

il 'cheating' nel sistema italiano di valutazione

metodi di rilevamento del cheating basato sui
pattern di risposta non usuali

~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~

i determinanti dei comportamenti opportunistici degli insegnanti

workshop
'metodi di identificazione, analisi e trattamento del cheating'

8 febbraio 2013



gerard ferrer-esteban



contenuti



il rilevamento del cheating

il fenomeno del cheating

un metodo basato sui pattern non usuali

comparazione tra metodi

proposte sul metodo

i determinanti del cheating

dati e modello empirico

risultati: incentivi e deterrenti

proposte per prevenire il cheating

il fenomeno del *cheating*



the more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor (campbell, 1976).

. fenomeno multiforme

. cheating e accountability

. ricerca sul *cheating*

rilevare il *cheating*



rischio nell'individuazione del cheating:

metodi basati sui livelli e deviazione delle performance

** errore tipo i: α o falsi positivi (quando si segnala come cheater una classe che non ha barato)

** errore tipo ii: β o falsi negativi (quando non si identifica una classe che ha barato)

il caso dell'indice invalsi (quintano, castellano, longobardi, 2009)



Indicatore per individuare il cheating (Jacob & Levitt, 2003)

. indicatore di sequenze di risposta sospette (3 misure combinate)

- . pattern più generali di similarità tra le risposte degli studenti (correlazione delle risposte del test)
- . grado di varianza della correlazione delle risposte tra gli item
- . casi nei quali gli studenti mancano item facili e rispondono correttamente agli item difficili

metodo



- misura 1**

pattern più generali di similarità tra le risposte degli studenti

1) calcolo dei residui per ogni possibile risposta che lo studente potrebbe aver fornito in ogni item

2) somma dei residui di ogni risposta j tra gli studenti dentro della classe c

$$e_{jisc} = 0 - \frac{e^{\hat{\beta}_j x_s}}{\sum_{j=1}^J e^{\hat{\beta}_j x_s}} \text{ if } j \neq k$$

$$= 1 - \frac{e^{\hat{\beta}_j x_s}}{\sum_{j=1}^J e^{\hat{\beta}_j x_s}} \text{ if } j = k$$

dove e_{jisc} è il residuo per la risposta j nell'item I per lo studente s nella classe c (4 residui per item)

J è il numero di possibili risposte (2 o 4)
 x è il vettore che comprende caratteristiche studenti, score passati e futuri e background

$$e_{jic} = \sum_s e_{jisc}$$

metodo



- **misura 1**

- 3) somma dei quadrati dei residui delle 4 possibili risposte di ogni singolo item aggregato a livello classe

$$v_{ic} = \frac{\sum_j e_{jic}^2}{ns_c}$$

- 4) media di classe di questa varianza

$$M1_c = \bar{v}_c = \frac{\sum_i v_{ic}}{ni}$$

dove ni è il numero di item nel test



- **misura 2**

varianza nel grado di correlazione tra gli item

- 1) in base alla misura di varianza tra le risposte calcolata precedentemente (m2), calcoliamo

$$M2_c = \sigma_{v_c}^2 = \frac{\sum_i (v_{ic} - \bar{v}_c)^2}{ni}$$

metodo



• misura 3

- . gli studenti mancano item facili e rispondono correttamente agli item difficili
 - . misura in cui il pattern di risposta di uno studente è diverso da un altro studente con lo stesso score aggregato
- 1) calcolare la frazione \bar{q}_i^A di studenti in ogni livello aggregato di performance che ha risposto agli item correttamente

$$\bar{q}_i^A = \frac{\sum_{s \in \{z: A_z = A_s\}} q_{isc}}{nS_A}$$

q_{isc} è uguale a 1 se lo studente s della classe c ha risposto il item i correttamente

A_s è uguale alla performance aggregata dello studente s al test



- **misura 3**

- 1) misura in cui il pattern di risposta dello studente s differisce del pattern di risposta degli altri studenti con lo stesso score aggregato

$$Z_{sc} = \sum_i (q_{isc} - \bar{q}_i^A)^2$$

- 2) sottrazione della deviazione media degli studenti con lo stesso score aggregato \bar{Z}^A e somma degli studenti dentro della classe

$$M3_c = \sum_s (Z_{sc} - \bar{Z}^A)$$



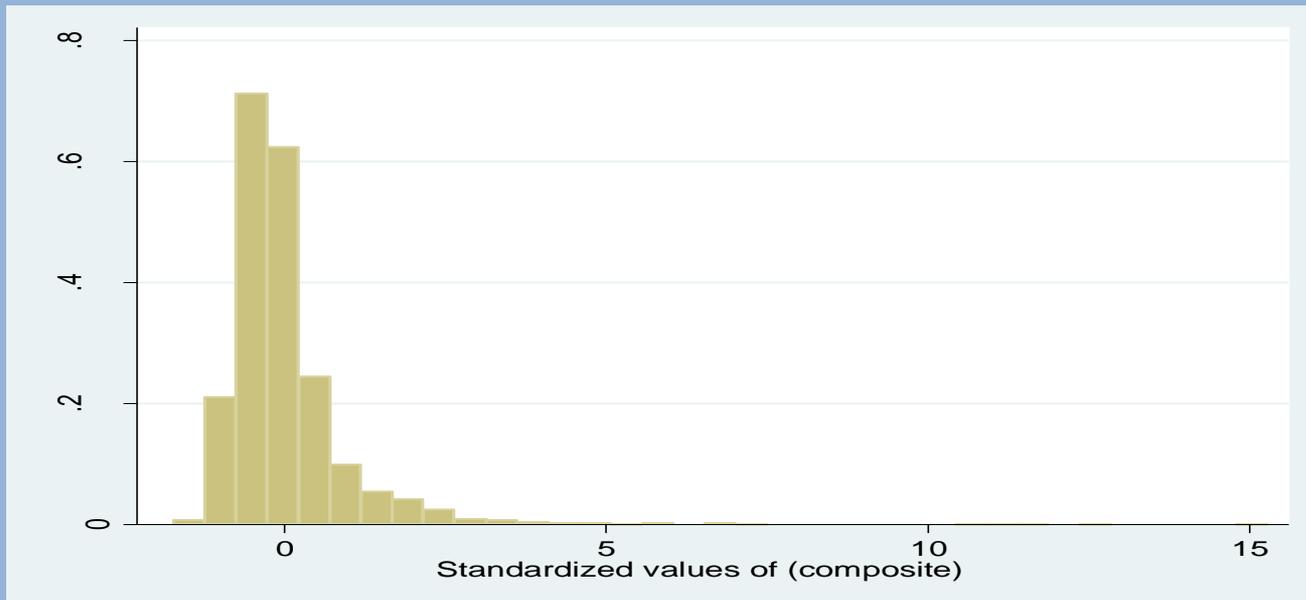
indicatore di sequenze di risposta sospette: sintesi

. due indicatori compositi:

** somma dei quadrati dei rank value delle 3 misure

$$CHEAT_{cbt} = (rank_M1_{c,b,t})^2 + (rank_M2_{c,b,t})^2 + (rank_M3_{c,b,t})^2$$

** somma dei valori normalizzati delle 3 misure



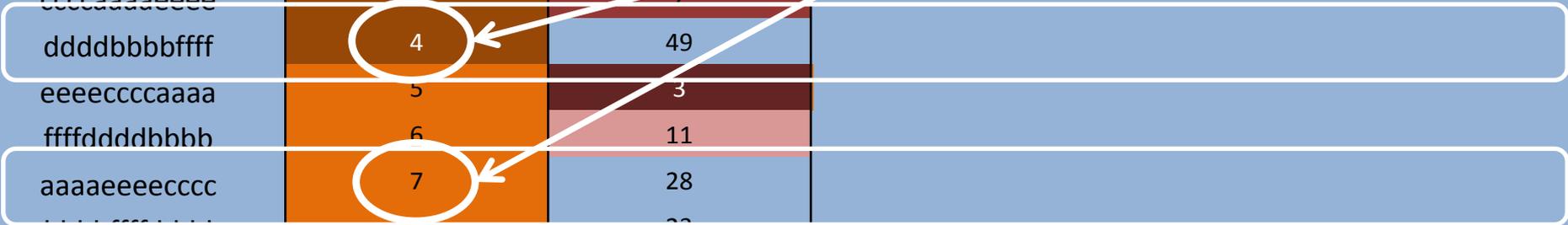
comparazione tra metodi

simulazione con un dataset di scuole N=100 estratte in modo casuale SNV 2010-11



scuole	ranking metodo invalsi	ranking metodo basato sui pattern di risposta non usuali
aaaaeeecccc	1	1
bbbfffdddd	2	2
ccccaaaaeee	3	7
ddddbbbbffff	4	49
eeeccccaaaa	5	3
ffffddd dbbb	6	11
aaaaeeecccc	7	28
bbbfffdddd	8	23
ccccaaaaeee	9	5
ddddbbbbffff	10	33
aaaaeeecccc	11	69
bbbfffdddd	12	64
ccccaaaaeee	13	15
ddddbbbbffff	14	41
eeeccccaaaa	15	29
ffffddd dbbb	16	6
aaaaeeecccc	17	58
bbbfffdddd	18	31
ccccaaaaeee	19	4

*potenziali errori di tipo I:
falsi positivi*



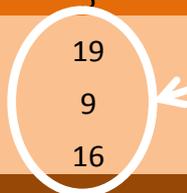
comparazione tra metodi

simulazione con un dataset di scuole N=100 estratte in modo casuale SNV 2010-11



scuole	ranking metodo invalsi	ranking metodo basato sui pattern di risposta non usuali
aaaaeeeecccc	1	1
bbbbffffdddd	2	2
ccccaaaaeeee	5	3
ddd dbbbffff	19	4
eeecccccaaaa	9	5
ffffddd dbbb	16	6
aaaaeeeecccc	3	7
bbbbffffdddd	70	8
ccccaaaaeeee	60	9
ddd dbbbffff	22	10
aaaaeeeecccc	6	11
bbbbffffdddd	39	12
ccccaaaaeeee	29	13
ddd dbbbffff	69	14
eeecccccaaaa	13	15
ffffddd dbbb	32	16
aaaaeeeecccc	20	17
bbbbffffdddd	43	18
ccccaaaaeeee	72	19
ddd dbbbffff	30	20

*potenziali errori di tipo II:
falsi negativi*



comparazione tra metodi



coincidenza tra metodi delle classi rilevate come 'cheater'

metodo basato sui pattern di risposta non usuali

p95

		Classi N = 3361		<i>potenziali</i>	
		non sospette	sospette	falsi positivi	
metodo invalsi	non sospette	3011	136	84,6%	
	sospette	181	33	80,5%	

potenziale errore di tipo I: falsi positivi

potenziale errore di tipo II: falsi negativi

p90

		non sospette	sospette	<i>potenziali</i>	
metodo invalsi	non sospette	2334	199	falsi positivi	83,3%
	sospette	690	138	falsi negativi	59,1%

proposte sul metodo



*proposte per la correzione del cheating **dopo** lo svolgimento delle prove
(prima della correzione alle scuole)*

sul metodo

*** metodi basati sulla logica del guadagno cognitivo (fluttuazioni dei punteggi non aspettati nel tempo) → necessità di controllare i fattori esogeni

*** metodi basati sulla probabilità condizionata per la rilevazione dei pattern di risposta non usuali

validazione del metodo

*** confronto di metodi

*** testare il metodo tramite simulazioni per minimizzare il rischio di errore di tipo I e II



oltre ai divari territoriali: i determinanti del cheating

dati



survey	servizio nazionale di valutazione (snv): 'prove invalsi'
anno scol	2010-2011
discipline	matematica (43 item) e italiano (81 item)
studenti	67174
classi	3361
scuole	739

modello empirico

oltre ai divari territoriali: i determinanti del cheating



$$\begin{aligned} cheat_{cs} = & \alpha_c + \gamma_1 T_r + \gamma_2 C_{cs} + \gamma_3 D_{cs} + \gamma_4 A_{cs} \\ & + \gamma_5 (D_{cs} A_{cs}) + \gamma_6 S_s + \varepsilon_{cs} \end{aligned}$$

dove

$cheat_{cs}$ è il livello di cheating della classe c nella scuola s

T_r è la variabile di ubicazione territoriale –regione- della scuola

C_{cs} è un vettore della composizione della classe, relativa alle caratteristiche sociodemografiche degli studenti – SES, genere, immigration background – e ripetenti

D_{cs} è un vettore relativo ai fattori che possono attuare come deterrenti del cheating, e contiene dummies sul controllo durante lo svolgimento delle prove Invalsi

A_{cs} è una proxy di strategia di cheating non basata sull'alterazione dei risultati delle prove, è rappresentata dallo share di assenze degli studenti nello svolgimento delle prove

P_s è un vettore che contiene dei fattori a livello di scuola che possono avere un ruolo nel spiegare il cheating: età media degli insegnanti e la dimensione media delle classi (controllo cheating studenti)

La regressione controlla i termini di interazione $D_{cs}A_{cs}$ tra le dummies dei deterrenti (controllori in classe) e il livello di assenza degli studenti nelle prove Invalsi

ε_{cs} è il error term

risultati

oltre ai divari territoriali: i determinanti del cheating



	(1)	(2)	(3)	(4)	(5)	(6)
	CHEAT	CHEAT	CHEAT	CHEAT	CHEAT	CHEAT
SES aggregato per classe						-0.296 ^{***}
% donne / classe	-0.00331	-0.00302	-0.00287	-0.00276	-0.00266	-0.00234
% stud immigr 1 gen/ classe	0.00061 ^{***}	0.00052 ^{***}	0.00061 [*]	0.00051 [*]	0.00078 ^{**}	-0.00561 [*]
% stud immigr 2 gen/ classe						-0.0111 ^{***}
% stud ripetenti / classe	-0.00791	-0.00782	-0.00904	-0.00925	-0.00823	-0.00819 ^{**}
Controllo in classe (ref: non controllo)						-0.297 ^{***}
Controllo test high-stake (ref: non controllo)						-0.276 ^{**}
Assenti alla prova di matematica						0.102 [*]
Età media degli insegnanti				0.0180	0.0222	0.0224
Class size				0.0285 ^{**}	0.0244 [*]	0.0246 [*]
Controllo in classe x strategia cheating						-0.103 [*]
Controllo in classe x strategia cheating						-0.173 [*]
Region fixed effect					Yes	Yes
_cons	0.375 ^{**}	0.429 ^{**}	0.241 ^{**}	-1.277 ^{**}	-1.324 ^{**}	-1.368 ^{**}
N	3361	3361	3361	3361	3361	3361

più bassa è la composizione sociale della classe, più alta è la tendenza a fare 'cheating'

l'aggregazione di studenti immigrati e studenti ripetenti in classe è associato a fare meno 'cheating'

il controllo in classe è un efficace deterrente del cheating, anche nelle scuole in cui i test sono high-stake per gli insegnanti

le classi che tendono a 'barare' svolgono anche strategie parallele di cheating non basate sull'alterazione dei test

la presenza di osservatori è un deterrente efficace anche per controbilanciare l'effetto di altre misure di cheating



prevenzione del cheating durante lo svolgimento delle prove

è evidente che il deterrente più efficace per contenere il cheating è la presenza di un controllore durante lo svolgimento delle prove – ma è anche il più insostenibile finanziariamente

approccio basato sul contrasto di interessi tra gli insegnanti di diversi gradi:

*** in un sistema di valutazione basato sul valore aggiunto, gli insegnanti sarebbero i **controllori ideali** per i test dei livelli inferiori

*** se i **livelli di apprendimento** misurati in uscita dalle scuole elementari fossero sovrastimati a causa del *cheating*, sarebbe difficile, per gli insegnanti della scuola media, creare valore aggiunto

vantaggio: le risorse disponibili localmente libererebbero inoltre i fondi per i controlli al livello superiore

grazie per l'attenzione



FONDAZIONE GIOVANNI AGNELLI WWW.FGA.IT

 FONDAZIONEAGNELLI

 @FONDAGNELLI

 FONDAGNELLI

