Perugia, 13 maggio 2011

DIFFERENTIAL VARIABILITY OF TEST SCORES AMONG SCHOOLS: A MULTILEVEL ANALYSIS OF THE 5TH GRADE INVALSI TEST USING HETEROSCHEDASTIC RANDOM EFFECTS

Leonardo Grilli **Claudia Sani**

Dipartimento di Statistica "Giuseppe Parenti"

Università degli Studi di Firenze

Outline

- Objectives
- Description of sample
- Analysis
  - Preliminary analysis
  - Multilevel analysis
    - Results
    - Analysis of predicted random effects
- Conclusions and future developments

Objectives

Evaluation of the performance of the Italian school system through the learning levels of the pupils

↓

Is the system fair?

↓

Investigation of individual and contextual factors that affect the variability of the scores across schools

Methodological novelty: we use a multilevel model with heteroschedastic random effects

Dataset

We analyse data from INVALSI. The dataset is composed by:

- Results of the mathematics test administered to 5th grade pupils (about 11 years old) at the end of the 2008/2009 year (41 *items*) → summarized by the Rasch score
- Results of pupil's questionnaire for measuring socio-economic factors
- Some demographic features of pupils (provided by school offices)

↓

The sample includes about 1000 schools and 40000 pupils

Steps of the analysis

1. Preliminary analysis for describing pupil and school features, and how they are related to the Rasch score of the math test:
   - Box-plot
   - Anova
   - Scheffé test for multiple comparisons
2. Multilevel regression models

   Dependent variable: Rasch score of the math test

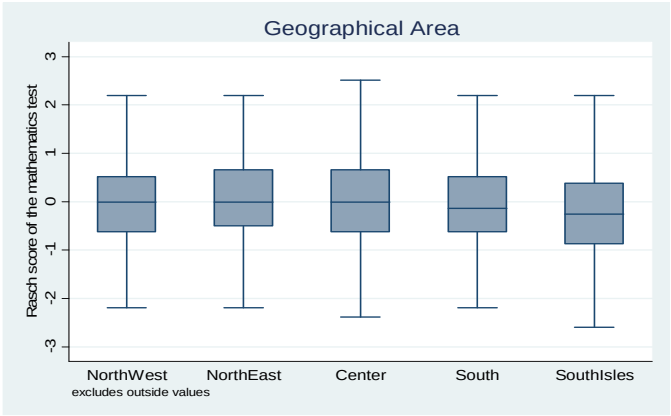
Note: analysis carried out using the software **Stata**

Analysed variables

- **Demographic variables** →
  - Gender
  - Foreigner *vs* Italian
  - Year of birth
- **Socio-cultural variables** →
  - Availability of computer, encyclopaedia, internet
  - Number of books at home
  - Help with homework
  - Hours playing video games
  - Hours reading
- **Home environment** →
  - With whom pupil lives
  - Language spoken at home
- **Wealth** →
  - Presence of alarm at home
  - Number of bathrooms at home
  - Number of cars at home
- **School climate** → Unease score (scale 0 - 4)
- **Geographical area of schools**

Distribution of Rasch math score

Min	Max	Mean	Std. Dev.	Percentil 25%	Percentil 50%	Percentil 75%
-5.664	4.683	0.009	0.944	-0.626	-0.135	0.515



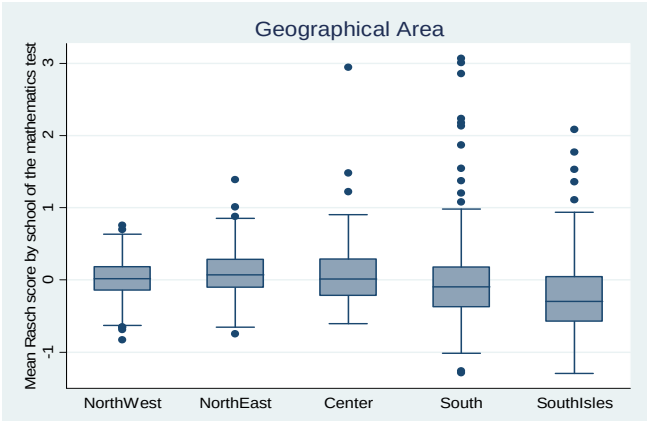
Std. Dev.	0.88	0.92	0.93	1.04	0.94

L. Grilli - C.Sani

Note: Basilicata and Calabria belong to South-Isles area

7

Distribution of ***school mean*** Rasch math score



Std. Dev.	0.30	0.34	0.41	0.67	0.54

L. Grilli - C.Sani

8

Multilevel model

Two-level linear model with heteroschedastic random effects

$$Y_{ij} = \gamma_{00} + \sum_{p=1}^{r} \beta_p X_{pij} + \sum_{t=1}^{s} \gamma_{0t} Z_{tj} + u_{0j}^{(k)} + \varepsilon_{ij}^{(m)}$$

level 1 errors

$$\varepsilon_{ij}^{(m)} \sim N\left(0; \left(\sigma_{\varepsilon}^{(m)}\right)^2\right)$$

level 1 variables

level 2 variables

level 2 errors

$$u_{0j}^{(k)} \sim N\left(0; \left(\sigma_{u_0}^{(k)}\right)^2\right)$$

i = 1, 2, …., 38708 → pupils (level 1)

j = 1, 2, …., 932 → schools (level 2)

Number of pupils by school: Min 1; Median 37; Max 123

m = 1, 2 → male, female

k = 1, 2,.., 5 → North-West, North-East, Center, South, South-Isles

Model selection

1. Empty models
   - with homoschedastic errors
   - with pupil-level errors depending on gender
   - with school-level errors depending on the geographical area
2. Models with pupil-level covariates (heteroschedastic errors)
3. Models with pupil-level *and* school-level covariates (heteroschedastic errors)

Significant pupil-level covariates

- **Demographic variables**
  Gender
  Foreigner *vs* Italian
- **Socio-cultural variables**
  Availability of encyclopaedia
  Hobby of reading
  Help with homework
  Number of books at home
- **Wealth**
  Number of bathrooms at home
- **School climate**
  Unease score (scale 0 - 4)

Note: statistical significance of coefficients has been assessed using the Wald test at 5% level

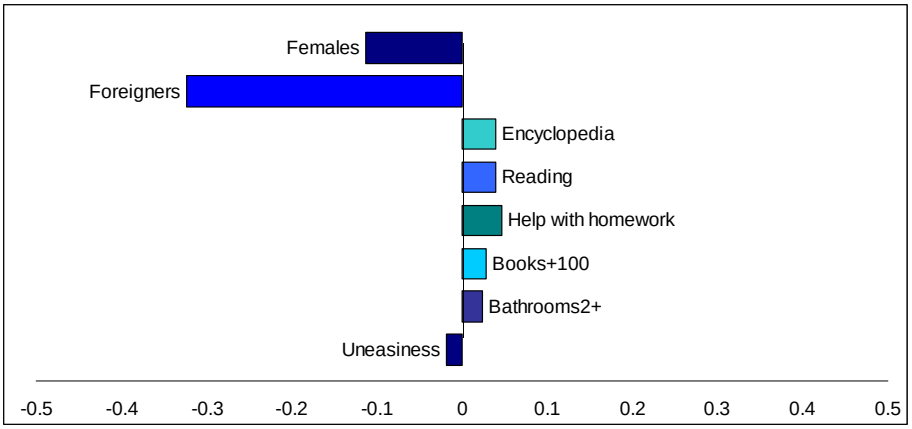
Significant school-level covariates

- **Geographical areas**
  North-West
  North-East
  Center (reference cat.)
  South
  South-Isles
- **Contextual variables**
  MS-Encyclopaedia
  MS-Reading
  MS-Bathrooms
  MS-Unease

Obtained as school means of pupil variables

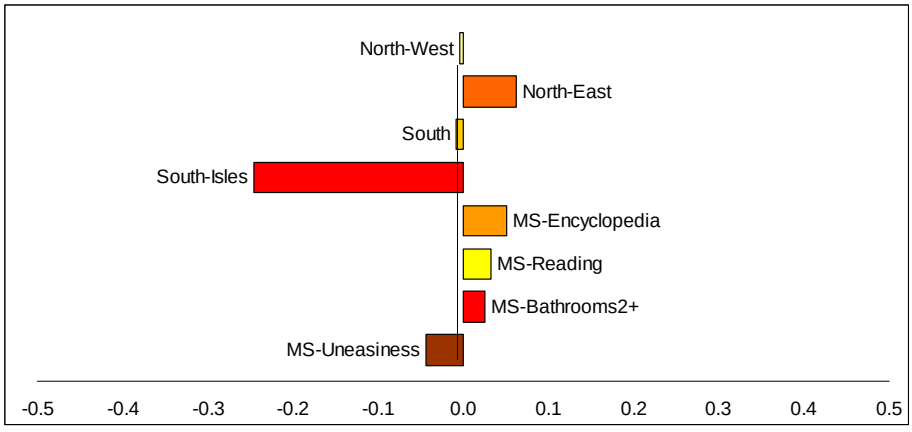
Note: statistical significance of coefficients has been assessed using the Wald test at 5% level

Coefficients of pupil-level covariates



Parameters estimated via Maximum Likelihood (command xtmixed)

Coefficients of school-level covariates



MS-Ency, MS-Read, MS-Bath: **+10% in the school percentage**
MS-Uneasiness: **+0.1 in the school mean of the score**

Standard deviations of model errors

Std. Dev. pupil-level errors

Male	0.886
Female	0.805

Std. Dev. school-level errors

North-West	0.197
North-East	0.237
Center	0.351
South	**0.652**
South-Isles	**0.486**

Parameters estimated via Maximum Likelihood (command xtmixed)

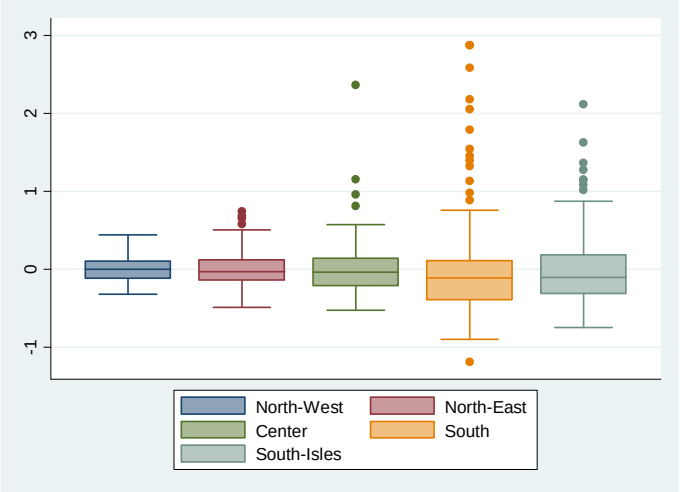
Intraclass Correlation Coefficient (ICC)

$$\rho_{m,k} = \frac{\left(\sigma_{u_0}^{(k)}\right)^2}{\left(\sigma_{u_0}^{(k)}\right)^2 + \left(\sigma_{\varepsilon}^{(m)}\right)^2} = \frac{\text{variance between schools}}{\text{total variance}} \qquad \rho_{m,k} \in [0,1]$$

ICC	North-West	North-East	Center	South	South-Isles
Males	4.7%	6.7%	13.5%	35.1%	23.1%
Females	5.7%	8.0%	16.0%	39.6%	26.7%

In the South area about 40% of the residual variance depends on the school attended, compared to about 5% in the North area

Predicted random effects



Predictions obtained with the Empirical Bayes method: the shrinkage property prevents schools with few pupils from having extreme values

Comparison between models with and without outlying schools

	N schools	N positive outliers	N negative outliers	ICC Model with outliers	ICC Model without outliers
North-West	216	0	0	4.7%	4.7%
North-East	179	5	2	6.7%	4.0%
Center	186	4	0	13.5%	7.2%
South	175	13	1	35.1%	11.9%
South-Isles	176	8	0	23.1%	14.5%

Outlying schools in the South area explain a great part of residual variance

Best and worst schools

- The *ranking of the schools based on the mean math score* is similar to the *ranking based on predicted random effects* (i.e. adjusted for the covariates)
- In particular, the positions of the top 20 schools are nearly unchanged
- Both best and worst schools belong to South and South-Isles areas (a consequence of the high variability)

self-selection of students and teachers in Southern Italy (driven by unobserved variables)

Differences in the expected score: underprivileged vs. privileged pupil; ineffective vs. effective school

Mean Score	Observed variables		Not observed variables	
	Pupil underprivileged → privileged	School ineffective → effective	Pupil underprivileged → privileged	School ineffective → effective
North-West	+0.707	+0.569	+3.544	+0.788
North-East	+0.707	+0.569	+3.544	+0.948
Center	+0.707	+0.569	+3.544	+1.404
South	+0.707	+0.569	+3.544	+2.608
South-Isles	+0.707	+0.569	+3.544	+1.944

Observed variables (model covariates)

Underprivileged/ineffective → Negative pattern of covariates

Privileged/effective → Positive pattern of covariates

Not observed variables (model errors)

Underprivileged/ineffective → -2 Std. Dev.

Privileged/effective → +2 Std. Dev.

Conclusions /1

- The math test score depends on several pupil-level factors (gender, to be foreigner, availability of encyclopaedias, hobby of reading, help with homework, number of books at home, number of bathrooms at home, uneasiness at school), as well as the geographical area of school and the contextual variables
- The South-Isles area has the lowest mean score (-0.195 versus 0.097 of the North-East area), whereas the South area has a mean score similar to Northern Italy but a huge between-school standard deviation (0.652 versus 0.197 of the North-West)
- In the South area about 40% of residual variance depends on the school, whereas in Northern Italy about 5% → The variance among schools increases when going from North to South → The goal of fairness is not attained in Southern Italy

Conclusions /2

- In the South area, a great part of variance depends on 14 outlying schools (8%): all of them – except one – have exceptionally positive results
  - the self-selection process seems to be asymmetric ('positive' selection into excellent schools more than 'negative' selection into worse schools)
  - the existence of positive outlying schools responsible for a substantial part of the variability makes the overall picture less problematic than what might appear at first sight

Future developments

- Analyse the Invalsi test score on Italian language and compare it with the math test (similar conclusions?)
- Analyse the test scores for several years (are the results stable over time?)
- Extend the model to more than two levels (for example, province as third level)
- Use models with random effects having an asymmetric distribution (to account for schools with exceptionally positive results)

References

- Cipollone P., Sestito P. (2010). *Il capitale umano*. Il Mulino, Bologna.
- Goldstein H. (2000). *Multilevel Statistical Models*. 3rd ed., Hodder Arnold, Londra.
- Grilli L., Bacci S., Bertaccini B. (2007). *Offerta formativa di secondo livello e domanda di lavoro qualificato*. Università di Firenze, Dipartimento di Statistica "G. Parenti". Rapporto di ricerca, "Finvali 2005"- progetto 047. Invalsi, Roma.
- Hox J. (2002). *Multilevel Analysis. Techniques and Applications*. Lawrence Erlbaum Associates, Inc., New Jersey.
- Invalsi (2009). *Servizio Nazionale di Valutazione a.s. 2008-2009. Rilevazione degli apprendimenti. Scuola primaria. Prime analisi*. Scaricabile dal sito web http://www.invalsi.it.
- Rabe-Hesketh S., Skrondal A. (2008). *Multilevel and Longitudinal Modelling Using Stata*. 2nd ed., Stata Press, Texas.
- Snijders T.A.B., Bosker R.J. (1999). *Multilevel analysis: an introduction to basic and advanced multilevel modelling*. Sage Publications, Londra.
- Stata Corp. (2009). *Stata 11 Reference Manual*. Stata Press, College Station, Texas.
- Steele F., Goldstein H. (2007). *Multilevel Models in Psychometrics*. In: Handbook of Statistics, Vol.26. Elsevier B.V, pp.401-420.