

When the Cat Is Near, the Mice Won't Play: The Effects of External Examiners in Italian Schools

Marco Bertoni

Università di Padova e London School of Economics

Giorgio Brunello

Università di Padova, CESifo e IZA

Lorenzo Rocco

Università di Padova

Workshop INVALSI

**“Metodi di identificazione, analisi
e trattamento del cheating”**

Roma, 8 Febbraio 2013

Introduzione

- I sistemi di *accountability* dell'istruzione basati su test standardizzati sono sempre più diffusi.
- **Problema:** meccanismi incentivanti con target definiti molto precisamente portano gli agenti a rispondere in modo strategico.

Introduzione

- Jacob & Levitt, 2003

Chicago: una frazione di insegnanti **cambia le risposte degli studenti** ai test in seguito all'introduzione di *high stakes tests*.

Ri-testati in un setting controllato, i risultati scendono drasticamente

- Figlio & Winicki, 2005

Florida: le scuole aumentano il **contenuto calorico dei menu della mensa** nei giorni di test.

- Bhokari & Schneider, 2011

US: in Stati dove le conseguenze dei test sono più alte è più alta la frequenza di diagnosi di ADHD e la prescrizione di **farmaci psico-stimolanti**

- Figlio, 2005 - Jacob, 2005 - Cullen & Reback, 2006

US: studenti peggiori **riclassificati come disabili, contati come assenti o sospesi** per non includerli nei partecipanti al test e migliorare i risultati

Introduzione

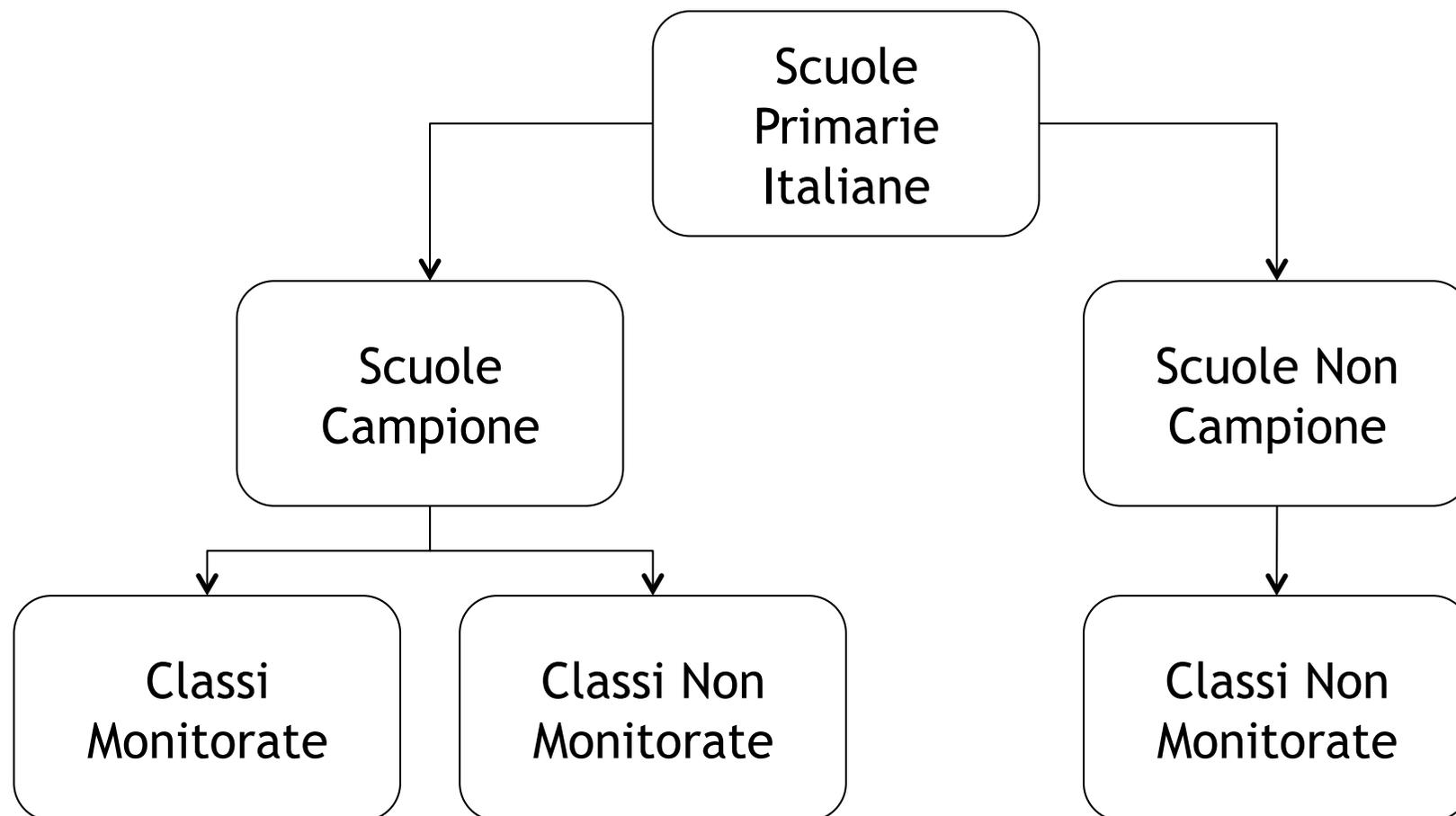
- A meno che...



Gli effetti del monitoraggio

- Questo studio sfrutta un **esperimento naturale** presente nell'indagine degli apprendimenti INVALSI SNV per la scuola primaria.
- **Assegnazione casuale di osservatori esterni** ad un campione di classi, per assicurare la corretta implementazione delle procedure di verifica degli apprendimenti

Monitoraggio nell'indagine INVALSI



Domande di ricerca

1. Valutazione dell'efficacia del monitoraggio - **effetti diretti**
2. Ci sono degli **effetti indiretti** del monitoraggio sulle classi non supervisionate della stessa scuola?
3. **Quali fattori** sono associati agli effetti del monitoraggio?

Risultati principali

- Si riscontrano **effetti diretti ed indiretti del monitoraggio negativi** e statisticamente significativi sul risultato medio delle classi, più forti nel **Sud Italia**
 - **Effetti diretti**: riduzione del *cheating/copying* nelle classi supervisionate
 - **Effetti indiretti**: possibili spiegazioni sono la paura che l'osservatore possa girare per le classi o un interesse verso la performance relativa delle diverse classi di una singola scuola.
- Differenze territoriali negli effetti del monitoraggio sono associate a **differenti dotazioni di capitale sociale**

L'indagine INVALSI SNV

- **Chi è interessato?**
 - Scuole Primarie - Classi seconde e quinte
 - 2009/2010: prima **indagine censuaria** - dati sull'intera popolazione di studenti
- **Che cosa indaga?**
 - Scopo: **misurare il valore aggiunto nelle scuole**
 - Strumenti: test standardizzati in matematica e italiano
 - **La posta in gioco è dubbia**: idealmente test come cartine al tornasole per l'autovalutazione della performance delle scuole, nella pratica conseguenze per le scuole e i docenti non chiare, specie per SNV 2009/2010
 - **Conseguenze reputazionali**: i risultati non sono pubblici, ma le scuole ne sono a conoscenza e possono renderli pubblici

Osservatori esterni

Rapporto INVALSI → Necessità di garantire la corretta implementazione delle procedure di verifica → Selezione di un campione di classi in cui i test vengono somministrati in presenza di osservatori esterni

- **L'osservatore:**

- E' presente durante la somministrazione del test, **riporta e trasmette le risposte degli studenti** a INVALSI

- **Nelle altre classi:**

- I test vengono somministrati da docenti della scuola, ma non della classe e della materia della prova
- **Trascrizione e trasmissione delle risposte a carico dei docenti**

Campionamento

- **Disegno di campionamento a due stadi** - stratificato per regione

1. Selezione delle **scuole**

- Campionamento probabilistico - probabilità di inclusione proporzionale al numero di studenti nelle classi interessate.
- In totale, 1385 scuole campionate

2. Selezione delle **classi** nelle scuole campionate

- Per ogni scuola trattata, selezione di una o due classi per grado, a seconda della dimensione della scuola. In totale, 2000 classi campionate
- Campionamento casuale delle scuole entro ogni scuola trattata

Campionamento

	Totale scuole	Totale classi	Totale studenti	Scuole campione	Classi campione	Studenti campione
Classi Seconde	7,700	30,175	555,347	1,385	2,000	39,299
Classi Quinte	7,700	30,476	565,064	1,385	2,000	39,643

Identificazione

- **Trattamento bi-dimensionale**, a livello di scuola e classe: S_j, C_i
- Risultati potenziali di classe:

$$Y_{(C=1,S=1)}, Y_{01}, Y_{00}$$

Randomizzazione condizionata

- Definiamo RD = dummy regionali
 $RS = RD * \#$ studenti nelle classi 2^e / 5^e
 $RC = RD * \#$ classi 2^e / 5^e
 $R = [RD, RS, RC]$
- Dato lo schema di campionamento:

$$Y_{11}, Y_{01}, Y_{00} \perp S_j, C_i \mid R$$

Test di bilanciamento

- Verifichiamo la bontà della randomizzazione ai due stadi del campionamento sfruttando la disponibilità di **variabili pre-determinate**

$$1) X_j = \alpha + \beta t_j + \rho RD_j + \sigma RS_j + \varepsilon_j$$

$$2) X_{ij} = \alpha + \beta t_{ij} + \delta R_{ij} + \varepsilon_{ij}$$

- In generale, la randomizzazione ha avuto successo
- **Controlli sulle osservabili** per migliorare il bilanciamento e la precisione delle stime

Stima

- Modello di Regressione Lineare:

$$Y_{ij} = \alpha + \beta C_i S_j + \gamma S_j + \boldsymbol{\delta}'\mathbf{R} + \boldsymbol{\theta}'\mathbf{X} + \varepsilon_{ij}$$

$$\text{ED: } \beta = E[Y_{ij}|C_i=1, S_j=1, \mathbf{R}, \mathbf{X}] - E[Y_{ij}|C_i=0, S_j=1, \mathbf{R}, \mathbf{X}]$$

$$\text{EI: } \gamma = E[Y_{ij}|C_i=0, S_j=1, \mathbf{R}, \mathbf{X}] - E[Y_{ij}|C_i=0, S_j=0, \mathbf{R}, \mathbf{X}]$$

$$\text{ET: } \beta + \gamma = E[Y_{ij}|C_i=1, S_j=1, \mathbf{R}, \mathbf{X}] - E[Y_{ij}|C_i=0, S_j=0, \mathbf{R}, \mathbf{X}]$$

- Dati medi di classe pesati per la dimensione della classe
- Standard error corretti per *clustering* a livello di scuola

Risultati principali

- Score medio di classe - Matematica - Classi V
- Y = % risposte corrette

	(1) Italy	(2) North	(3) Centre	(4) South
Direct Effect	-2.79*** (0.25)	-0.99*** (0.28)	-2.27*** (0.48)	-4.92*** (0.50)
Indirect Effect	-0.81*** (0.28)	-0.70*** (0.27)	-0.73 (0.45)	-1.04* (0.61)
Overall Effect	-3.59*** (0.29)	-1.69*** (0.31)	-2.99*** (0.54)	-5.96*** (0.60)
Observations	27,325	11,541	4,886	10,898
R-squared	0.15	0.2	0.15	0.14
Additional covariates	Yes	Yes	Yes	Yes
Mean - Untreated Schools	65.1	63.9	64.0	66.8

Risultati principali

- Condizioni psicologiche degli studenti

	(1) I was already anxious before starting the test	(2) I was so nervous I couldn't find the answers	(3) While answering , I felt like I was doing badly	(4) While answering, I was calm
Direct Effect	0.25 (0.42)	-0.92*** (0.29)	-0.08 (0.39)	0.64 (0.39)
Indirect Effect	0.25 (0.31)	0.01 (0.21)	0.36 (0.28)	-0.01 (0.29)
Overall Effect	0.50 (0.41)	-0.90*** (0.28)	0.28 (0.38)	0.63* (0.38)
Observations	27,141	27,142	27,141	27,140
R-squared	0.07	0.11	0.1	0.07
Additional covariates	Yes	Yes	Yes	Yes
Mean - Untreated Schools	61.0	19.2	50.7	53.1

Risultati principali

- Distribuzione degli score entro classe

	(1) Standard Deviation	(2) Coefficient of Variation	(3) Bottom quartile	(4) Median	(5) Top quartile
Direct Effect	0.76*** (0.09)	2.14*** (0.21)	-3.70*** (0.31)	-3.07*** (0.29)	-2.26*** (0.27)
Indirect Effect	0.03 (0.08)	0.30 (0.18)	-0.55* (0.31)	-0.56* (0.29)	-0.61** (0.26)
Overall Effect	0.79*** (0.09)	2.44*** (0.22)	-4.26*** (0.33)	-3.63*** (0.32)	-2.88*** (0.30)
Observations	27,325	27,325	27,325	27,325	27,325
R-squared	0.18	0.15	0.12	0.1	0.09
Additional covariates	Yes	Yes	Yes	Yes	Yes
Mean - Untreated Schools	14.1	22.8	55.6	65.6	75.2

Risultati principali

- Eterogeneità nei pattern di risposte entro classe
- Gli studenti danno le stesse risposte?

- Herfindal Index

$$H = \frac{1 - \sum_{a=1}^A s_a^2}{1 - \frac{1}{A}}$$

- Range: 0-1. Valori più elevati → Maggiore eterogeneità

Risultati principali

- Eterogeneità nei pattern di risposte entro classe
- $Y = \text{Herfindal Index}$

	(1) Italy	(2) North	(3) Centre	(4) South
Direct Effect	3.93*** (0.32)	1.24*** (0.32)	2.63*** (0.60)	7.32*** (0.64)
Indirect Effect	0.82** (0.34)	0.64** (0.31)	0.51 (0.58)	1.22* (0.73)
Overall Effect	4.75*** (0.35)	1.88*** (0.35)	3.14*** (0.62)	8.54*** (0.719)
Observations	27,325	11,541	4,886	10,898
R-squared	0.2	0.17	0.13	0.15
Additional covariates	Yes	Yes	Yes	Yes
Mean - Untreated Schools	52.8	57.3	55.7	46.9

Sensitività

- GLM - frazione come variabile dipendente
Papke e Wooldridge (1996)
- Correzione per popolazione finita
- Esclusione covariate
- Differenze nelle assenze



Estensioni

Quali fattori sono legati agli effetti del monitoraggio?

- **Interazione** delle variabili di trattamento con:
 - Dimensione della classe
 - Percentuale di insegnanti di ruolo nella scuola
 - Media indice ESCS nella classe (background familiare)
- **Rank correlation** entro classe tra score INVALSI e voto semestre precedente più elevata nelle classi monitorate

Estensioni

	(1)	(2)	(3)
	Interacted with Class Size	Interacted with % Tenured Teachers	Interacted with ESCS
Direct Effect	-3.41*** (0.41)	-1.34*** (0.29)	-2.65*** (0.33)
Interacted Direct Effect	0.98* (0.53)	-2.98*** (0.50)	-0.15 (0.54)
Indirect Effect	-0.94*** (0.36)	-0.66** (0.29)	-0.67** (0.31)
Interacted Indirect Effect	0.22 (0.41)	-0.19 (0.54)	-0.30 (0.44)
Overall Effect	-4.35*** (0.43)	-2.00*** (0.33)	-3.32*** (0.36)
Interacted Overall Effect	1.20** (0.51)	-3.17*** (0.57)	-0.45 (0.51)
Observations	27,325	26,313	27,323
R-squared	0.15	0.15	0.15
Additional covariates	Yes	Yes	Yes
Mean - Untreated Schools	65.1	64.9	65.1

Eterogeneità regionale

- Perché gli effetti del monitoraggio sono **più forti nel Sud Italia?**

- Ci può essere un **legame negativo** tra gli effetti del **monitoraggio** e il **capitale sociale** locale...

“Civic capital consists of those persistent and shared beliefs and values that help a group overcome free rider outcomes”. (GSZ, 2010)

- Capitale sociale **bridging vs. bonding**

Capitale sociale

- Verifichiamo questa ipotesi utilizzando **misure indirette di capitale sociale a livello provinciale (GSZ, 2004)**
 - **Donazioni di sangue**/100.000 abitanti (1995);
 - Partecipazione media ai **referendum** (1946-1989).
- Per evitare distorsioni da variabili omesse, includiamo nei modelli il **PIL** pro capite e il **tasso di disoccupazione** a livello provinciale



Capitale sociale

	(1)	(2)	(3)	(4)	(5)
	Baseline	Interacted with Blood Donations	Interacted with Blood Donations and Macro Variables	Interacted with Turnover at Referenda	Interacted with Turnover at Referenda and Macro Variables
Direct Effect	-2.78*** (0.25)	-2.48*** (0.24)	-2.64*** (0.26)	-2.63*** (0.24)	-2.69*** (0.25)
Interacted Direct Effect		0.81*** (0.11)	0.41*** (0.12)	0.25*** (0.04)	0.14** (0.06)
Indirect Effect	-0.82*** (0.28)	-0.85*** (0.26)	-0.93*** (0.29)	-0.80*** (0.26)	-0.88*** (0.20)
Interacted Indirect Effect		-0.06 (0.12)	-0.13 (0.13)	0.01 (0.04)	-0.02 (0.07)
Overall Effect	-3.60*** (0.30)	-3.33*** (0.28)	-3.57*** (0.31)	-3.43*** (0.28)	-3.57** (0.30)
Interacted Overall Effect		0.75*** (0.13)	0.28** (0.14)	0.26*** (0.04)	0.12* (0.07)
Observations	27,178	27,178	27,178	27,178	27,178
R-squared	0.15	0.15	0.15	0.15	0.15
Additional covariates	Yes	Yes	Yes	Yes	Yes
Mean - Untreated Schools	65.1	65.1	65.1	65.1	65.1

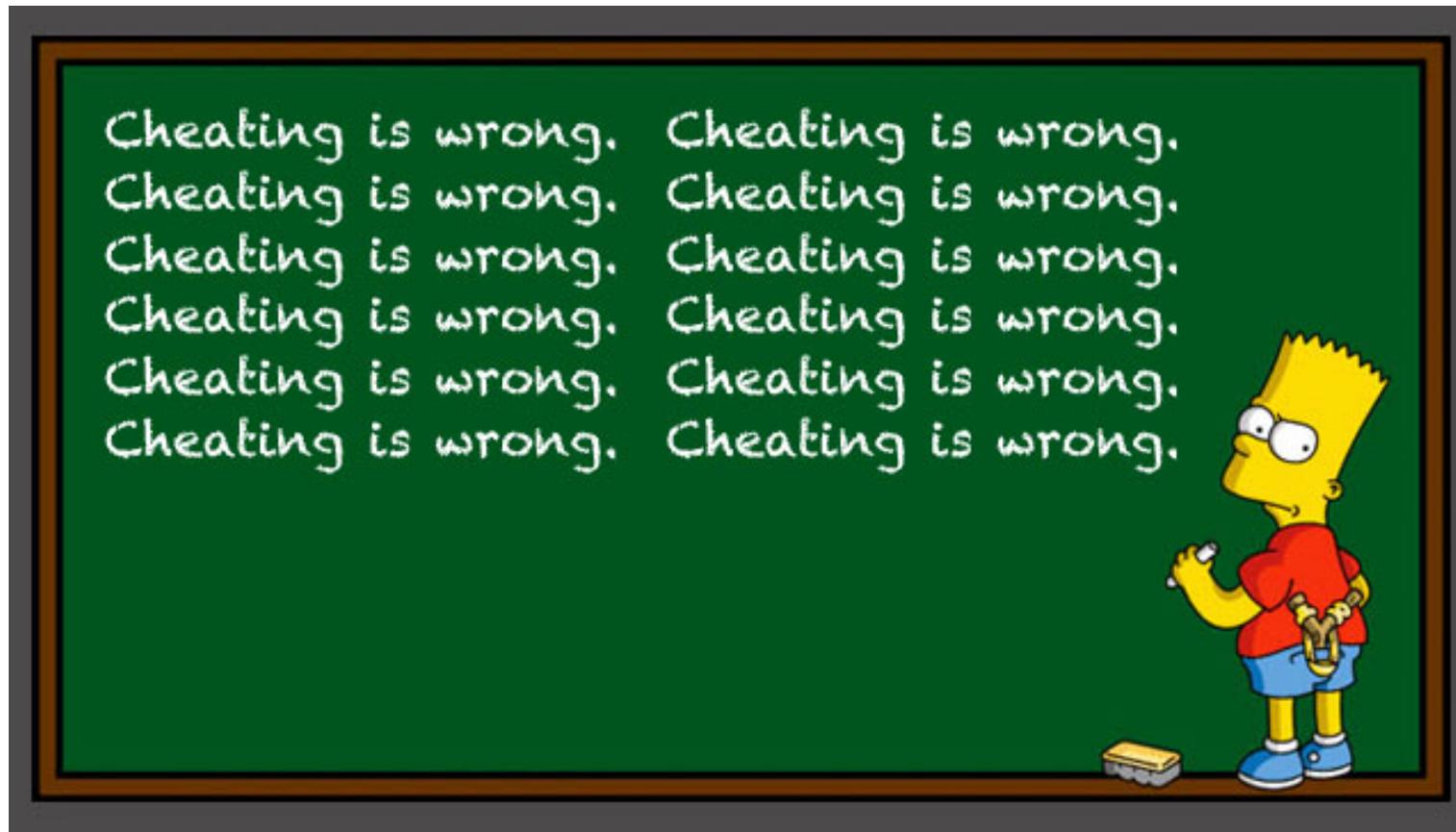
Conclusioni

- Questa analisi studia gli **effetti diretti ed indiretti del monitoraggio nei test standardizzati**, sfruttando un esperimento naturale presente nell'indagine INVALSI SNV.
- Stimiamo **effetti diretti negativi** del monitoraggio sugli score medi delle classi. Interpretiamo questi risultati come **riduzione del cheating/copying** - più forte nel Sud Italia
- **Effetti indiretti e negativi** del monitoraggio sono coerenti con una irrazionale paura che l'osservatore possa "girare per le classi" o con un interesse verso la performance relativa delle classi in una scuola
- L'eterogeneità regionale negli effetti del monitoraggio è negativamente correlata con **misure di capitale sociale**

Discussione

- Valutazione **economicità** utilizzo osservatori
 - **Benefici**: provata efficacia nella riduzione del cheating
 - **Costi**: 200€ per osservatore per giorno
 - Possibili difficoltà nell'estensione all'intera popolazione
- L'utilizzo degli osservatori **non è l'unica alternativa**
 - *Question reshuffling*
 - *Computer-based testing*

Grazie!



GLM

- Score medio di classe - Matematica - Classi V
- Y = % risposte corrette



	(1) Italy	(2) North	(3) Centre	(4) South
Direct Effect	-2.74*** (0.25)	-0.97*** (0.28)	-2.25*** (0.47)	-4.73*** (0.48)
Indirect Effect	-0.80*** (0.28)	-0.70*** (0.27)	-0.72 (0.45)	-1.04* (0.60)
Overall Effect	-3.54*** (0.29)	-1.67*** (0.30)	-2.97*** (0.53)	-5.77*** (0.57)
Observations	27,325	11,541	4,886	10,898
Additional covariates	Yes	Yes	Yes	Yes
Mean - Untreated Schools	65.1	63.9	64.0	66.8

Correzione per popolazioni finite

- **Score medio di classe - Matematica - Classi V**
- **Y = % risposte corrette**



	(1) Italy	(2) North	(3) Centre	(4) South
Direct Effect	-2.89*** (0.12)	-1.08*** (0.14)	-2.35*** (0.23)	-5.05*** (0.24)
Indirect Effect	-0.83*** (0.13)	-0.71*** (0.13)	-0.70*** (0.21)	-1.06*** (0.27)
Overall Effect	-3.72*** (0.14)	-1.79*** (0.14)	-3.05*** (0.25)	-6.11*** (0.28)
Observations	27,325	11,541	4,886	10,898
R-squared	0.15	0.19	0.15	0.15
Additional covariates	Yes	Yes	Yes	Yes
Mean - Untreated Schools	65.1	63.9	64.0	66.8

Esclusione covariate

- Score medio di classe - Matematica - Classi V
- Y = % risposte corrette



	(1) Italy	(2) North	(3) Centre	(4) South
Direct Effect	-2.82*** (0.26)	-0.85*** (0.30)	-2.04*** (0.49)	-5.29*** (0.52)
Indirect Effect	-0.70** (0.30)	-0.82*** (0.31)	-0.46 (0.51)	-0.70 (0.65)
Overall Effect	-3.52*** (0.31)	-1.68*** (0.34)	-2.50*** (0.58)	-5.99*** (0.64)
Observations	27,325	11,541	4,886	10,898
R-squared	0.03	0.01	0.01	0.03
Additional covariates	No	No	No	No
Mean - Untreated Schools	65.1	63.9	64.0	66.8

Assenze

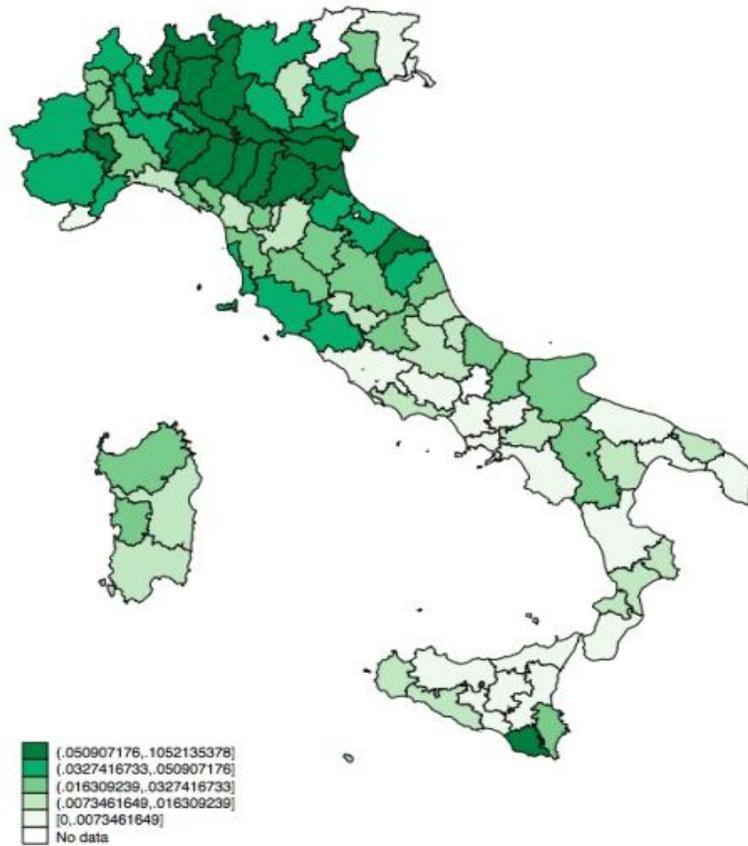
- Score medio di classe - Matematica - Classi V
- $Y = \% \text{ studenti assenti}$



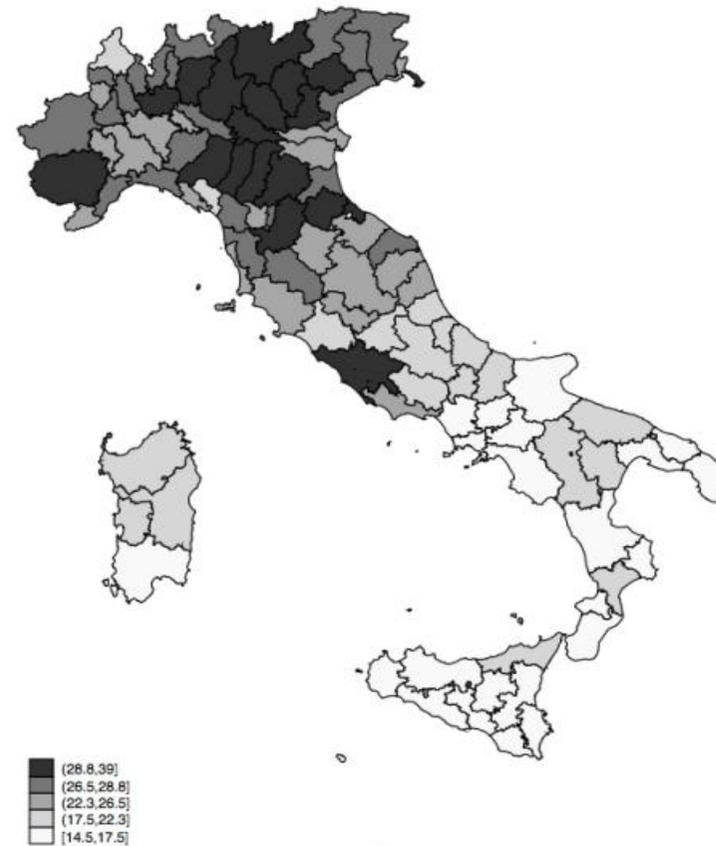
	(1) Italy	(2) North	(3) Centre	(4) South
Direct Effect	-0.53** (0.24)	-0.50 (0.40)	-0.47 (0.47)	-0.55 (0.40)
Indirect Effect	-0.10 (0.24)	0.44 (0.36)	-0.44 (0.42)	-0.51 (0.44)
Overall Effect	-0.63** (0.25)	-0.06 (0.40)	-0.91** (0.46)	-1.06** (0.42)
Observations	27,325	11,541	4,886	10,898
R-squared	0.03	0.02	0.03	0.03
Additional covariates	No	No	No	No
Mean - Untreated Schools	11.0	10.4	11.7	11.4

Capitale sociale

Donazioni di sangue

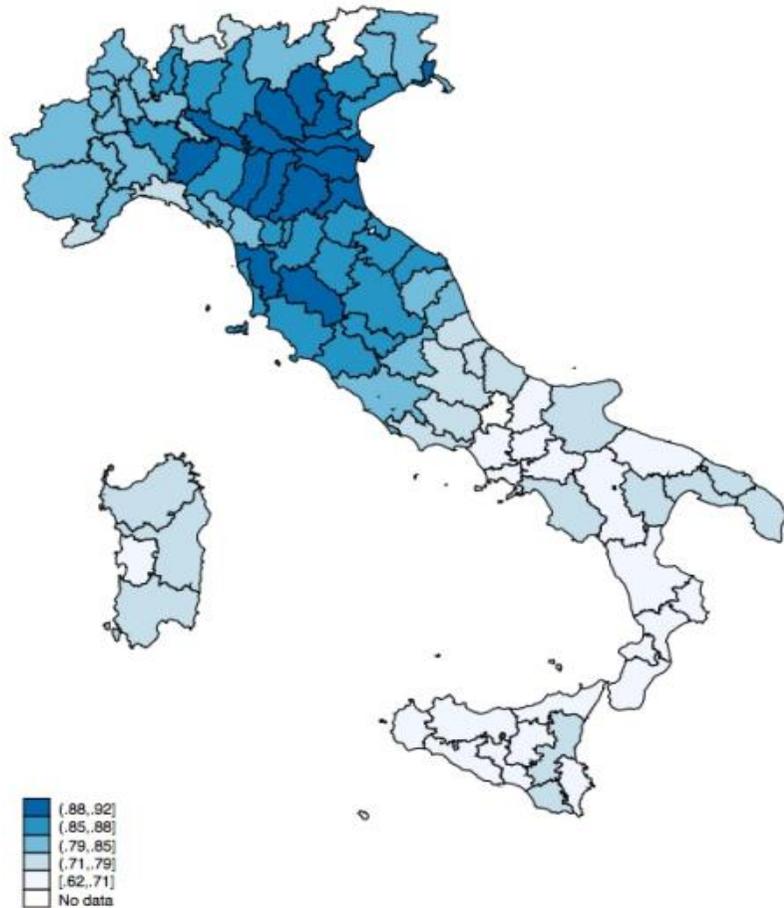


PIL pro capite



Capitale sociale

Partecipazione referendum



Tasso di disoccupazione

