0003)	(0.0015)	(0.0007)	(0.0005)	(0.0020)
enrolment	✗	✗	✗	✗	✗	✗	✗	✗	✗
enrolment squared	✗	✗	✗	✗	✗	✗	✗	✗	✗
interactions							✗	✗	✗
N	135,781	84,974	50,807	134,811	84,421	50,390	134,811	84,421	50,390
					B. language				
Class size	-0.0016***	-0.0011***	-0.0025***	-0.0018***	-0.0008**	-0.0042***	-0.0014**	-0.0003	-0.0040**
	(0.0002)	(0.0002)	(0.0005)	(0.0005)	(0.0003)	(0.0014)	(0.0006)	(0.0005)	(0.0018)
enrolment	✗	✗	✗	✗	✗	✗	✗	✗	✗
enrolment squared	✗	✗	✗	✗	✗	✗	✗	✗	✗
interactions							✗	✗	✗
N	135,790	84,977	50,813	134,820	84,424	50,396	134,820	84,424	50,396

Notes: The left hand side panel presents OLS estimates of the effect of class size on cheating. The right hand side panel presents IV/2SLS estimates of the effect of class size on cheating instrumenting class size with Maimonides' rule. The unit of observation is the class in all regressions. Robust standard errors, clustered on school and grade, are shown in parentheses. Control variables include: % female students, % students behind and ahead grade level, % immigrants, % fathers at least high school graduate, % unemployed mothers, % mother NILF, grade and year dummies, and dummies for missing values in these variables. All regressions have sampling strata controls (grade enrolment at institution, region dummies and their interactions). * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 4: Covariate balance in the cheating experiment

	Italy		North/Centre		South	
	Control Mean	Treatment Difference	Control Mean	Treatment Difference	Control Mean	Treatment Difference
	(1)	(2)	(3)	(4)	(5)	(6)
			A. Administrative data			
Class size	20.0474	0.0455	20.2485	0.0274	19.7202	0.0754
	[3.372]	(0.0295)	[3.3184]	(0.0364)	[3.4323]	(0.0501)
Grade enrolment at school	54.0695	-0.3640	50.6014	-0.5005	59.7131	-0.1213
	[30.5072]	(0.3322)	[27.4249]	(0.3959)	[34.2033]	(0.5956)
% in class sitting the test	0.942	0.0001	0.937	0.0008	0.9502	-0.0010
	[0.0608]	(0.0005)	[0.0627]	(0.0006)	[0.0566]	(0.0007)
% in school sitting the test	0.9404	-0.0000	0.9355	0.0007	0.9482	-0.0011
	[0.0512]	(0.0005)	[0.0528]	(0.0006)	[0.0473]	(0.0008)
% in institution sitting the test	0.9382	-0.0002	0.9332	0.0005	0.9462	-0.0013*
	[0.0429]	(0.0004)	[0.0423]	(0.0005)	[0.0426]	(0.0007)
% female students	0.4811	0.0012	0.4831	0.0004	0.4779	0.0026
	[0.1204]	(0.0009)	[0.117]	(0.0011)	[0.1257]	(0.0017)
% students in urban areas						
			B. Survey data			
instruction time: up to 30 hours per week	0.6196	0.0093**	0.4878	0.0068	0.8342	0.0138**
	[0.4855]	(0.0041)	[0.4999]	(0.0054)	[0.3719]	(0.0059)
instruction time: more than 30 hours per week	0.3412	0.0038	0.4834	0.0033	0.1098	0.0045
	[0.4741]	(0.0039)	[0.4997]	(0.0054)	[0.3127]	(0.0051)
behind grade level	0.0255	0.0007**	0.0299	0.0008*	0.0183	0.0001
	[0.0437]	(0.0003)	[0.0467]	(0.0005)	[0.0374]	(0.0003)
ahead grade level	0.0123	0.0001	0.0046	0.0001	0.0248	0.0001
	[0.0343]	(0.0003)	[0.0171]	(0.0002)	[0.0486]	(0.0006)
immigrant 1st generation	0.0405	-0.0004	0.0562	-0.0011	0.0148	0.0005
	[0.0639]	(0.0005)	[0.0715]	(0.0007)	[0.0367]	(0.0005)

Table 4: Covariate balance in the cheating experiment (cont.)

	Italy		North/Centre		South	
	Control Mean	Treatment Difference	Control Mean	Treatment Difference	Control Mean	Treatment Difference
	(1)	(2)	(3)	(4)	(5)	(6)
immigrant 2nd generation	0.0968	0.0007	0.1375	0.0001	0.0304	0.0017**
	[0.1194]	(0.0010)	[0.1295]	(0.0015)	[0.055]	(0.0007)
father HS drop-out	0.4275	0.0147***	0.4072	0.0117***	0.4604	0.0200***
	[0.258]	(0.0025)	[0.2433]	(0.0029)	[0.2771]	(0.0046)
mother employed	0.4428	0.0088***	0.5326	0.0073**	0.2966	0.0115***
	[0.2668]	(0.0024)	[0.2583]	(0.0032)	[0.2098]	(0.0036)
			Survey Non responses			
Missing data on instruction time	0.0392	-0.0131***	0.0288	-0.0101***	0.056	-0.0183***
	[0.1940]	(0.0016)	[0.1673]	(0.0017)	[0.2300]	(0.0033)
Missing data on students behind/ahead grade level	0.0162	-0.0030***	0.0126	-0.0017*	0.0221	-0.0053***
	[0.1133]	(0.0009)	[0.0996]	(0.0010)	[0.1324]	(0.0019)
Missing data on father's education	0.2234	-0.0218***	0.2255	-0.0191***	0.22	-0.0266***
	[0.3415]	(0.0035)	[0.3407]	(0.0044)	[0.3427]	(0.0057)
Missing data on mother's occupation	0.1953	-0.0169***	0.1967	-0.0090**	0.1929	-0.0309***
	[0.3284]	(0.0034)	[0.3258]	(0.0043)	[0.3326]	(0.0055)
Missing data on country of origin	0.0331	-0.0115***	0.0257	-0.0083***	0.0451	-0.0171***
	[0.1644]	(0.0013)	[0.1439]	(0.0014)	[0.1925]	(0.0027)
N	135,796		84,982		50,814	

Notes: “Control Mean” columns report averages and standard deviations for variables in the left-most column. “Treatment Difference” columns report coefficients from regressions of each variable in the left-most column on a treatment dummy, grade and year dummies and sampling strata controls (grade enrolment at institution, region dummies and their interactions). Standard deviations for the control group are in square brackets, robust standard errors are in parentheses, and p values for F tests are in curly braces. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 5: First stage and reduced form observer effects

	First stage			Reduced form		
	Italy	**North/Centre**	**South**	**Italy**	**North/Centre**	**South**
	(1)	(2)	(3)	(4)	(5)	(6)
			A. math			
Observer at institution	-0.0297***	-0.0107***	-0.0629***	-0.1137***	-0.0765***	-0.1826***
	(0.0018)	(0.0012)	(0.0044)	(0.0056)	(0.0050)	(0.0125)
Means	*0.064*	*0.020*	*0.137*	*0.005*	*-0.077*	*0.142*
(sd)	*(0.244)*	*(0.138)*	*(0.344)*	*(0.636)*	*(0.501)*	*(0.794)*
N	135,781	84,974	50,807	135,794	84,980	50,814
			B. language			
Observer at institution	-0.0250***	-0.0122***	-0.0477***	-0.0813***	-0.0542***	-0.1313***
	(0.0016)	(0.0012)	(0.0039)	(0.0044)	(0.0041)	(0.0096)
Means	*0.055*	*0.022*	*0.109*	*0.009*	*-0.007*	*0.035*
(sd)	*(0.227)*	*(0.147)*	*(0.312)*	*(0.521)*	*(0.428)*	*(0.648)*
N	135,790	84,977	50,813	135,796	84,982	50,814

Notes: The left hand side panel presents first stage effects of cheating on observer at institution. The right hand side panel presents RF effects of score on observer at institution. The specification employs a quadratic polynomial in grade enrolment, segment dummies and their interactions. The unit of observation is the class in all regressions. Robust standard errors, clustered on school and grade, are shown in parentheses. Control variables include: % female students, % students behind and ahead grade level, % immigrants, % fathers at least high school graduate, % unemployed mothers, % mother NILF grade and year dummies, and dummies for missing values in these variables. All regressions have sampling strata controls (grade enrolment at institution, region dummies and their interactions). * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 6: Twin first stages

	A. cheating					
	math			language		
	Italy	**North/Centre**	**South**	**Italy**	**North/Centre**	**South**
	(1)	(2)	(3)	(4)	(5)	(6)
Maimonides' rule	-0.0010***	-0.0004	-0.0022**	-0.0007**	-0.0002	-0.0017**
	(0.0004)	(0.0003)	(0.0009)	(0.0003)	(0.0003)	(0.0008)
Observer at institution	-0.0294***	-0.0105***	-0.0624***	-0.0245***	-0.0118***	-0.0469***
	(0.0018)	(0.0012)	(0.0044)	(0.0016)	(0.0013)	(0.0039)
N	134,811	84,421	50,390	134,820	84,424	50,396

	B. class size		
	Italy	**North/Centre**	**South**
	(1)	(2)	(3)
Maimonides' rule	0.5035***	0.5343***	0.4416***
	(0.0061)	(0.0078)	(0.0101)
Observer at institution	0.0241	0.0374	0.0104
	(0.0236)	(0.0267)	(0.0442)
N	134,826	84,429	50,397

Notes: This table shows the first stage regressions for cheating and class size on Maimonides' rule and observer at institution. The specification employs a quadratic polynomial in grade enrolment, segment dummies and their interactions. The unit of observation is the class in all regressions. Robust standard errors, clustered on school and grade, are shown in parentheses. Control variables include: % female students, % students behind and ahead grade level, % immigrants, % fathers at least high school graduate, % unemployed mothers, % mother NILF grade and year dummies, and dummies for missing values in these variables. All regressions have sampling strata controls (grade enrolment at institution, region dummies and their interactions). * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 7: OLS and IV/ 2SLS estimates of the effect of class size and cheating on scores

	OLS			IV/2SLS			IV/2SLS (overidentified)		
	Italy	**North/Centre**	**South**	**Italy**	**North/Centre**	**South**	**Italy**	**North/Centre**	**South**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
				A. math					
Class size	0.0015**	-0.0002	0.0036***	0.0018	0.0012	0.0013	0.0010	0.0010	0.0005
	(0.0006)	(0.0007)	(0.0012)	(0.0022)	(0.0031)	(0.0044)	(0.0020)	(0.0029)	(0.0039)
Cheating	1.4165***	1.4082***	1.4156***	3.8379***	7.2466***	2.9031***	3.8101***	7.2323***	2.8950***
	(0.0058)	(0.0088)	(0.0068)	(0.1914)	(0.7679)	(0.1617)	(0.1900)	(0.7663)	(0.1602)
Overid test [P-value]							[0.3820]	[0.7282]	[0.4301]
N	135,790	84,977	50,813	134,820	84,424	50,396	134,811	84,421	50,390
				B. language					
Class size	0.0027***	0.0003	0.0056***	0.0009	0.0002	0.0002	0.0019	0.0006	0.0031
	(0.0005)	(0.0006)	(0.0010)	(0.0018)	(0.0021)	(0.0038)	(0.0016)	(0.0018)	(0.0034)
Cheating	1.1807***	1.0896***	1.2127***	3.2746***	4.5242***	2.7603***	3.2391***	4.2512***	2.7629***
	(0.0046)	(0.0069)	(0.0058)	(0.1842)	(0.4604)	(0.1804)	(0.1823)	(0.4229)	(0.1801)
Overid test (P-value)							[0.2357]	[0.1726]	[0.3547]
N	135,790	84,977	50,813	134,820	84,424	50,396	134,820	84,424	50,396

Notes: The left hand side panel presents OLS estimates of the effect of scores on class size and cheating. The center panel presents IV/2SLS estimates instrumenting class size and cheating with Maimonide's rule and observer at institution.The right hand side panel presents overidentified IV/2SLS estimates which include also a dummy for grade enrolment being in a 10 percent window below and above each cutoff as instrument. The specification employs a quadratic polynomial in grade enrolment, segment dummies and their interactions. The unit of observation is the class in all regressions. Robust standard errors, clustered on school and grade, are shown in parentheses. Control variables include: % female students, % students behind and ahead grade level, % immigrants, % fathers at least high school graduate, % unemployed mothers, % mother NILF, grade and year dummies, and dummies for missing values in these variables. All regressions have sampling strata controls (grade enrolment at institution, region dummies and their interactions). * significant at 10%; ** significant at 5%; *** significant at 1%.

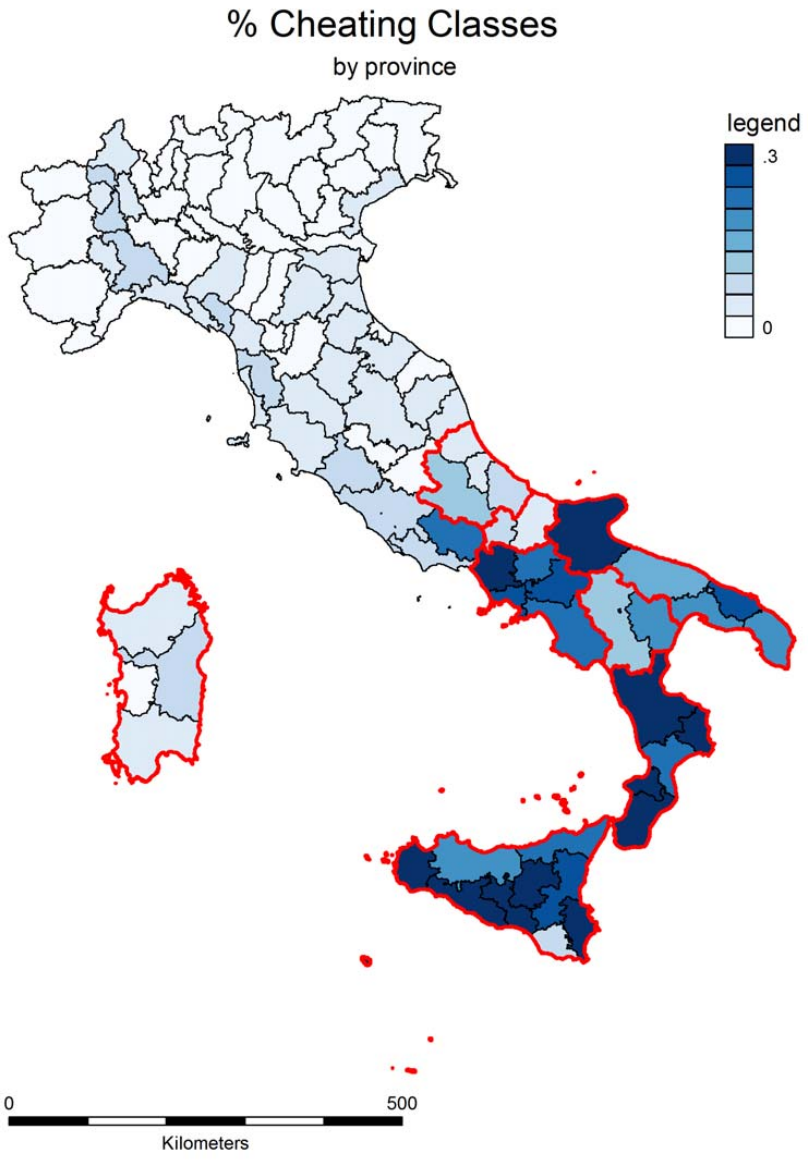
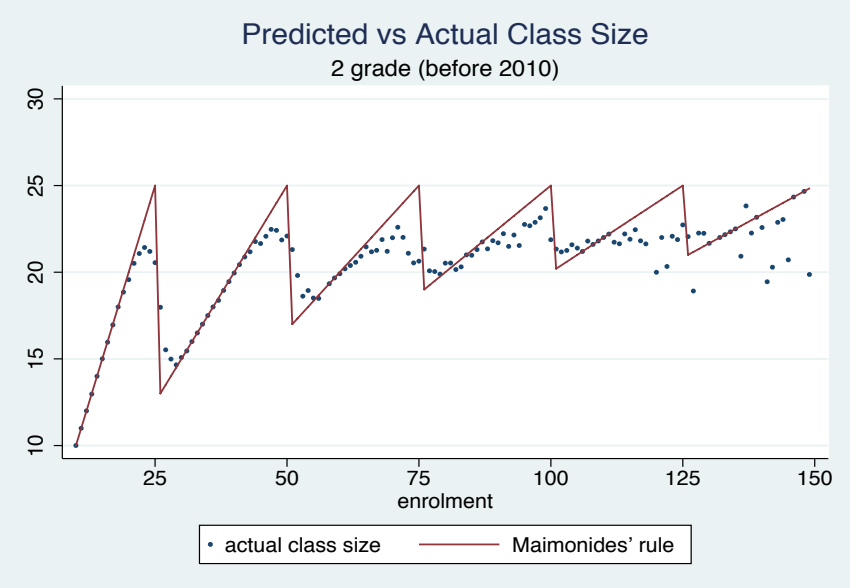


Figure 1: Cheating map



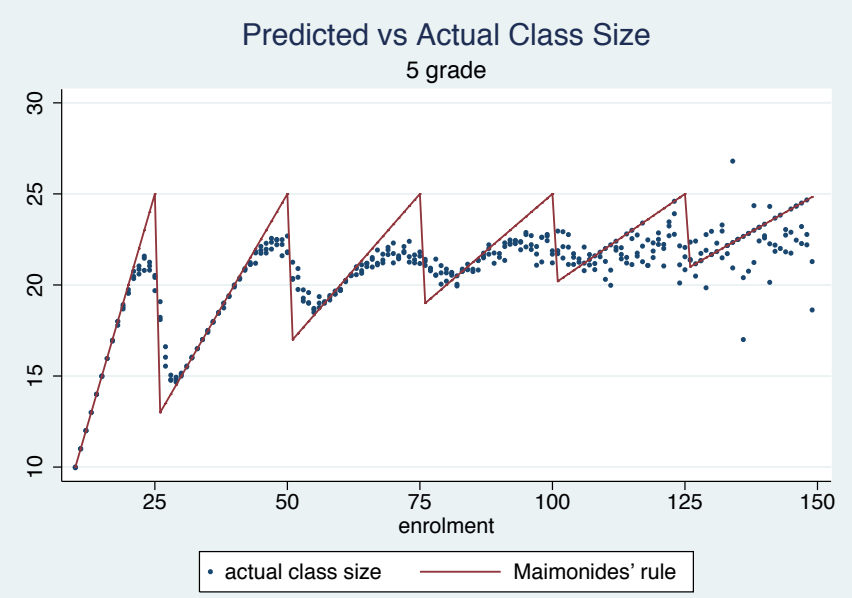


Figure 2: Class size by enrollment, actual class size and as predicted by the Maimonides' rule for 2 and 5 grades in non-reform years

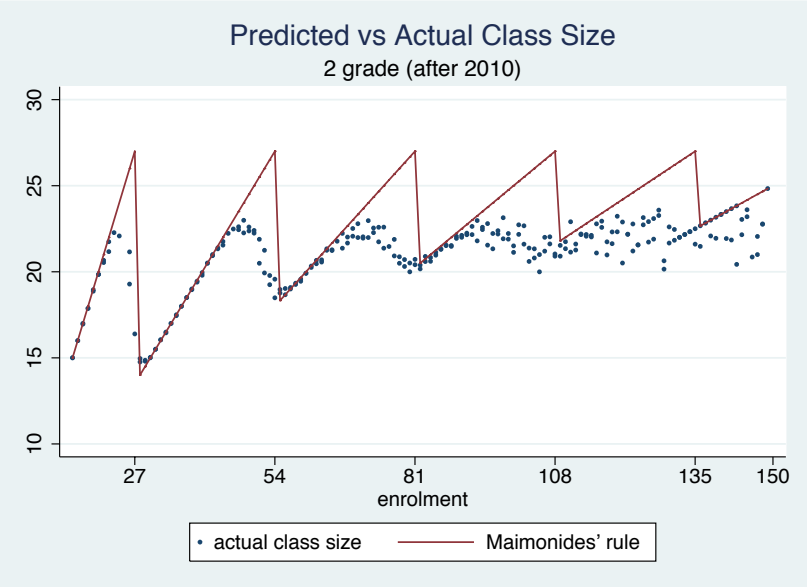
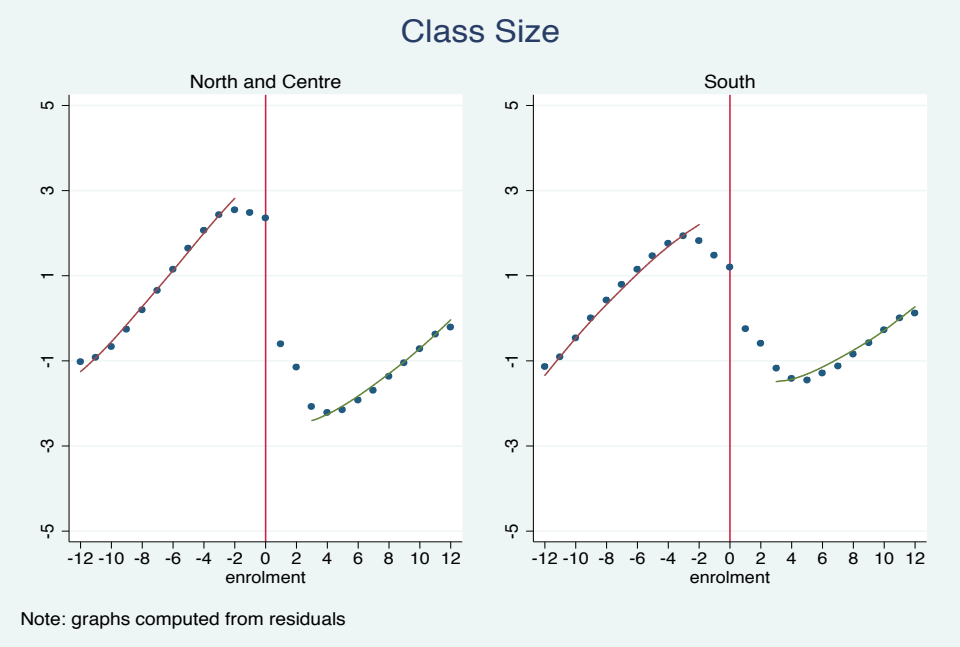


Figure 3: Class size by enrollment, actual class size and as predicted by the Maimonides' rule for 2 grade in reform years

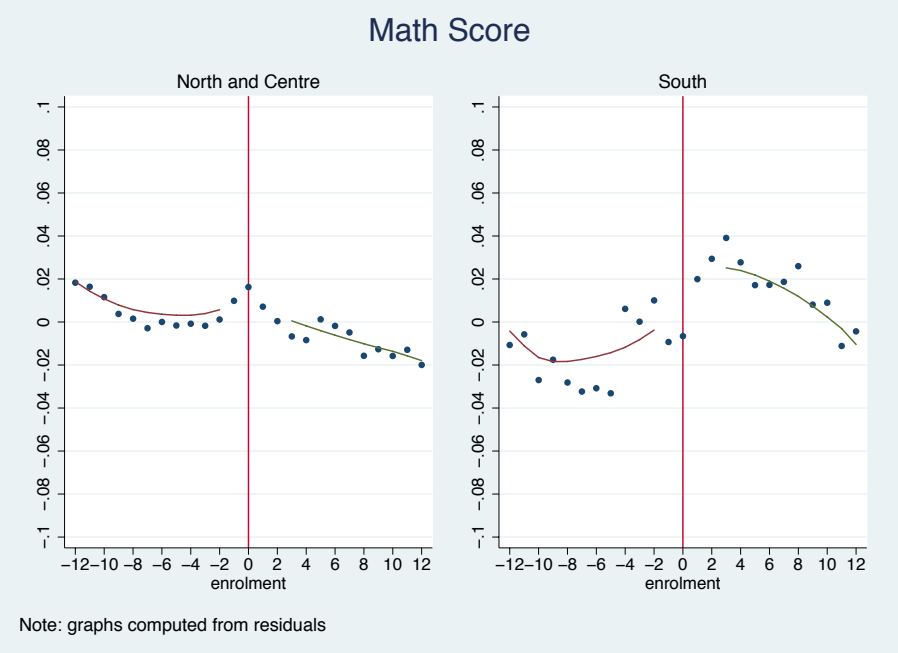


Panel A: Non parametric estimates for class size for 2 grade



Panel B: Non parametric estimates for class size for 5 grade

Figure 4: Non parametric estimates for class size



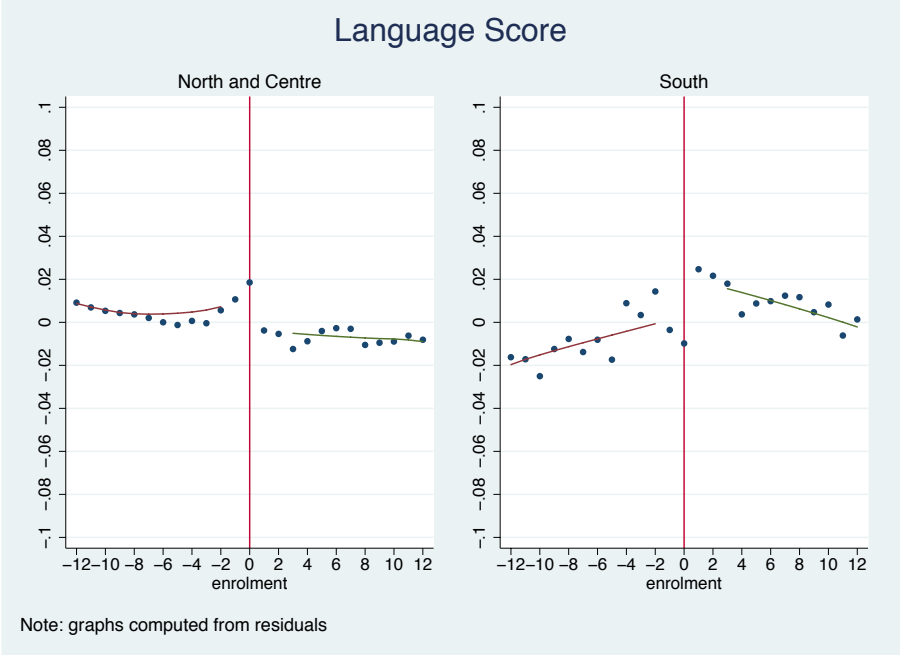
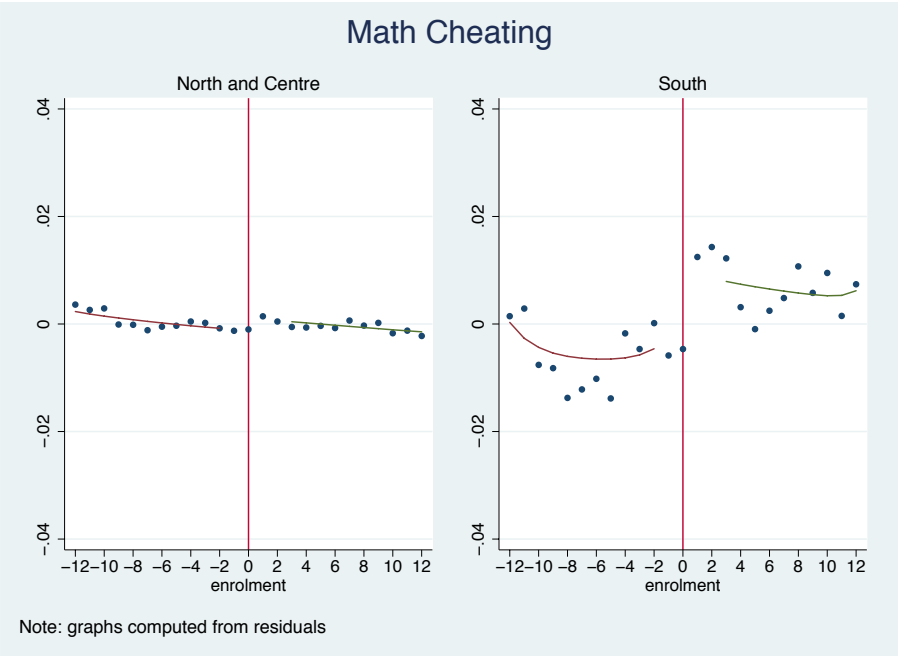


Figure 5: Non parametric estimates for score (math and language)



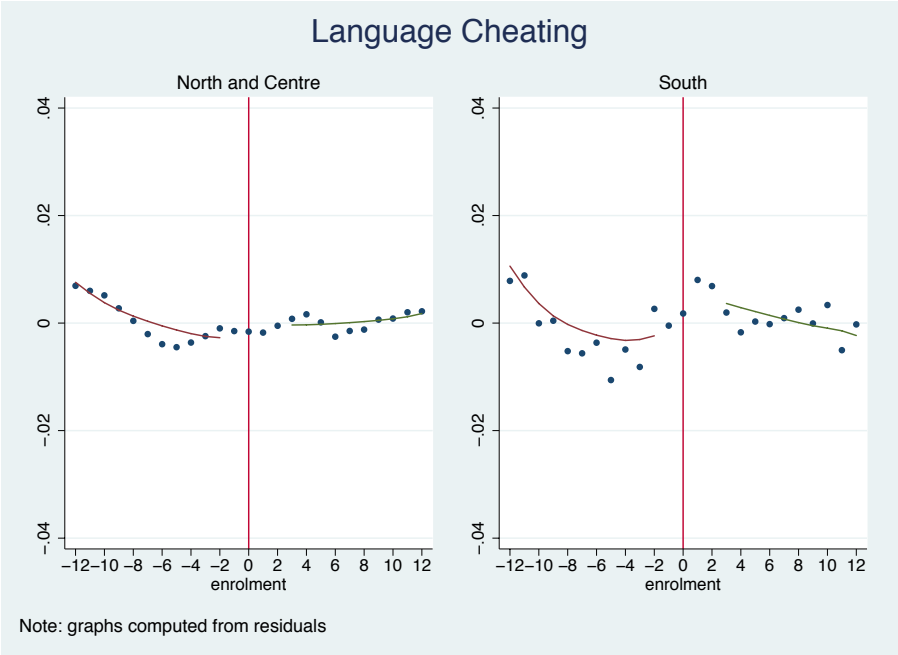


Figure 6: Non parametric estimates for cheating (math and language)

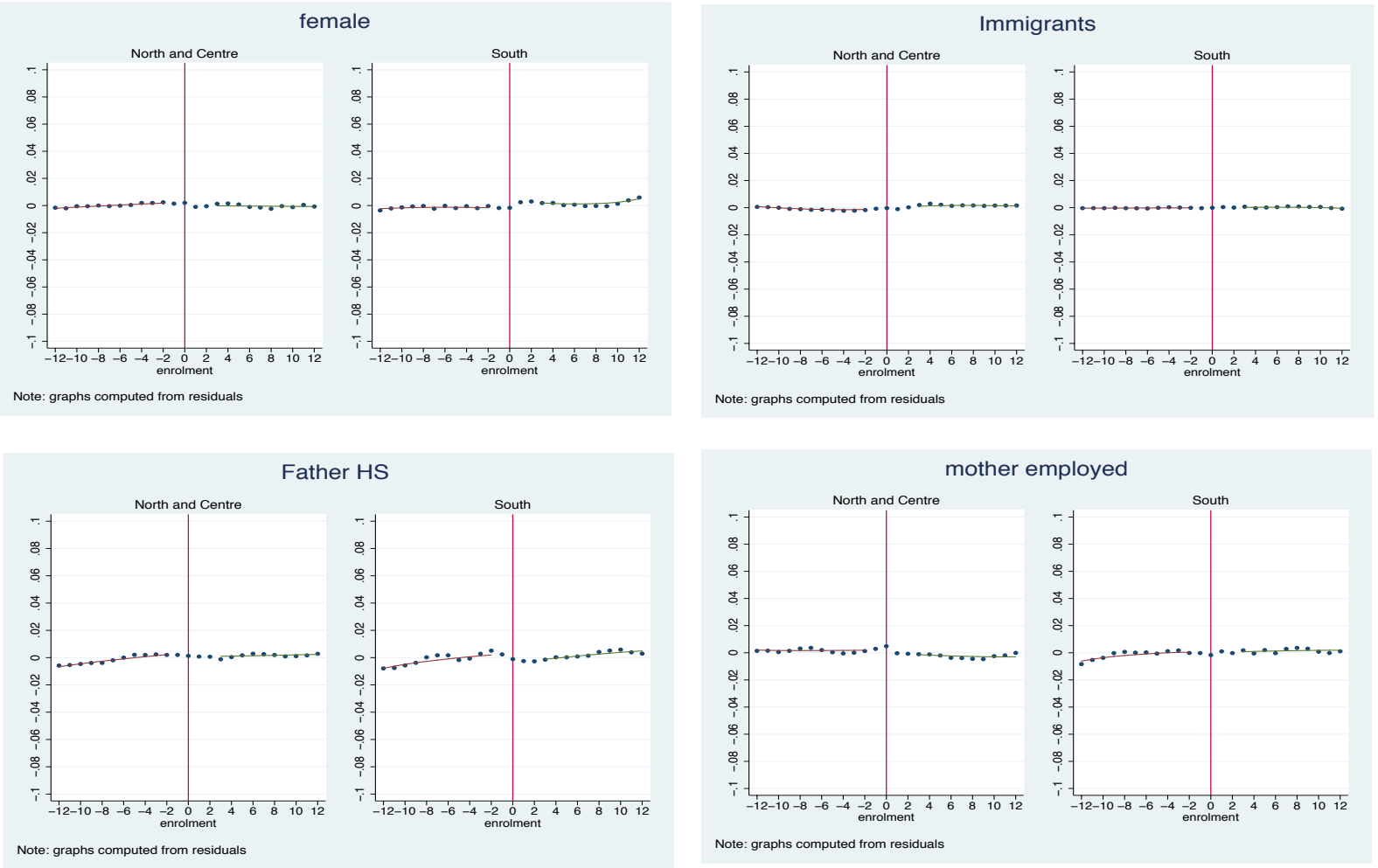


Figure 7: Non parametric estimates for selected covariates

References

Angrist, J. D., and V. Lavy (1999): "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement," *Quarterly Journal of Economics*, 114(2), 533–575.

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

Bonesronning, H. (2003): "Class size effects on student achievement in Norway: Patterns and explanations," *Southern Economic Journal*.

Bratti, M., D. Checchi, and A. Filippin (2007): "Territorial differences in Italian students' mathematical competences: Evidence from PISA," *Giornale degli Economisti e Annali di Economia*, 66(3), 299–335.

Brunello, G., and D. Checchi (2005): "School quality and family background in Italy," *Economics of Education Review*, 24, 563–577.

Chetty, R., J. Friedman, N. Hilger, E. Saez, D. Schanzenbach, and D. Yagan (2011): "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR," *Quarterly Journal of Economics*, 126(4), 1593–1660.

Dobbelsteen, S., J. Levin, and H. Oosterbeek (2002): "The causal effect of class size on scholastic achievement: Distinguishing the pure class size effect from the effect of changes in class composition," *Oxford Bulletin of Economics and Statistics*, 64(1), 17–38.

Gary-Bobo, R. J., and M.-B. Mahjoub (2006): "Estimation of class-size effects, using Maimonides' rule: the case of French junior high schools," CEPR Discussion Papers 5754.

Guiso, L., P. Sapienza, and L. Zingales (2004): "The Role of Social Capital in Financial Development," *American Economic Review*, 94(3), 526–556.

Hanushek, E. A. (1995): "Interpreting recent research on schooling in developing countries," *The World Bank Research Observer*, X, 227–246.

Hoxby, C. (2000): "Does Competition among Public Schools Benefit Students and Taxpayers?," *American Economic Review*, 90(5), 1209–1238.

Ichino, A., and G. Maggi (2000): "Work Environment and Individual Background: Explaining Regional Shirking Differentials in a Large Italian Firm," *Quarterly Journal of Economics,*, 115(3), 933–959.

Imbens, G., and K. Kalyanaraman (2012): "Optimal Bandwidth Choice for the Regression Discontinuity Estimator," *Review of Economic Studies*, 79(3), 933–959.

INVALSI (2010): "Sistema Nazionale di Valutazione - A.S. 2009/2010, Rilevazione degli apprendimenti," *Technical Report*.

Jacob, B., and S. Levitt (2003): "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating," *Quarterly Journal of Economics*, 118(3), 843–77.

Krueger, A. (1999): "Experimental estimates of education production functions," *Quarterly Journal of Economics*, 114, 497–532.

Leuven, E., H. Oosterbeek, and M. Ronning (2008): "Quasi-experimental estimates of the effect of class size achievement in Norway," *The Scandinavian Journal of Economics*, 110(4), 663–693.

Nannicini, T., A. Stella, G. Tabellini, and U. Troiano (2013): "Social Capital and Political Accountability," *American Economic Journal: Economic Policy*, 5, 1957–1969.

Piketty, T. (2004): "Should we reduce class size or school segregation? Theory and evidence from France," presentation at the Roy Seminars, Association pour le dévelopement de la recherche en économie et en statistique (ADRES), 22 November, available at: http://www.adres.polytechnique.fr/SEMINAIRE/221104b.pdf.

Putnam, R., R. Leonardi, and R. Nanetti (1993): *Making Democracy Work*. Princeton University Press, Princeton.

Quintano, C., R. Castellano, and S. Longobardi (2009): "A Fuzzy Clustering Approach to Improve the Accuracy of Italian Student Data. An Experimental Procedure to Correct the Impact of the Outliers on Assessment Test Scores," *Statistica & Applicazioni*, Vol.VII(2), 149–171.

Urquiola, M., and E. Verhoogen (2009): "Class size caps, sorting, and the regression discontinuity design," *American Economic Review*, 99(1), 179–215.

Woessmann, L. (2005): "Educational production in Europe," *Economic Policy*, 43, 445–493.

Appendix

Table A.1: Reduced form estimates of the effect of Maimonides' rule on scores, cheating and class size

	math			language		
	Italy	**North/Centre**	**South**	**Italy**	**North/Centre**	**South**
	(1)	(2)	(3)	(4)	(5)	(6)
			A. scores			
Maimonides' rule	-0.0031***	-0.0020**	-0.0064***	-0.0019**	-0.0006	-0.0048***
	(0.0010)	(0.0010)	(0.0022)	(0.0008)	(0.0008)	(0.0018)
Means	*0.005*	*-0.077*	*0.142*	*0.009*	*-0.007*	*0.035*
(sd)	*(0.636)*	*(0.501)*	*(0.794)*	*(0.521)*	*(0.428)*	*(0.648)*
N	134,824	84,427	50,397	134,826	84,429	50,397
			B. cheating			
Maimonides' rule	-0.0010***	-0.0004	-0.0024***	-0.0007**	-0.0002	-0.0018**
	(0.0004)	(0.0003)	(0.0009)	(0.0003)	(0.0003)	(0.0008)
Means	*0.064*	*0.020*	*0.137*	*0.055*	*0.022*	*0.109*
(sd)	*(0.244)*	*(0.138)*	*(0.344)*	*(0.227)*	*(0.147)*	*(0.312)*
N	134,811	84,421	50,390	134,820	84,424	50,396
			C. cheating (INVALSI)			
Maimonides' rule	-0.0007**	-0.0003*	-0.0015**	-0.0007***	-0.0003	-0.0014**
	(0.0003)	(0.0002)	(0.0007)	(0.0003)	(0.0002)	(0.0006)
Means	*0.065*	*0.025*	*0.132*	*0.057*	*0.028*	*0.105*
(sd)	*(0.191)*	*(0.099)*	*(0.271)*	*(0.174)*	*(0.104)*	*(0.243)*
N	134,819	84,426	50,393	134,820	84,424	50,396
			D. class size			
Maimonides rule	0.5035***	0.5343***	0.4416***			
	(0.0061)	(0.0078)	(0.0101)			
Means	*20.12*	*20.29*	*19.83*			
(sd)	*(3.375)*	*(3.332)*	*(3.428)*			
N	134,826	84,429	50,397			

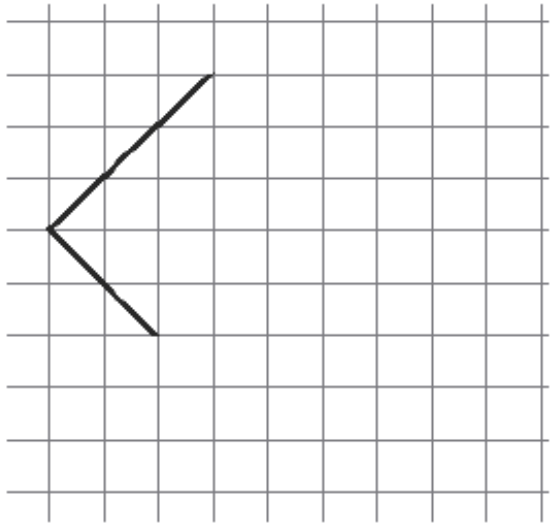
Notes: This table shows results from RF effects of scores (Panel A), cheating (Panel B), cheating (INVALSI) (Panel C) and class size (Panel D) on Maimonides' rule. The specification employs a quadratic polynomial in grade enrolment, segment dummies and their interactions. The unit of observation is the class in all regressions. Robust standard errors, clustered on school and grade, are shown in parentheses. Control variables include: % female students, % students behind and ahead grade level, % immigrants, % fathers at least high school graduate, % unemployed mothers, % mother NILF, grade and year dummies, and dummies for missing values in these variables. All regressions have sampling strata controls (grade enrolment at institution, region dummies and their interactions). * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.2: First stage regressions for overidentified IV/2SLS

	class size			cheating math			cheating language		
VARIABLES	**Italy**	**North/Centre**	**South**	**Italy**	**North/Centre**	**South**	**Italy**	**North/Centre**	**South**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Maimonides	0.7236***	0.7559***	0.6630***	-0.0011**	-0.0006	-0.0021	-0.0014***	-0.0007*	-0.0027**
	(0.0064)	(0.0076)	(0.0114)	(0.0005)	(0.0004)	(0.0013)	(0.0005)	(0.0004)	(0.0011)
observer	0.0259	0.0367	0.0159	-0.0294***	-0.0105***	-0.0624***	-0.0245***	-0.0118***	-0.0470***
	(0.0232)	(0.0262)	(0.0438)	(0.0018)	(0.0012)	(0.0044)	(0.0016)	(0.0013)	(0.0039)
10% above cutoff	2.9442***	3.0384***	2.8504***	-0.0007	-0.0033	0.0022	-0.0095*	-0.0073*	-0.0131
	(0.0748)	(0.0948)	(0.1201)	(0.0060)	(0.0039)	(0.0135)	(0.0052)	(0.0040)	(0.0115)
within 10% from cutoff	-1.4102***	-1.2452***	-1.6925***	-0.0020	0.0009	-0.0068	0.0059*	0.0065***	0.0045
	(0.0527)	(0.0629)	(0.0913)	(0.0035)	(0.0024)	(0.0085)	(0.0032)	(0.0024)	(0.0077)
N	134,826	84,429	50,397	134,811	84,421	50,390	134,820	84,424	50,396

Notes: This table shows the first stage regressions for class size and cheating on observer at institution, Maimonides' rule and a dummy for grade enrolment being in a 10 percent window below and above each cutoff. The specification employs a quadratic polynomial in grade enrolment, segment dummies and their interactions. The unit of observation is the class in all regressions. Robust standard errors, clustered on school and grade, are shown in parentheses. Control variables include: % female students, % students behind and ahead grade level, % immigrants, % fathers at least high school graduate, % unemployed mothers, % mother NILF grade and year dummies, and dummies for missing values in these variables. All regressions have sampling strata controls (grade enrolment at institution, region dummies and their interactions). * significant at 10%; ** significant at 5%; *** significant at 1%.

D23. Osserva la seguente figura.



a. Completa la figura in modo da ottenere un quadrato.

b. Spiega come hai fatto per disegnare il quadrato.

.....

.....

.....

Figure 8: Example of open question in math test - V grade 2010/11

C3. Leggi questa frase: "L'uccello notturno emise un suono talmente acuto che fece molta paura agli abitanti del bosco".
Indica nella tabella quali parole sono nomi e quali non lo sono. Metti una crocetta per ogni riga della tabella.

		È un nome	Non è un nome
a.	L'	☐	☐
b.	uccello	☐	☐
c.	notturno	☐	☐
d.	emise	☐	☐
e.	un	☐	☐
f.	suono	☐	☐
g.	talmente	☐	☐
h.	acuto	☐	☐
i.	che	☐	☐
l.	fece	☐	☐
m.	molta	☐	☐
n.	paura	☐	☐
o.	agli	☐	☐
p.	abitanti	☐	☐
q.	del	☐	☐
r.	bosco	☐	☐

Figure 9: Example of open question in language test - V grade 2010/11

Servizio Nazionale di Valutazione a.s. 2010/11

CLASSE:

Scheda Risposte Studente n°

Risultati delle prove

Codice istituto: Codice Scuola:

Codice plesso: Livello:

Codice Classe: **NON CAMPIONE**

Codice studente: Numero progressivo studente:

PROVA ITALIANO (1)

A1	☐A	☐B	☐C	☐D	☐NV	**C1_a1**	☐0	☐1	☐NV		
A2	☐A	☐B	☐C	☐D	☐NV	**C1_a2**	☐0	☐1	☐NV		
A3	☐A	☐B	☐C	☐D	☐NV	**C1_b1**	☐0	☐1	☐NV		
A4	☐A	☐B	☐C	☐D	☐NV	**C1_b2**	☐0	☐1	☐NV		
A5	☐A	☐B	☐C	☐D	☐NV	**C1_b3**	☐0	☐1	☐NV		
A6	☐A	☐B	☐C	☐D	☐NV	**C2**	☐A	☐B	☐C	☐D	☐NV
A7	☐A	☐B	☐C	☐D	☐NV	**C3_a**	☐Nome		☐Non_Nome		☐NV
A8	☐A	☐B	☐C	☐D	☐NV	**C3_b**	☐Nome		☐Non_Nome		☐NV
A9	☐A	☐B	☐C	☐D	☐NV	**C3_c**	☐Nome		☐Non_Nome		☐NV
A10	☐A	☐B	☐C	☐D	☐NV	**C3_d**	☐Nome		☐Non_Nome		☐NV
A11	☐A	☐B	☐C	☐D	☐NV	**C3_e**	☐Nome		☐Non_Nome		☐NV
A12	☐A	☐B	☐C	☐D	☐NV	**C3_f**	☐Nome		☐Non_Nome		☐NV
A13	☐A	☐B	☐C	☐D	☐NV	**C3_g**	☐Nome		☐Non_Nome		☐NV
A14	☐A	☐B	☐C	☐D	☐NV	**C3_h**	☐Nome		☐Non_Nome		☐NV
A15	☐A	☐B	☐C	☐D	☐NV	**C3_i**	☐Nome		☐Non_Nome		☐NV
A16	☐A	☐B	☐C	☐D	☐NV	**C3_l**	☐Nome		☐Non_Nome		☐NV
A17	☐A	☐B	☐C	☐D	☐NV	**C3_m**	☐Nome		☐Non_Nome		☐NV
B1	☐A	☐B	☐C	☐D	☐NV	**C3_n**	☐Nome		☐Non_Nome		☐NV
B2	☐A	☐B	☐C	☐D	☐NV	**C3_o**	☐Nome		☐Non_Nome		☐NV
B3	☐A	☐B	☐C	☐D	☐NV	**C3_p**	☐Nome		☐Non_Nome		☐NV
B4	☐A	☐B	☐C	☐D	☐NV	**C3_q**	☐Nome		☐Non_Nome		☐NV
B5	☐A	☐B	☐C	☐D	☐NV	**C3_r**	☐Nome		☐Non_Nome		☐NV
B6	☐A	☐B	☐C	☐D	☐NV	**C4**	☐0	☐1	☐NV		
B7	☐A	☐B	☐C	☐D	☐NV	**C5**	☐0	☐1	☐NV		
B8	☐A	☐B	☐C	☐D	☐NV	**C6**	☐A	☐B	☐C	☐D	☐NV
B9	☐A	☐B	☐C	☐D	☐NV	**C7**	☐0	☐1	☐NV		
B10	☐A	☐B	☐C	☐D	☐NV	**C8**	☐A	☐B	☐C	☐D	☐NV
B11	☐A	☐B	☐C	☐D	☐NV	**C9**	☐A	☐B	☐C	☐D	☐NV
B12	☐A	☐B	☐C	☐D	☐NV	**C10**	☐0	☐1	☐NV		
B13	☐A	☐B	☐C	☐D	☐NV						
B14	☐A	☐B	☐C	☐D	☐NV						
B15	☐A	☐B	☐C	☐D	☐NV						

PROVA MATEMATICA (1)

D1_a	☐V	☐F	☐NV		
D1_b	☐V	☐F	☐NV		
D1_c	☐V	☐F	☐NV		
D1_d	☐V	☐F	☐NV		
D2	☐A	☐B	☐C	☐D	☐NV
D3	☐0	☐1	☐NV		
D4_a	☐A	☐B	☐C	☐D	☐NV
D4_b	☐0	☐1	☐NV		
D5	☐A	☐B	☐C	☐D	☐NV
D6	☐A	☐B	☐C	☐D	☐NV
D7	☐A	☐B	☐C	☐D	☐NV
D8	☐A	☐B	☐C	☐D	☐NV
D9	☐0	☐1	☐NV		
D10	☐A	☐B	☐C	☐D	☐NV
D11	☐A	☐B	☐C	☐D	☐NV
D12	☐0	☐1	☐NV		
D13	☐A	☐B	☐C	☐D	☐NV
D14	☐A	☐B	☐C	☐D	☐NV
D15	☐A	☐B	☐C	☐D	☐NV
D16_a	☐0	☐1	☐NV		
D16_b	☐0	☐1	☐NV		
D17_a	☐V	☐F	☐NV		
D17_b	☐V	☐F	☐NV		
D17_c	☐V	☐F	☐NV		
D17_d	☐V	☐F	☐NV		
D18	☐A	☐B	☐C	☐D	☐NV
D19	☐0	☐1	☐NV		
D20	☐A	☐B	☐C	☐D	☐NV
D21_a	☐0	☐1	☐NV		
D21_b	☐0	☐1	☐NV		
D22	☐A	☐B	☐C	☐D	☐NV
D23_a	☐0	☐1	☐NV		
D23_b	☐0	☐1	☐NV		
D24_a	☐0	☐1	☐NV		
D24_b	☐0	☐1	☐NV		
D24_c	☐0	☐1	☐NV		
D25	☐0	☐1	☐NV		
D26	☐A	☐B	☐C	☐D	☐NV
D27	☐A	☐B	☐C	☐D	☐NV
D28_a	☐km	☐m	☐cm	☐mm	☐NV
D28_b	☐km	☐m	☐cm	☐mm	☐NV
D28_c	☐km	☐m	☐cm	☐mm	☐NV
D29_a	☐V	☐F	☐NV		
D29_b	☐V	☐F	☐NV		
D29_c	☐V	☐F	☐NV		
D29_d	☐V	☐F	☐NV		
D30	☐A	☐B	☐C	☐D	☐NV

(1) Barrare NV per risposta non valida (2 risposte o risposta incomprensibile) e non barrare nulla in caso di risposta omessa
(**ATTENZIONE Non spillare, non modificare per nessun motivo i dati precompilati della scheda)**

Figure 10: Answer sheet for V grade in 2010/11