In a Small Moment: Class Size and Moral Hazard in the Mezzogiorno

Josh Angrist (MIT and NBER)

Erich Battistin (Queen Mary University, IRVAPP and IZA)

Daniela Vuri (University of Rome "Tor Vergata" and IZA)

June 2014

A Tale of Two Inputs

- Social scientists, educators and parents have long been concerned with causal effects of class size, a key input in education production
- Small classes are costly, so evidence on their effectiveness is welcome
- Class size research typically measures effectiveness with standardized test scores

A Tale of Two Inputs

- Social scientists, educators and parents have long been concerned with causal effects of class size, a key input in education production
- Small classes are costly, so evidence on their effectiveness is welcome
- Class size research typically measures effectiveness with standardized test scores

- Standardized assessments may prove unreliable
- As testing regimes have proliferated, so has the temptation to cut corners or cheat, an unintended consequence demonstrated by Jacob and Levitt (2003) and Dee et al. (2011), among others

A Tale of Two Inputs

- Social scientists, educators and parents have long been concerned with causal effects of class size, a key input in education production
- Small classes are costly, so evidence on their effectiveness is welcome
- Class size research typically measures effectiveness with standardized test scores

- Standardized assessments may prove unreliable
- As testing regimes have proliferated, so has the temptation to cut corners or cheat, an unintended consequence demonstrated by Jacob and Levitt (2003) and Dee et al. (2011), among others

- Moral hazard is an unwelcome input in measured education production, as we've seen recently in Atlanta, where district officials face indictment for test-related fraud
- This paper documents and diagnoses a surprising interaction between class size and moral hazard in Italian primary schools

Causal Class Size Effects

- The STAR randomized trial revealed important achievement gains from smaller classes (Krueger 1999; Chetty et al. 2011)
    - Such randomized evaluations are, as yet, exceedingly rare

Causal Class Size Effects

- The STAR randomized trial revealed important achievement gains from smaller classes (Krueger 1999; Chetty et al. 2011)
  - Such randomized evaluations are, as yet, exceedingly rare
- Researchers have therefore turned to quasi-experimental designs
  - Angrist and Lavy (1999) and Hoxby (2000) exploit the variation in class size generated by rules for classroom assignment in a regime with class size caps (Maimonides' rule caps Israeli class size at 40)
  - Maimonides' legacy has since appeared in many countries

Causal Class Size Effects

- The STAR randomized trial revealed important achievement gains from smaller classes (Krueger 1999; Chetty et al. 2011)
    - Such randomized evaluations are, as yet, exceedingly rare
- Researchers have therefore turned to quasi-experimental designs
    - Angrist and Lavy (1999) and Hoxby (2000) exploit the variation in class size generated by rules for classroom assignment in a regime with class size caps (Maimonides' rule caps Israeli class size at 40)
    - Maimonides' legacy has since appeared in many countries
- In contemporary Italy, Maimonides' Rule applies with caps of 25 or 27
    - As you'll soon see, RD estimates using this suggest small class size boost scores ... mostly in the South

Maimonides in the Mezzogiorno

- Southern Italy - *the Mezzogiorno* - is distinguished by high unemployment, low per-capita income, crime, lags in development ...
  - and widespread manipulation of standardized test scores (**Figure 1**)

Maimonides in the Mezzogiorno

- Southern Italy - *the Mezzogiorno* - is distinguished by high unemployment, low per-capita income, crime, lags in development ...
  - and widespread manipulation of standardized test scores (**Figure 1**)
- We show here that returns to class size in the Mezzogiorno reflect some sort of score manipulation (i.e., something other than honest answers by students), not learning

Maimonides in the Mezzogiorno

- Southern Italy - *the Mezzogiorno* - is distinguished by high unemployment, low per-capita income, crime, lags in development ...
  - and widespread manipulation of standardized test scores (**Figure 1**)
- We show here that returns to class size in the Mezzogiorno reflect some sort of score manipulation (i.e., something other than honest answers by students), not learning
- We investigate the how and why of this
  - Italy is the original low-stakes labor market. Teachers' pay depends only on seniority, without regard to qualification, performance or conduct. Why cheat?
  - We uncover moral hazard in teacher effort, apparently unrelated to accountability: manipulation by *shirking* more than cheating
  - A caution for the interpretation of causal class size effects, unrelated to the specifics of research design
  - Manipulation arises not only where accountability pressures are high

Data

- In 2009, Italy introduced nationwide achievement tests
- We analyze data on 2nd and 5th graders in public schools (**background**) for years 2009/10, 2010/11 and 2011/12
  - We drop classes smaller than the official minimum (10 before 2010, and 15 afterwards)
  - Our sample is limited to schools with grade enrollment of 160 or less (about 2.6 mil students, in 140,000 classes)

Data

- In 2009, Italy introduced nationwide achievement tests
- We analyze data on 2nd and 5th graders in public schools (**background**) for years 2009/10, 2010/11 and 2011/12
  - We drop classes smaller than the official minimum (10 before 2010, and 15 afterwards)
  - Our sample is limited to schools with grade enrollment of 160 or less (about 2.6 mil students, in 140,000 classes)
- These data include (summary statistics in Table 1):
  - **Test scores**: number of correct answers; standardized by subject (math and language), year of survey, and grade
  - **Student data**: includes gender, citizenship, and information on parents' employment status and educational background
  - **Class size**: defined as administrative enrollment at the beginning of the school year

Maimonides' Rules

- Min and max
  - Until 2008/09, the min and max were 10 and 25
  - Rolling forward with first grade in 2009/10, new min=15 and max=27
  - The higher limit applies to our 2nd graders in 2010/11-2011/12
  - The law allows a 2-3 student deviation (10%); it's "flexible Maimonides" in practice

Maimonides' Rules

- Min and max
  - Until 2008/09, the min and max were 10 and 25
  - Rolling forward with first grade in 2009/10, new min=15 and max=27
  - The higher limit applies to our 2nd graders in 2010/11-2011/12
  - The law allows a 2-3 student deviation (10%); it's "flexible Maimonides" in practice
- Ignoring flexibility, Maimonides' Rule predicts the size of any class i, in grade g, at school k in year t, as:

$$f_{igkt} = \frac{r_{gkt}}{\left[int\left(\left(r_{gkt}-1\right)/c_{gt}\right)+1\right]}$$

where r_{gkt} is grade-level enrollment and c_{gt} is effective max

Maimonides' Rules

- Min and max
  - Until 2008/09, the min and max were 10 and 25
  - Rolling forward with first grade in 2009/10, new min=15 and max=27
  - The higher limit applies to our 2nd graders in 2010/11-2011/12
  - The law allows a 2-3 student deviation (10%); it's "flexible Maimonides" in practice

- Ignoring flexibility, Maimonides' Rule predicts the size of any class i, in grade g, at school k in year t, as:

$$f_{igkt} = \frac{r_{gkt}}{\left[int\left(\left(r_{gkt} - 1\right) / c_{gt}\right) + 1\right]}$$

where r_{gkt} is grade-level enrollment and c_{gt} is effective max

- Figure 2 plots average class size and f_{igkt} against r_{gkt}

Class Size Effects

Graphical Analysis of Score Effects

- We begin with nonparametric visual IV, focusing on enrollment in a [-12,12] window around Maimonides's cutoffs
  - The figures also plot LLR fits for points more than 2 kids away from the cutoff on either side
    - The edge kernel and an optimal bandwidth were used for smoothing [the dots plot an MA(+1,-1), but the LLR is fit to micro data]

Graphical Analysis of Score Effects

- We begin with nonparametric visual IV, focusing on enrollment in a [-12,12] window around Maimonides's cutoffs
  - The figures also plot LLR fits for points more than 2 kids away from the cutoff on either side
    - The edge kernel and an optimal bandwidth were used for smoothing [the dots plot an MA(+1,-1), but the LLR is fit to micro data]
- Every picture tells a story ...
  - **First stages:** Class size in **Figure 3** (grade 2) and **Figure 4** (grade 5)
  - **Reduced forms:** Test scores in **Figure 5** (math) and **Figure 6** (language)

Graphical Analysis of Score Effects

- We begin with nonparametric visual IV, focusing on enrollment in a [-12,12] window around Maimonides's cutoffs
  - The figures also plot LLR fits for points more than 2 kids away from the cutoff on either side
    - The edge kernel and an optimal bandwidth were used for smoothing [the dots plot an MA(+1,-1), but the LLR is fit to micro data]
- Every picture tells a story ...
  - **First stages:** Class size in **Figure 3** (grade 2) and **Figure 4** (grade 5)
  - **Reduced forms:** Test scores in **Figure 5** (math) and **Figure 6** (language)
- These figures suggest class size effects are nonparametrically identified by Maimonides cutoffs

Empirical Framework

- We use a flexible parametric setup that exploits Maimonides-induced changes in slope as well as discontinuities, while facilitating an investigation of multivariate causal models
- y_{igkt}, the average score in class i in grade g at school k in year t, is determined by the running variable, r_{gkt}, and class size, s_{igkt}:

$$y_{igkt} = \rho_0(t,g) + \beta s_{igkt} + \rho_1 r_{gkt} + \rho_2 r_{gkt}^2 + \varepsilon_{igkt}, \tag{1}$$

  where $\rho_0(t,g)$ captures year and grade effects
  - f_{igkt} provides instruments for s_{igkt}

Empirical Framework

- We use a flexible parametric setup that exploits Maimonides-induced changes in slope as well as discontinuities, while facilitating an investigation of multivariate causal models
- y_{igkt}, the average score in class i in grade g at school k in year t, is determined by the running variable, r_{gkt}, and class size, s_{igkt}:

$$y_{igkt} = \rho_0(t,g) + \beta s_{igkt} + \rho_1 r_{gkt} + \rho_2 r_{gkt}^2 + \varepsilon_{igkt}, \tag{1}$$

  where $\rho_0(t,g)$ captures year and grade effects
  - f_{igkt} provides instruments for s_{igkt}
- Details
  - The estimating equation controls for demographic and sampling strata variables (used in the monitoring experiment)
  - We also allow the coefficients on r_{gkt} to vary across windows centered around each cutoff, and include a full set of window dummies - we call this "the interacted specification"
  - Standard errors are clustered by institution

Achievement Estimates

- First stage estimates (s_{igkt} on f_{igkt}) are in **Table A1**
  - A one-student increase in predicted class size increases actual class size by about half a student, in both North/Central and Southern Italy

Achievement Estimates

- First stage estimates (s_{igkt} on f_{igkt}) are in **Table A1**
    - A one-student increase in predicted class size increases actual class size by about half a student, in both North/Central and Southern Italy
- **Table 2** reports OLS and 2SLS estimates of the effect of class size on test scores
    - OLS estimates show small negative class size effects in N/C region, positive in the South
    - 2SLS estimates suggest smaller classes boost achievement, with a precisely estimated effect of about 0.05σ in math and 0.04σ in language for a 10 student reduction
        - The interacted specification generates similar results, with a slight loss of precision

Achievement Estimates

- First stage estimates (s_{igkt} on f_{igkt}) are in **Table A1**
  - A one-student increase in predicted class size increases actual class size by about half a student, in both North/Central and Southern Italy
- **Table 2** reports OLS and 2SLS estimates of the effect of class size on test scores
  - OLS estimates show small negative class size effects in N/C region, positive in the South
  - 2SLS estimates suggest smaller classes boost achievement, with a precisely estimated effect of about 0.05σ in math and 0.04σ in language for a 10 student reduction
    - The interacted specification generates similar results, with a slight loss of precision
- The estimated returns to class size are over twice as large in the South: the largest is $+0.13\sigma$ in math for a 10 student reduction (reported in column 9, from the interacted model)

Maimonides and Manipulation

Measuring Manipulation

- We identify manipulation using a procedure similar to that used by INVALSI
- Class-level indicators of compromised scores are defined using within-class information on:
  - average and standard deviation of test scores
  - proportion of items missing
  - variability in response patterns (measured by a Herfindahl index)

Measuring Manipulation

- We identify manipulation using a procedure similar to that used by INVALSI
- Class-level indicators of compromised scores are defined using within-class information on:
  - average and standard deviation of test scores
  - proportion of items missing
  - variability in response patterns (measured by a Herfindahl index)
- A principal component analysis flags classes with abnormally high performance, small dispersion of test scores, low proportion of missing items, and a high concentration in response patterns
- We code a dummy variable indicating classrooms where manipulation seems likely (in the spirit of Jacob and Levitt, 2003)

Effects of Class Size on Manipulation

- Manipulation rates near enrollment cutoffs are plotted in **Figure 7** (for math) and **Figure 8** (for language)

Effects of Class Size on Manipulation

- Manipulation rates near enrollment cutoffs are plotted in **Figure 7** (for math) and **Figure 8** (for language)
- **Table 3** reports OLS and 2SLS estimates of the effect of class size on score manipulation in a format paralleling that of Table 2
  - OLS estimates show manipulation is negatively correlated with class size, with stronger effects in the South
  - 2SLS estimates for the South are again especially large; estimates of effects elsewhere are negative though mostly not significant
  - *Small classes boost manipulation as well as measured achievement;* we'll soon outline a model explaining this

Effects of Class Size on Manipulation

- Manipulation rates near enrollment cutoffs are plotted in **Figure 7** (for math) and **Figure 8** (for language)

- **Table 3** reports OLS and 2SLS estimates of the effect of class size on score manipulation in a format paralleling that of Table 2
  - OLS estimates show manipulation is negatively correlated with class size, with stronger effects in the South
  - 2SLS estimates for the South are again especially large; estimates of effects elsewhere are negative though mostly not significant
  - *Small classes boost manipulation as well as measured achievement;* we'll soon outline a model explaining this

- We next show that the manipulation declines sharply with external monitoring - an important result for our purposes because this identifies the culprits!

Monitoring and Manipulation

The Monitoring Experiment

- Tests are usually proctored by teachers from the same school (though not the same class)
- About 20% of institutions are randomly assigned external monitors, who supervise test administration and are responsible for score sheet transcription in selected classes

The Monitoring Experiment

- Tests are usually proctored by teachers from the same school (though not the same class)
- About 20% of institutions are randomly assigned external monitors, who supervise test administration and are responsible for score sheet transcription in selected classes

- **Table 5** reports monitoring effects on manipulation and scores
  - Central office monitoring reduces score manipulation
  - The fact that monitors matter suggests teachers are the problem; from the point of view of students, honest teachers should be monitors too

The Monitoring Experiment

- Tests are usually proctored by teachers from the same school (though not the same class)
- About 20% of institutions are randomly assigned external monitors, who supervise test administration and are responsible for score sheet transcription in selected classes

- Table 5 reports monitoring effects on manipulation and scores
  - Central office monitoring reduces score manipulation
  - The fact that monitors matter suggests teachers are the problem; from the point of view of students, honest teachers should be monitors too

- We check random assignment by comparing covariate means across institutions with and without monitors (see Table 4)
  - Good balance in administrative variables
  - Variables collected from school staff are moderately imbalanced, a result we think is explained by the effect of monitoring on data quality

Two Causal Channels

- Tables 3 and 5 motivate a 2SLS setup with two endogenous variables, class size (s_{igkt}) and manipulation (m_{igkt}):

$$y_{igkt} = \rho_0(t,g) + \beta_1 s_{igkt} + \beta_2 m_{igkt} + \rho_1 r_{gkt} + \rho_2 r_{gkt}^2 + \eta_{igkt} \qquad (2)$$

- Excluded IVs: Maimonides' Rule (f_{igkt}) and a dummy for institutions with randomly assigned monitors (M_{igkt})
- First-stage equations for class size and manipulation (**Table 6**):

$$s_{igkt} = \lambda_{10}(t,g) + \mu_{11} f_{igkt} + \mu_{12} M_{igkt} + \lambda_{11} r_{gkt} + \lambda_{12} r_{gkt}^2 + \xi_{ik}$$
$$m_{igkt} = \lambda_{20}(t,g) + \mu_{21} f_{igkt} + \mu_{22} M_{igkt} + \lambda_{21} r_{gkt} + \lambda_{22} r_{gkt}^2 + \upsilon_{ik}$$

Two Causal Channels

- Tables 3 and 5 motivate a 2SLS setup with two endogenous variables, class size (s_{igkt}) and manipulation (m_{igkt}):

$$y_{igkt} = \rho_0(t,g) + \beta_1 s_{igkt} + \beta_2 m_{igkt} + \rho_1 r_{gkt} + \rho_2 r_{gkt}^2 + \eta_{igkt} \qquad (2)$$

- Excluded IVs: Maimonides' Rule (f_{igkt}) and a dummy for institutions with randomly assigned monitors (M_{igkt})
- First-stage equations for class size and manipulation (**Table 6**):

$$s_{igkt} = \lambda_{10}(t,g) + \mu_{11} f_{igkt} + \mu_{12} M_{igkt} + \lambda_{11} r_{gkt} + \lambda_{12} r_{gkt}^2 + \xi_{ik}$$
$$m_{igkt} = \lambda_{20}(t,g) + \mu_{21} f_{igkt} + \mu_{22} M_{igkt} + \lambda_{21} r_{gkt} + \lambda_{22} r_{gkt}^2 + \upsilon_{ik}$$

- To boost precision, we add dummy IVs indicating values of the running variable that fall within 10% of each cutoff
  - Over-identified first stage estimates appear in **Table A2**

Two-Endos Estimates

- Manipulation may *interact* with class size in education production as well as channeling additive class size effects
- We therefore report estimates adding $s_{igkt} * m_{igkt}$ to (2) and using $f_{igkt} * M_{igkt}$ and the extra dummy instruments (for 10% tolerance) interacted with M_{igkt} as instruments
- **Table 7** reports 2SLS estimates of (2)
  - The class size effect disappears, with reasonably precise zeros; confidence intervals exclude the earlier results
  - We don't need interactions to explain away class size effects

Two-Endos Estimates

- Manipulation may *interact* with class size in education production as well as channeling additive class size effects
- We therefore report estimates adding $s_{igkt} * m_{igkt}$ to (2) and using $f_{igkt} * M_{igkt}$ and the extra dummy instruments (for 10% tolerance) interacted with M_{igkt} as instruments
- **Table 7** reports 2SLS estimates of (2)
    - The class size effect disappears, with reasonably precise zeros; confidence intervals exclude the earlier results
    - We don't need interactions to explain away class size effects

- The return to class size generated by Maimonides-type instruments is due entirely to the causal effect of class size on score manipulation, most likely (as explained next) by teachers origins

Threats to Validity

Manipulation Misclassification

- Measurement issues
  - 2SLS estimates of manipulation effects on scores are too big
  - Classification error attenuates first stage estimates, so the corresponding second stage estimates are proportionally inflated
  - As noted by Kane, Rouse, and Staiger (1999), instrumenting doesn't fix non-classical classification error
- We can show that as long as misclassification rates are independent of instruments, mismeasurement of manipulation leaves 2SLS estimates of *class size* effects in (2) unaffected
  - The manipulation effect is inflated by $[\pi_1 + \pi_0 - 1]^{-1}$, where π_j is the probability that score manipulation is correctly detected and we assume $\pi_j > .5$, i.e. score manipulation is a better indicator of actual manipulation than a coin toss

Sorting Near Cutoffs

- As always, endogenous running variable manipulation threatens RD; we look for signs of this in covariate discontinuities

Sorting Near Cutoffs

- As always, endogenous running variable manipulation threatens RD; we look for signs of this in covariate discontinuities
- Maimonides Rule predicts covariates, but it also predicts monitoring
  - Maimonides predicts monitoring because typically (unless enrollment exceeds 100), only one class is monitored: when class size gets smaller, the odds of being monitored go down
  - **Table 8** reports regression estimates of the effect of Maimonides on covariates, with the same controls as used to produce the estimates in Tables 2 and 3

Sorting Near Cutoffs

- As always, endogenous running variable manipulation threatens RD; we look for signs of this in covariate discontinuities
- Maimonides Rule predicts covariates, but it also predicts monitoring
  - Maimonides predicts monitoring because typically (unless enrollment exceeds 100), only one class is monitored: when class size gets smaller, the odds of being monitored go down
  - **Table 8** reports regression estimates of the effect of Maimonides on covariates, with the same controls as used to produce the estimates in Tables 2 and 3
- Maimonides effects on covs parallel the monitoring effects on covariates shown in Table 4: where we see one, we see the other

Sorting Near Cutoffs

- As always, endogenous running variable manipulation threatens RD; we look for signs of this in covariate discontinuities
- Maimonides Rule predicts covariates, but it also predicts monitoring
  - Maimonides predicts monitoring because typically (unless enrollment exceeds 100), only one class is monitored: when class size gets smaller, the odds of being monitored go down
  - **Table 8** reports regression estimates of the effect of Maimonides on covariates, with the same controls as used to produce the estimates in Tables 2 and 3
- Maimonides effects on covs parallel the monitoring effects on covariates shown in Table 4: where we see one, we see the other
- Covariate discontinuities are absent in monitored institutions, suggesting these are indeed driven by the same behavior that drives score manipulation

Origins of Manipulation

Who Manipulates?

- The large effect of monitoring on scores suggests the problem is **teachers** and not students
  - Honest teacher-proctors are the same as monitors to cheating students; Monitors, like substitute teachers, might *facilitate* student cheating
  - Manipulation *decreases* with class size, at odds with the idea that large classes facilitate student cheating
  - Students never see their scores

Who Manipulates?

- The large effect of monitoring on scores suggests the problem is **teachers** and not students
  - Honest teacher-proctors are the same as monitors to cheating students; Monitors, like substitute teachers, might *facilitate* student cheating
  - Manipulation *decreases* with class size, at odds with the idea that large classes facilitate student cheating
  - Students never see their scores

- In addition to test proctoring, score transcription is probably an important channel for teacher manipulation
  - Teachers copy students' original answer sheets onto a machine readable **scheda risposta**
  - Some questions are open: transcribers determine whether answers are correct, missing, or invalid (see examples for **math** and **language**)
  - Transcription is essentially a form of local grading, as with NY Regents

How Class Size Affects Teacher Manipulation

- **Through test administration**:
  - Small classes reduced the odds of monitoring (typically only one class per selected institution is monitored)
  - In large classes, proportionally fewer students are assisted; inappropriate proctor aid also becomes less discrete
- **Through transcription**:
  - The number of teachers transcribing scores probably increases with class size, limiting manipulation through peer monitoring
  - Some teachers either cheat or simply shirk by curbstoning; this is less accurately done in large classes
    - Accuracy may fall with class size w/o regard to cheating, but the relationship between class size and scores disappears once manipulation is accounted for

How Class Size Affects Teacher Manipulation

- **Through test administration**:
  - Small classes reduced the odds of monitoring (typically only one class per selected institution is monitored)
  - In large classes, proportionally fewer students are assisted; inappropriate proctor aid also becomes less discrete
- **Through transcription**:
  - The number of teachers transcribing scores probably increases with class size, limiting manipulation through peer monitoring
  - Some teachers either cheat or simply shirk by curbstoning; this is less accurately done in large classes
    - Accuracy may fall with class size w/o regard to cheating, but the relationship between class size and scores disappears once manipulation is accounted for
- Finally, we ask: *Why* do teachers manipulate?
  - Accountability concerns
  - Shirking and sloppiness

Why Manipulate?

- A **model** of item-level scores discriminates between two alternatives

Why Manipulate?

- A **model** of item-level scores discriminates between two alternatives
- **Accountability concerns**: motivate cheating on difficult items, where students do poorly without help
  - This induces a nonlinear relation between difficulty and scores, tested in **Figure 10**
- **Shirking and sloppiness**: curbstoning transcribers do this more often and less accurately on high effort items
  - This induces grading-effort interactions in the relationship between item difficulty and scores, explored in **Figure 11**

Why Manipulate?

- A **model** of item-level scores discriminates between two alternatives
- **Accountability concerns**: motivate cheating on difficult items, where students do poorly without help
  - This induces a nonlinear relation between difficulty and scores, tested in **Figure 10**
- **Shirking and sloppiness**: curbstoning transcribers do this more often and less accurately on high effort items
  - This induces grading-effort interactions in the relationship between item difficulty and scores, explored in **Figure 11**
- **Table 9** reports estimates of a model that allows for these behaviors

Why Manipulate?

- A **model** of item-level scores discriminates between two alternatives
- **Accountability concerns**: motivate cheating on difficult items, where students do poorly without help
  - This induces a nonlinear relation between difficulty and scores, tested in **Figure 10**
- **Shirking and sloppiness**: curbstoning transcribers do this more often and less accurately on high effort items
  - This induces grading-effort interactions in the relationship between item difficulty and scores, explored in **Figure 11**
- **Table 9** reports estimates of a model that allows for these behaviors
- The results suggests that **moral hazard in effort** (shirking and sloppiness) is the primary explanation for score manipulation

Wrap Up

- Maimonides Rule identifies class size effects in Italy: the first stage is beautiful, the 2SLS estimates it generates, precise
    - Class size effects are much larger in the Mezzogiorno

Wrap Up

- Maimonides Rule identifies class size effects in Italy: the first stage is beautiful, the 2SLS estimates it generates, precise
  - Class size effects are much larger in the Mezzogiorno
- Maimonides also reveals class size effects on score manipulation; a monitoring experiment suggests the problem is teachers

Wrap Up

- Maimonides Rule identifies class size effects in Italy: the first stage is beautiful, the 2SLS estimates it generates, precise
  - Class size effects are much larger in the Mezzogiorno
- Maimonides also reveals class size effects on score manipulation; a monitoring experiment suggests the problem is teachers
- Models with two endogenous variables show that class size effects are driven entirely by score manipulation
  - Manipulation would seem to come from workplace malfeasance rather than accountability concerns
  - Here, manipulation arguably arises from a *lack* of accountability

Wrap Up

- Maimonides Rule identifies class size effects in Italy: the first stage is beautiful, the 2SLS estimates it generates, precise
  - Class size effects are much larger in the Mezzogiorno
- Maimonides also reveals class size effects on score manipulation; a monitoring experiment suggests the problem is teachers
- Models with two endogenous variables show that class size effects are driven entirely by score manipulation
  - Manipulation would seem to come from workplace malfeasance rather than accountability concerns
  - Here, manipulation arguably arises from a *lack* of accountability
- Broader lessons: Score manipulation mimics real learning effects, even in a strong design; manipulation arises without accountability
- Questions: Would simple grading reforms eliminate manipulation? Why *don't* small classes boost learning in Italian schools?

Tables and Figures

Table I. Descriptive statistics

	grade 2 (2009-2011)			grade 5 (2009-2011)		
	Italy	North/Centre	South	Italy	North/Centre	South
	(1)	(2)	(3)	(4)	(5)	(6)
	A. Class characteristics					
female	0.49	0.49	0.49	0.49	0.49	0.49
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
immigrant	0.10	0.14	0.03	0.10	0.14	0.03
	(0.30)	(0.35)	(0.17)	(0.30)	(0.34)	(0.18)
father HS	0.34	0.34	0.33	0.32	0.33	0.30
	(0.47)	(0.48)	(0.47)	(0.47)	(0.47)	(0.46)
mother employed	0.57	0.68	0.39	0.55	0.66	0.38
	(0.49)	(0.47)	(0.49)	(0.50)	(0.47)	(0.49)
pct correct: math	47.9	46.1	51.1	64.2	63.3	65.6
	(14.6)	(12.9)	(16.7)	(12.9)	(10.9)	(15.5)
pct correct: language	69.8	69.2	70.8	74.2	74.3	74.1
	(10.9)	(9.2)	(13.3)	(8.9)	(7.5)	(10.8)
class size	20.1	20.3	19.9	19.7	19.9	19.3
	(3.40)	(3.35)	(3.48)	(3.72)	(3.67)	(3.76)
score manipulation: math	0.06	0.02	0.14	0.07	0.02	0.14
score manipulation: language	0.05	0.02	0.11	0.06	0.02	0.11
	B. School characteristics					
enrollment	40.5	38.8	43.8	38.9	37.3	41.7
	(25.2)	(23.0)	(28.6)	(25.2)	(22.8)	(28.9)
Number of schools	34,591	22,863	11,728	37,476	24,225	13,251

Table 2. OLS and IV/2SLS Estimates of the Effect of Class Size on Test 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	-0.0078	-0.0224***	0.0091	-0.0519***	-0.0436***	-0.0957***	-0.0609***	-0.0417**	-0.1294**
	(0.0070)	(0.0067)	(0.0146)	(0.0134)	(0.0115)	(0.0362)	(0.0196)	(0.0171)	(0.0507)
Enrollment	✗	✗	✗	✗	✗	✗	✗	✗	✗
Enrollment squared	✗	✗	✗	✗	✗	✗	✗	✗	✗
Interactions							✗	✗	✗
N	140,010	87,498	52,512	140,010	87,498	52,512	140,010	87,498	52,512
				B. Language					
Class size	0.0029	-0.0188***	0.0328***	-0.0395***	-0.0313***	-0.0641**	-0.0409***	-0.0215	-0.0937**
	(0.0055)	(0.0053)	(0.0114)	(0.0106)	(0.0092)	(0.0289)	(0.0155)	(0.0136)	(0.0403)
Enrollment	✗	✗	✗	✗	✗	✗	✗	✗	✗
Enrollment squared	✗	✗	✗	✗	✗	✗	✗	✗	✗
Interactions							✗	✗	✗
N	140,010	87,498	52,512	140,010	87,498	52,512	140,010	87,498	52,512

Table 3. OLS and IV/2SLS Estimates of the Effect of Class Size on Score Manipulation

	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	-0.0163***	-0.0074***	-0.0309***	-0.0186***	-0.0042	-0.0542***	-0.0179***	-0.0053	-0.0471**
	(0.0025)	(0.0017)	(0.0058)	(0.0047)	(0.0031)	(0.0143)	(0.0069)	(0.0045)	(0.0202)
Enrollment	✗	✗	✗	✗	✗	✗	✗	✗	✗
Enrollment squared	✗	✗	✗	✗	✗	✗	✗	✗	✗
Interactions							✗	✗	✗
N	139,996	87,491	52,505	139,996	87,491	52,505	139,996	87,491	52,505
				B. Language					
Class size	-0.0166***	-0.0120***	-0.0244***	-0.0202***	-0.0116***	-0.0400***	-0.0161**	-0.0059	-0.0379**
	(0.0023)	(0.0018)	(0.0051)	(0.0043)	(0.0032)	(0.0128)	(0.0063)	(0.0048)	(0.0177)
Enrollment	✗	✗	✗	✗	✗	✗	✗	✗	✗
Enrollment squared	✗	✗	✗	✗	✗	✗	✗	✗	✗
Interactions							✗	✗	✗
N	140,003	87,493	52,510	140,003	87,493	52,510	140,003	87,493	52,510

Table 4. Covariate Balance in the Monitoring 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 on Schools						
Class size	19.812	0.0348	20.031	0.0179	19.456	0.0623
	[3.574]	(0.0303)	[3.511]	(0.0374)	[3.646]	(0.0515)
Grade enrollment at school	53.119	-0.4011	49.804	-0.5477	58.483	-0.1410
	[30.663]	(0.3289)	[27.562]	(0.3913)	[34.437]	(0.5909)
% in class sitting the test	0.939	0.0001	0.934	0.0006	0.947	-0.0007
	[0.065]	(0.0005)	[0.066]	(0.0006)	[0.062]	(0.0008)
% in school sitting the test	0.938	-0.0001	0.933	0.0005	0.946	-0.0010
	[0.054]	(0.0005)	[0.055]	(0.0006)	[0.051]	(0.0008)
% in institution sitting the test	0.937	-0.0001	0.932	0.0005	0.945	-0.0010
	[0.045]	(0.0004)	[0.043]	(0.0005)	[0.045]	(0.0007)
B. Data Provided by School Staff						
Female students	0.482	0.0012	0.483	0.0004	0.479	0.0027*
	[0.121]	(0.0009)	[0.1179]	(0.0011)	[0.126]	(0.0016)
Immigrant students	0.097	0.0010	0.137	0.0004	0.031	0.0020***
	[0.120]	(0.0010)	[0.13]	(0.0014)	[0.056]	(0.0007)
Father HS	0.25	0.0060***	0.258	0.0061***	0.238	0.0056**
	[0.168]	(0.0016)	[0.163]	(0.0019)	[0.176]	(0.0027)
Mother employed	0.441	0.0085***	0.532	0.0067**	0.295	0.0117***
	[0.267]	(0.0024)	[0.258]	(0.0031)	[0.210]	(0.0035)
C. Non-Response Indicators						
Missing data on father's education	0.223	-0.0217***	0.225	-0.0186***	0.221	-0.0271***
	[0.341]	(0.0034)	[0.340]	(0.0043)	[0.343]	(0.0057)
Missing data on mother's occupation	0.195	-0.0168***	0.196	-0.0083**	0.194	-0.0316***
	[0.328]	(0.0033)	[0.325]	(0.0042)	[0.333]	(0.0054)
Missing data on country of origin	0.033	-0.0115***	0.025	-0.0078***	0.045	-0.0178***
	[0.163]	(0.0013)	[0.143]	(0.0014)	[0.192]	(0.0026)
N	140,010		87,498		52,512	

Table 5. Monitoring Effects on Score Manipulation and Test Scores

	Score manipulation			Test scores		
	Italy	North/Centre	South	Italy	North/Centre	South
	(1)	(2)	(3)	(4)	(5)	(6)
			A. Math			
Monitor at institution (M_{igkt})	-0.029***	-0.010***	-0.062***	-0.112***	-0.075***	-0.180***
	(0.002)	(0.001)	(0.004)	(0.006)	(0.005)	(0.012)
Means	0.064	0.020	0.139	0.007	-0.074	0.141
(sd)	(0.246)	(0.139)	(0.346)	(0.637)	(0.502)	(0.796)
N	139,996	87,491	52,505	140,010	87,498	52,512
			B. Language			
Monitor at institution (M_{igkt})	-0.025***	-0.012***	-0.047***	-0.081***	-0.054***	-0.131***
	(0.002)	(0.001)	(0.004)	(0.004)	(0.004)	(0.009)
Means	0.055	0.023	0.110	0.01	-0.005	0.035
(sd)	(0.229)	(0.149)	(0.313)	(0.523)	(0.428)	(0.649)
N	140,003	87,493	52,510	140,010	87,498	52,512

Table 6. Twin First Stages

	A. Score Manipulation					
	Math			Language		
	Italy (1)	North/Centre (2)	South (3)	Italy (4)	North/Centre (5)	South (6)
Maimonides' Rule (f_{igkt})	-0.0009**	-0.0003	-0.0019**	-0.0008**	-0.0003	-0.0015**
	(0.0004)	(0.0002)	(0.0009)	(0.0003)	(0.0003)	(0.0008)
Monitor at institution (M_{igkt})	-0.029***	-0.010***	-0.062***	-0.025***	-0.012***	-0.047***
	(0.002)	(0.001)	(0.004)	(0.002)	(0.001)	(0.004)
N	139,996	87,491	52,505	140,003	87,493	52,510
	B. Class size					
	Italy (1)	North/Centre (2)	South (3)			
Maimonides' Rule (f_{igkt})	0.513***	0.555***	0.433***			
	(0.0006)	(0.0008)	(0.0011)			
Monitor at institution (M_{igkt})	0.013	0.032	-0.009			
	(0.024)	(0.027)	(0.045)			
N	140,010	87,498	52,512			

↩

Table 7. IV/ 2SLS Estimates of the Effect of Class Size and Score Manipulation on Test Scores

	IV/2SLS			IV/2SLS (overidentified)			IV/2SLS (overidentified-interacted)		
	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	0.0075	-0.0029	0.0062	0.0024	-0.0113	0.0133	0.0116	0.0136	0.0473
	(0.0213)	(0.0298)	(0.0441)	(0.0190)	(0.0251)	(0.0378)	(0.0316)	(0.0482)	(0.0675)
Score manipulation	3.82***	7.33***	2.88***	3.82***	7.02***	2.87***	4.10***	9.21**	3.33***
	(0.19)	(0.79)	(0.16)	(0.19)	(0.73)	(0.16)	(0.96)	(4.41)	(0.86)
Class size * Score manipulation							-0.1464	-1.2700	-0.2273
							(0.4814)	(2.1598)	(0.4304)
Overid test [P-value]				[0.914]	[0.600]	[0.541]	[0.914]	[0.475]	[0.476]
N	139,996	87,491	52,505	139,996	87,491	52,505	139,996	87,491	52,505
					B. Language				
Class size	0.0121	0.0049	0.0127	0.0218	0.0109	0.0491	0.0325	0.0098	0.1337*
	(0.0173)	(0.0196)	(0.0385)	(0.0153)	(0.0174)	(0.0329)	(0.0308)	(0.0320)	(0.0800)
Score manipulation	3.29***	4.50***	2.80***	3.21***	4.34***	2.74***	3.59***	4.31*	4.18***
	(0.18)	(0.45)	(0.18)	(0.18)	(0.42)	(0.18)	(1.03)	(2.25)	(1.30)
Class size * Score manipulation							-0.2130	-0.0029	-0.7058
							(0.4980)	(1.0898)	(0.6214)
Overid test (P-value)				[0.129]	[0.796]	[0.036]	[0.216]	[0.844]	[0.109]
N	140,003	87,493	52,510	140,003	87,493	52,510	140,003	87,493	52,510

Table 8. Maimonides' Rule and Covariate Balance

	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 on Schools			
% in class sitting the test	0.9392	0.0000	0.9345	0.0001	0.9471	0.0000
	[0.0643]	(0.0001)	[0.0657]	(0.0001)	[0.061]	(0.0001)
% in school sitting the test	0.9386	0.0001	0.9339	0.0001	0.9464	0.0001
	[0.0534]	(0.0001)	[0.0548]	(0.0001)	[0.05]	(0.0001)
% in institution sitting the test	0.9374	-0.0001	0.9327	-0.0001	0.9451	-0.0000
	[0.0436]	(0.0001)	[0.0426]	(0.0001)	[0.0441]	(0.0001)
			B. Data Provided by School Staff			
Female	0.482	0.0000	0.4836	0.0002	0.4792	-0.0002
	[0.1205]	(0.0002)	[0.1176]	(0.0002)	[0.1251]	(0.0003)
Immigrant	0.0981	-0.0007***	0.1375	-0.0007***	0.0324	-0.0004***
	[0.1198]	(0.0002)	[0.1298]	(0.0003)	[0.0572]	(0.0001)
Father HS	0.2546	0.0006**	0.2613	0.0002	0.2434	0.0013***
	[0.1678]	(0.0003)	[0.1626]	(0.0003)	[0.1755]	(0.0005)
Mother employed	0.4503	0.0012***	0.5356	0.0010*	0.3082	0.0016***
	[0.2658]	(0.0004)	[0.2574]	(0.0005)	[0.2138]	(0.0006)
			C. Non-Response Indicators			
Missing data on father's education	0.2187	0.0003	0.2216	0.0015**	0.2139	-0.0018*
	[0.3361]	(0.0006)	[0.3358]	(0.0007)	[0.3367]	(0.0010)
Missing data on mother's occupation	0.1925	0.0002	0.1963	0.0014**	0.1861	-0.0019*
	[0.3239]	(0.0006)	[0.3231]	(0.0007)	[0.3251]	(0.0010)
Missing data on country of origin	0.0296	-0.0001	0.0232	-0.0001	0.0401	-0.0000
	[0.1544]	(0.0002)	[0.1361]	(0.0003)	[0.1804]	(0.0005)
N	140,010		87,498		52,512	

Table 9: Testing Alternative Models of Manipulation

	Sicily (1)	South (2)	Sicily (3)	South (4)	Sicily (5)	South (6)
			A. Math			
Percent correct (p_j)	0.698***	0.769***	0.643***	0.713***	0.725***	0.792***
	(0.017)	(0.015)	(0.109)	(0.090)	(0.021)	(0.018)
Percent correct squared (p_j^2)			0.047	0.047		
			(0.086)	(0.071)		
Open (e_j)					0.040	0.038*
					(0.024)	(0.020)
Percent correct (p_j) * open (e_j)					-0.066*	-0.054*
					(0.035)	(0.029)
N	229	1832	229	1832	229	1832
			B. Language			
Percent correct (p_j)	0.790***	0.829***	0.650***	0.735***	0.812***	0.851***
	(0.020)	(0.017)	(0.132)	(0.113)	(0.019)	(0.015)
Percent correct squared (p_j^2)			0.107	0.072		
			(0.092)	(0.078)		
Open (e_j)					0.094**	0.100***
					(0.038)	(0.030)
Percent correct (p_j) * open (e_j)					-0.115**	-0.116***
					(0.047)	(0.037)
N	314	2,512	314	2,512	314	2,512

Table A1. Reduced Form Estimates of the Effect of Maimonides' Rule on Class Size, Test Scores, and Score Manipulation

	Math			Language		
	Italy	North/Centre	South	Italy	North/Centre	South
	(1)	(2)	(3)	(4)	(5)	(6)
		A. Class size				
Maimonides' Rule	0.513***	0.555***	0.433***			
	(0.006)	(0.008)	(0.011)			
Means	19.88	20.07	19.58			
(sd)	(3.58)	(3.52)	(3.64)			
N	140,010	87,498	52,512			
			B. Test Scores			
Maimonides' Rule	-0.0031***	-0.0023**	-0.0056**	-0.0021***	-0.0012	-0.0041**
	(0.0010)	(0.0009)	(0.0022)	(0.0008)	(0.0008)	(0.0017)
Means	0.007	-0.074	0.141	0.01	-0.005	0.035
(sd)	(0.637)	(0.502)	(0.796)	(0.523)	(0.428)	(0.649)
N	140,010	87,498	52,512	140,010	87,498	52,512
			C. Score Manipulation			
Maimonides' Rule	-0.0009***	-0.0003	-0.0020**	-0.0008**	-0.0003	-0.0016**
	(0.0004)	(0.0002)	(0.0009)	-0.0003	-0.0003	-0.0008
Means	0.065	0.02	0.139	0.055	0.023	0.110
(sd)	(0.246)	(0.139)	(0.346)	(0.229)	(0.149)	(0.313)
N	139,996	87,491	52,505	140,003	87,493	52,510

Table A2. First Stage Estimates for Over-Identified Models

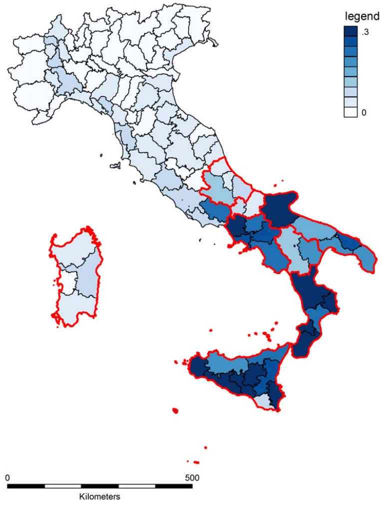
	Class size			Score manipulation math			Score manipulation language		
	Italy	North/Centre	South	Italy	North/Centre	South	Italy	North/Centre	South
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Maimonides' Rule (f_{igkt})	0.704***	0.753***	0.617***	-0.0009**	-0.0003	-0.0021*	-0.0014***	-0.0008**	-0.0024**
	(0.0059)	(0.0069)	(0.0107)	(0.0005)	(0.0003)	(0.0011)	(0.0004)	(0.0003)	(0.0010)
Monitor at institution (M_{igkt})	0.010	0.029	-0.013	-0.029***	-0.010***	-0.062***	-0.025***	-0.012***	-0.047***
	(0.023)	(0.026)	(0.044)	(0.002)	(0.001)	(0.004)	(0.002)	(0.001)	(0.004)
2 students below cutoff	-1.427***	-1.154***	-1.865***	0.002	-0.002	0.008	0.010**	0.005	0.018
	(0.083)	(0.101)	(0.138)	(0.005)	(0.003)	(0.012)	(0.005)	(0.004)	(0.011)
1 student below cutoff	-2.258***	-2.053***	-2.580***	0.001	0.001	0.000	0.007	0.009**	0.002
	(0.093)	(0.116)	(0.150)	(0.005)	(0.004)	(0.012)	(0.005)	(0.004)	(0.011)
1 student above cutoff	2.411***	3.026***	1.519***	0.000	0.003	-0.004	-0.001	-0.001	-0.001
	(0.097)	(0.132)	(0.138)	(0.006)	(0.005)	(0.013)	(0.005)	(0.004)	(0.012)
2 students above cutoff	1.247***	1.546***	0.826***	0.001	-0.004	0.007	-0.007	-0.005	-0.012
	(0.083)	(0.114)	(0.120)	(0.006)	(0.004)	(0.013)	(0.005)	(0.004)	(0.009)
N	140,010	87,498	52,512	139,996	87,491	52,505	140,003	87,493	52,510

↩

Table A3. Covariate Balance in Maimonides' Rule for Institutions with and without External Monitor

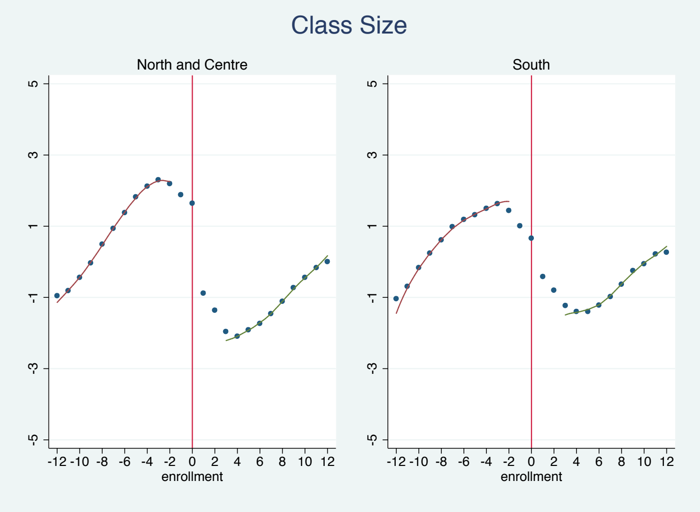
	Institutions with Monitor			Institutions without Monitor		
	Italy	North/Centre	South	Italy	North/Centre	South
	(1)	(2)	(3)	(4)	(5)	(6)
	A. Administrative Data on Schools					
% in class sitting the test	0.0001	0.0002	0.0000	0.0000	0.0000	0.0000
	(0.0002)	(0.0002)	(0.0003)	(0.0001)	(0.0001)	(0.0002)
% in school sitting the test	0.0003	0.0003	0.0002	0.0001	0.0001	0.0001
	(0.0002)	(0.0002)	(0.0003)	(0.0001)	(0.0001)	(0.0002)
% in institution sitting the test	-0.0000	-0.0000	0.0001	-0.0001*	-0.0002*	-0.0000
	(0.0001)	(0.0002)	(0.0003)	(0.0001)	(0.0001)	(0.0001)
	B. Data Provided by School Staff					
Female	-0.0003	-0.0006	0.0001	0.0001	0.0005*	-0.0003
	(0.0003)	(0.0004)	(0.0006)	(0.0002)	(0.0002)	(0.0003)
Immigrant	-0.0005	-0.0002	-0.0007**	-0.0007***	-0.0009***	-0.0003*
	(0.0003)	(0.0005)	(0.0003)	(0.0002)	(0.0003)	(0.0002)
Father HS	-0.0005	-0.0002	-0.0014	0.0010***	0.0003	0.0020***
	(0.0005)	(0.0006)	(0.0010)	(0.0003)	(0.0004)	(0.0005)
Mother employed	0.0001	0.0003	-0.0004	0.0015***	0.0012**	0.0022***
	(0.0008)	(0.0010)	(0.0012)	(0.0004)	(0.0006)	(0.0006)
	C. Non-Response Indicators					
Missing data on father's education	0.0014	0.0012	0.0019	0.0000	0.0016**	-0.0026**
	(0.0011)	(0.0013)	(0.0020)	(0.0007)	(0.0008)	(0.0012)
Missing data on mother's occupation	0.0018*	0.0017	0.0020	-0.0002	0.0012	-0.0028**
	(0.0011)	(0.0013)	(0.0019)	(0.0007)	(0.0008)	(0.0011)
Missing data on country of origin	0.0006	0.0003	0.0011	-0.0002	-0.0002	-0.0003
	(0.0004)	(0.0004)	(0.0008)	(0.0003)	(0.0003)	(0.0006)
N	34,325	22,174	12,151	105,685	65,324	40,361

Score Manipulation by Province

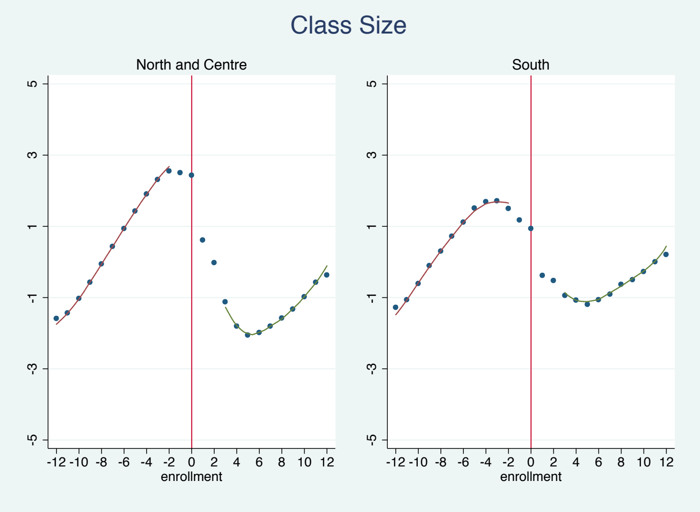



Predicted vs Actual Class Size
grade 2 (before 2010)
30
25
20
15
10
25
50
75
100
125
150
enrollment
grade 5
30
25
20
15
10
25
50
75
100
125
150
enrollment
grade 2 (after 2010)
30
25
20
15
10
27
54
81
108
135
enrollment

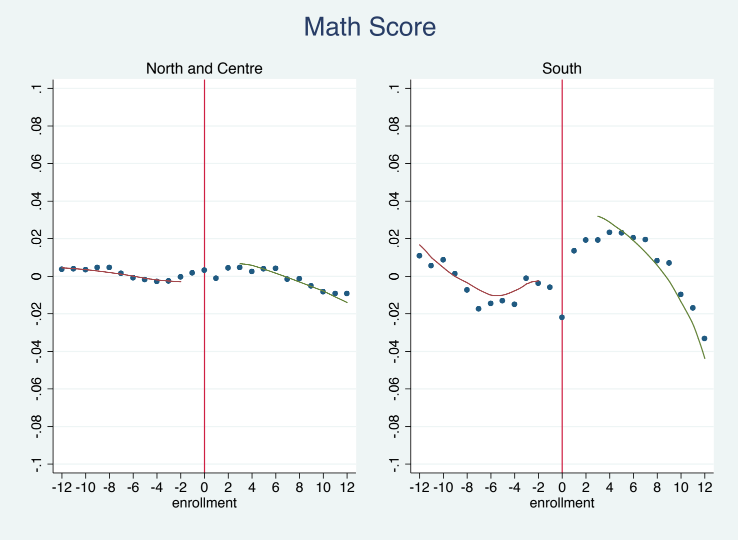
Class Size Around Cutoffs: Grade 2



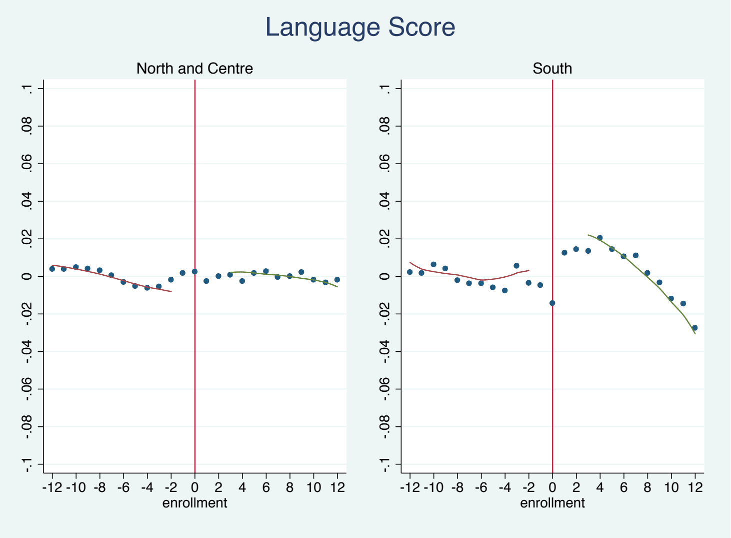
Class Size Around Cutoffs: Grade 5



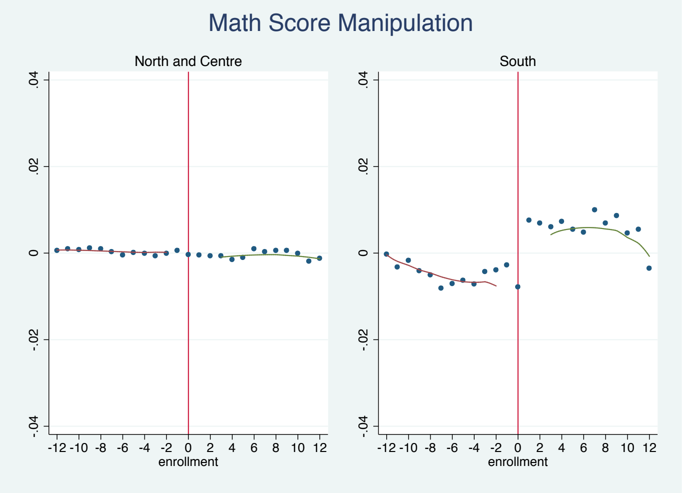
Math Scores Around Cutoffs



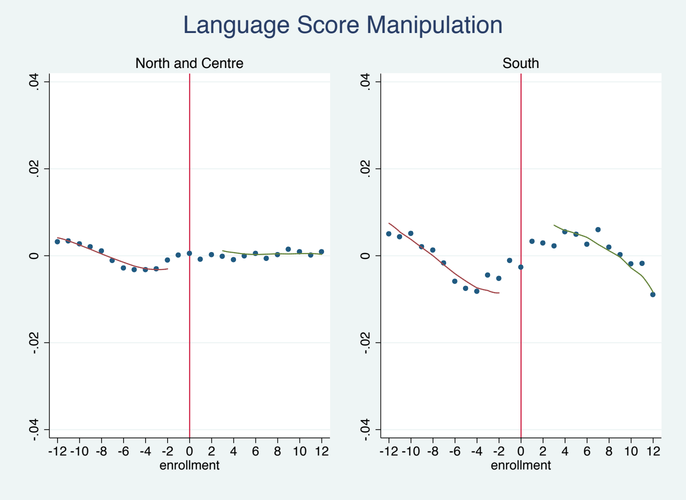
Language Scores Around Cutoffs



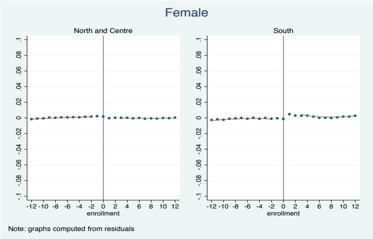
Math Score Manipulation

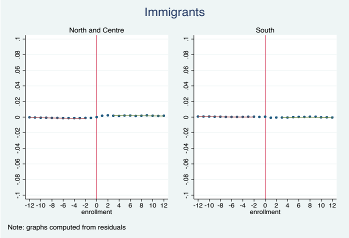


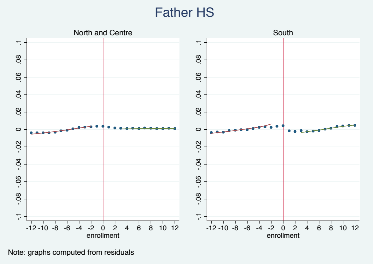
Language Score Manipulation

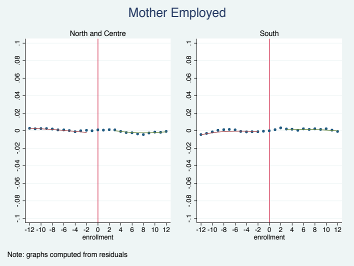


Covariates


Female
North and Centre
South
enrollment
enrollment
Note: graphs computed from residuals


Immigrants
North and Centre
South
enrollment
enrollment
Note: graphs computed from residuals


Father HS
North and Centre
South
enrollment
enrollment
Note: graphs computed from residuals


Mother Employed
North and Centre
South
enrollment
enrollment
Note: graphs computed from residuals

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
A2	☐A	☐B	☐C	☐D	☐NV
A3	☐A	☐B	☐C	☐D	☐NV
A4	☐A	☐B	☐C	☐D	☐NV
A5	☐A	☐B	☐C	☐D	☐NV
A6	☐A	☐B	☐C	☐D	☐NV
A7	☐A	☐B	☐C	☐D	☐NV
A8	☐A	☐B	☐C	☐D	☐NV
A9	☐A	☐B	☐C	☐D	☐NV
A10	☐A	☐B	☐C	☐D	☐NV
A11	☐A	☐B	☐C	☐D	☐NV
A12	☐A	☐B	☐C	☐D	☐NV
A13	☐A	☐B	☐C	☐D	☐NV
A14	☐A	☐B	☐C	☐D	☐NV
A15	☐A	☐B	☐C	☐D	☐NV
A16	☐A	☐B	☐C	☐D	☐NV
A17	☐A	☐B	☐C	☐D	☐NV
B1	☐A	☐B	☐C	☐D	☐NV
B2	☐A	☐B	☐C	☐D	☐NV
B3	☐A	☐B	☐C	☐D	☐NV
B4	☐A	☐B	☐C	☐D	☐NV
B5	☐A	☐B	☐C	☐D	☐NV
B6	☐A	☐B	☐C	☐D	☐NV
B7	☐A	☐B	☐C	☐D	☐NV
B8	☐A	☐B	☐C	☐D	☐NV
B9	☐A	☐B	☐C	☐D	☐NV
B10	☐A	☐B	☐C	☐D	☐NV
B11	☐A	☐B	☐C	☐D	☐NV
B12	☐A	☐B	☐C	☐D	☐NV
B13	☐A	☐B	☐C	☐D	☐NV
B14	☐A	☐B	☐C	☐D	☐NV
B15	☐A	☐B	☐C	☐D	☐NV

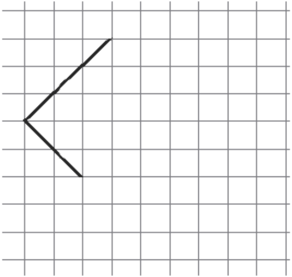
C1_a1	☐0	☐1	☐NV		
C1_a2	☐0	☐1	☐NV		
C1_b1	☐0	☐1	☐NV		
C1_b2	☐0	☐1	☐NV		
C1_b3	☐0	☐1	☐NV		
C2	☐A	☐B	☐C	☐D	☐NV
C3_a	☐Nome	☐Non_Nome	☐NV		
C3_b	☐Nome	☐Non_Nome	☐NV		
C3_c	☐Nome	☐Non_Nome	☐NV		
C3_d	☐Nome	☐Non_Nome	☐NV		
C3_e	☐Nome	☐Non_Nome	☐NV		
C3_f	☐Nome	☐Non_Nome	☐NV		
C3_g	☐Nome	☐Non_Nome	☐NV		
C3_h	☐Nome	☐Non_Nome	☐NV		
C3_i	☐Nome	☐Non_Nome	☐NV		
C3_l	☐Nome	☐Non_Nome	☐NV		
C3_m	☐Nome	☐Non_Nome	☐NV		
C3_n	☐Nome	☐Non_Nome	☐NV		
C3_o	☐Nome	☐Non_Nome	☐NV		
C3_p	☐Nome	☐Non_Nome	☐NV		
C3_q	☐Nome	☐Non_Nome	☐NV		
C3_r	☐Nome	☐Non_Nome	☐NV		
C4	☐0	☐1	☐NV		
C5	☐0	☐1	☐NV		
C6	☐A	☐B	☐C	☐D	☐NV
C7	☐0	☐1	☐NV		
C8	☐A	☐B	☐C	☐D	☐NV
C9	☐A	☐B	☐C	☐D	☐NV
C10	☐0	☐1	☐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**)

D23. Osserva la seguente figura.



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

b. Spiega come hai fatto per disegnare il quadrato.

....

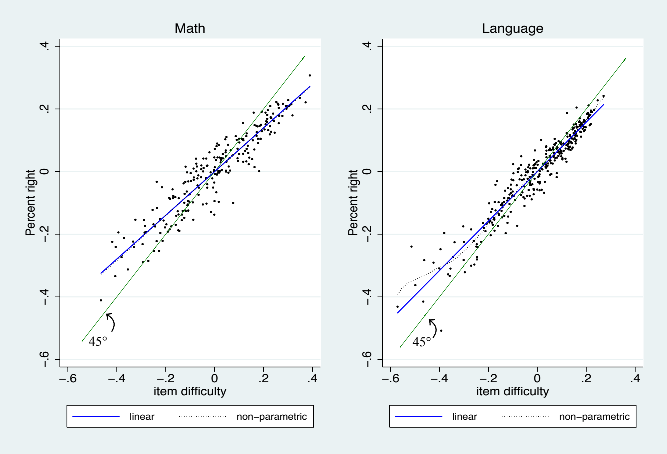
....

....

C4. Nella frase che segue inserisci le parole mancanti scegliendole da questa lista: *così, dove, perché, però, se, siccome*.

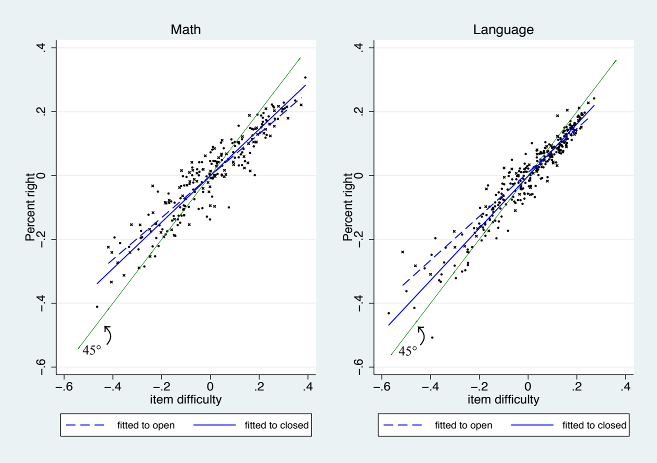
……………. non conoscevo la strada, ho chiesto a una signora ……… dovevo andare; …………….. non mi sono perso.

Looking for Nonlinearity



Notes: The figure plots average percent correct by item in Sicily against average percent correct in Veneto.

The Effect of Grading Effort



Notes: The figure plots average percent correct by item in Sicily against average percent correct in Veneto, with linear fit of the lines separately by item grading effort. Points plotted with a "×" refer to open question, points plotted with a "•" refer to closed questions.

Why Manipulate? An Item-Level Analysis

- For item j, let $1-p_j$ be difficulty, e_j be (Bernoulli) teacher grading effort, and m_{ij} an indicator for manipulation in class i. Manipulators score $g\left(e_j\right)$. Class i's percent correct on item j is:

$$y_{ij} = p_j + (g(e_j) - p_j)m_{ij} + \upsilon_{ij}$$

- Accountability concerns (dishonesty related to item difficulty): $m_{ij} = \kappa_0 + \kappa_1 p_j$, where $\kappa_1 < 0$ and $g\left(e_j\right) = \gamma_0$, implying

$$y_{ij} = \gamma_0 \kappa_0 + [\gamma_0 \kappa_1 + (1-\kappa_0)]p_j - \kappa_1 p_j^2 + \upsilon_{ij}$$

- Selective shirking & sloppiness (curbstone open items, perhaps less accurately): $m_{ij} = \kappa_0 + \kappa_1 e_j$, where $\kappa_1 > 0$ and $g(e_j) = \gamma_0 + \gamma_1 e_j$, where $\gamma_1 < 0$, implying

$$y_{ij} = \kappa_0 \gamma_0 + (\kappa_0 \gamma_1 + \kappa_1 \gamma_0 + \kappa_1 \gamma_1) e_j + (1-\kappa_0) p_j - \kappa_1 p_j e_j + \upsilon_{ij}$$

- Curbstoning (shirking unrelated to item difficulty and grading effort): $m_{ij} = \kappa_0$ and $g\left(e_j\right) = \gamma_0$ implying

$$y_{ij} = \gamma_0 \kappa_0 + (1-\kappa_0) p_j + \upsilon_{ij}$$

Background

- Families apply for school admission in February of the previous year in which their child is starting school or they wish to transfer
- Parents can apply to only one school in the province of residence. Applicants are accepted before the summer
- In cases of over-subscription, distance usually determines who has a first claim on seats
- Parents learn about class composition only in September, shortly before school starts
- Mobility across schools is limited after class formation because of administrative burdens and little negotiation power with the school principal