Estimating school and class effects on achievement - methodological insights and empirical evidence about Italian primary and secondary schools

Tommaso Agasisti, Politecnico di Milano School of Management

The policy issue (i)

- Analysing school performance… and using it for choice and accountability purposes
  - Choice (Family perspective): Key elements for supporting family choices ("objective" information)
  - Accountability (Ministry perspective): deciding about the allocation of resources across schools and for rewards/sanctions
- Market mechanisms and regulation can work better if better information is available
  - Public and private projects for creating rankings of schools…

The policy issue (ii)

- Key problems of using measures of school performance
  - Standardized tests, cognitive skills, etc. (limitations about the competences assessed)
  - Performance levels vs improvement ➜ School Value Added (VA), net of the role of individual-level factors

  - Stability of VA estimates over time
    - Can these numbers inform the choices and policies in the future (predictive power)?
  - Identification of determinants of school VA
    - Accuracy in prediction and factors associated with it

The policy issue (iii)

- If performance information is distorted, market incentives do not work
  - Perverse effects in allocation of public resources (sanctions/rewards)
  - Reputation of schools (reinforcing inefficient self-selection)
  - Ineffective policies (adverse imitation)

“Heads, you failed to learn. Tails, I failed to teach.”

Policy background (i)

- The Italian educational system
  - K-12 education is articulated in three cycles: primary (1-5), junior secondary (6-8) and secondary (9-13) schools
    - Students stay in the same class with same teachers within the cycle
  - Standardized tests (low-stakes) at grades 2, 5, 8 and 10 (reading and mathematics; from 2017/18, English)
  - Teachers' grades every year; high-stakes measures for passing each year and moving across cycles
    - Focus of the paper: students' results at grade 5
  - High regulation by the central government (teachers' allocation and pay, curricula structure and content)
  - Recent reform for increasing the autonomy of school principals

Policy background (ii)

- The educational system is characterized by important achievement gaps:
  - Italian vs immigrants (1st and 2nd generation)
    - 2nd generation higher scores than 1st, but much lower than Italian classmates
  - Disadvantaged vs advantaged students/schools
    - The indicator ESCS (Economic, Social and Cultural Status, see OECD)
  - Northern, Central and Southern Italy
    - Students/schools in the North outperform their counterparts in the South
  - Public vs private schools (not addressed here; few students in private)

In this presentation

Paper #1

Agasisti, T., & Minaya, V. (2018). Evaluating the Stability of School Performance Estimates for School Choice: Evidence for Italian Primary Schools, Italian Society Public Economics, Working Paper No. 67.

- Testing the stability of school VA estimates over time (across cohorts)

Paper #2

Schiltz, F., Sestito, P., Agasisti, T., & De Witte, K. (2018). The added value of more accurate predictions for school rankings. Economics of Education Review, 67, 207-215.

- Using machine learning techniques for improving accuracy of Value Added estimates

Paper #1

- School performance estimates have been used worldwide for both high-and low-stakes accountability purposes. It is expected that by evaluating school performance and making these results public, parents will use them to choose schools and schools will be motivated to increase performance. Using administrative data provided by INVALSI (National Evaluation Committee for Education), this paper explores the stability of performance estimates for Italian primary schools. We first construct school performance metrics using INVALSI standardized tests and quarterly teacher assessments, by taking advantage of a rich array of individual level variables (including prior achievement) that allow us to estimate a school-effect in a 'value added' perspective. We then explore how sensitive school ratings are to the choice of performance metric and the use of different models to account for compositional differences due to students' socioeconomic background. We find that school performance estimates are very robust whatever the models employed to control for compositional differences, but they are inconsistent across metrics and cohorts.

Siep
Società italiana di economia pubblica

WORKING PAPER
No 733
febbraio 2018

EVALUATING THE STABILITY OF SCHOOL PERFORMANCE ESTIMATES FOR SCHOOL CHOICE: EVIDENCE FOR ITALIAN PRIMARY SCHOOLS

Tommaso Agasisti, Politecnico di Milano School of Management

Veronica Minaya, Politecnico di Milano School of Management

JEL Classification: I21

Keywords: accountability, school choice, school value-added

società italiana di economia pubblica
c/o dipartimento di scienze politiche e sociali – Università di Pavia

Research questions

- This paper focuses on testing the robustness of value-added estimations from a methodological perspective, by looking at how some features of the modelling affect the estimated school effects: functional specification, persistence across cohorts, metrics used as output indicators

  - [RQ.1]
    Which specifications can be used for estimating "school effect" of primary schools?
    - Which assumptions are behind the different specification?
  - [RQ.2]
    How stable and robust are school effects' estimates?
    - Stability across specifications (models), over time (across cohorts) and across metrics

Data (i)

- Dataset provided by INVALSI Statistical Office
- Three cohorts of students at grade 5
  - 2013/14, 2014/15 and 2015/16
  - Sample restrictions: only students with cheating-corrected scores, for whom we have both grade 2 and 5 test scores
    - 790,000 students in 5,200 schools
    - 55% of all students, 75% of all schools

Data (ii)

- Output variables
  - INVALSI scores in Italian and Mathematics (grade 5)
  - WLE INVALSI scores (Rasch scores) in Italian and Mathematics (grade 5)
  - Teachers' grades (voto) in Italian and Mathematics (grade 5)

- Input variables – student and school variables
  - Student-level variables (gender, immigrant status, ESCS)
  - Class and school-level variables (averages at class and school level for student-level variables)

- Descriptive statistics

Variables	2013	2014	2015	Total
Demographics and SES				
Female	0.50	0.50	0.50	0.50
Italian	0.97	0.97	0.97	0.97
Regular Student	0.98	0.98	0.98	0.98
Average Student ESCS	0.14	0.10	0.11	0.11
Father in High Income Occupation	0.22	0.23	0.22	0.22
Mother in High Income Occupation	0.11	0.12	0.12	0.11
Parents Attended College	0.67	0.70	0.71	0.69
Outcome Metrics				
INVALSI Score in Ma, 5th Grade	65.26	58.62	56.50	59.71
INVALSI Score in Ma, 2nd Grade	60.47	60.91	58.76	59.86
INVALSI Score in It, 5th Grade	63.40	59.80	65.65	63.38
INVALSI Score in It, 2nd Grade	68.32	69.85	62.28	66.16
WLE in Math, 5th Grade	204.68	206.39	206.18	205.79
WLE in Italian, 5th Grade	203.12	205.91	203.12	203.88
Teacher's Score in Math, 5th Grade	7.92	7.94	7.96	7.94
Teacher's Score in Italian, 5th Grade	7.84	7.87	7.90	7.87
Number of Students	237,526	215,533	337,460	790,519
Number of Schools	3,706	3,256	4,929	5,427

VA models estimated in this paper (i)

- A glance to the variables used in the various models

Method and Covariate Choice							
	Student's Second Grade Score	Student Demographic Control	Class Means	Class Fixed Effects	School Means	School Fixed Effects	School Random Effects
Model 1							
Model 2	X					X	
Model 3	X	X				X	
Model 4	X	X	X			X	
Model 5	X	X		X			
Model 6	X	X	X		X		
Model 7	X	X	X				X

VA models estimated in this paper (ii)

- Model 1: Raw means
  - No controls; percentage of correct answers

- Model 2: VA model with grade 2 (<u>prior achievement</u>) as control

$$y_{isc} = \beta t_i + \gamma_c + schFE + \varepsilon_{isc}$$

  - y_{isc} is test score at grade V
  - t_i is the student prior achievement
  - γ_c is the cohort fixed effect
  - $schFE$ is the school (fixed) effect

VA models estimated in this paper (iii)

- Model 3: VA model with grade 2 (prior achievement) and student's characteristics as control

$$y_{isc} = \beta t_i + \delta X_i + \gamma_c + schFE + \varepsilon_{isc}$$

  - y_{isc} is test score at grade V
  - t_i is the student prior achievement
  - X_i is a vector of student's characteristics (gender, immigrant status, socioeconomic background)
  - γ_c is the cohort fixed effect
  - $schFE$ is the school (fixed) effect

VA models estimated in this paper (iv)

- Model 4: VA models with grade 2 (prior achievement) and student's characteristics as control, as well as class-average characteristics

$$y_{isc} = \beta t_i + \delta X_i + \partial \bar{X}_j + \gamma_c + schFE + \varepsilon_{isc}$$

  - y_{isc} is test score at grade V
  - t_i is the student prior achievement
  - X_i is a vector of student's characteristics (gender, immigrant status, socioeconomic background)
  - $\bar{X}_j$ is a vector with class-level characteristics (including class size and average test score at grade 2)
  - γ_c is the cohort fixed effect
  - $schFE$ is the school (fixed) effect

VA models estimated in this paper (v)

- Model 5: VA models with grade 2 (prior achievement) and student's characteristics as control, and class fixed-effects

$$y_{isc} = \beta t_i + \delta X_i + \text{classFE} + \gamma_c + schFE + \varepsilon_{isc}$$

  - y_{isc} is test score at grade V
  - t_i is the student prior achievement
  - X_i is a vector of student's characteristics (gender, immigrant status, socioeconomic background)
  - classFE is a the class (fixed) effect
  - γ_c is the cohort fixed effect
  - $schFE$ is the school (fixed) effect

VA models estimated in this paper (vi)

- Model 6: VA models with grade 2 (prior achievement) and student's characteristics as control, as well as class-average characteristics and without school fixed effects
  - *School VA is the average of residuals (at class level) for each school*

$$y_{isc} = \beta t_i + \delta X_i + \partial \bar{X}_j + \gamma_c + \varepsilon_{isc}$$

  - y_{isc} is test score at grade V
  - t_i is the student prior achievement
  - X_i is a vector of student's characteristics (gender, immigrant status, socioeconomic background)
  - $\bar{X}_j$ is a vector with class-level characteristics (including class size and average test score at grade 2)
  - γ_c is the cohort fixed effect

VA models estimated in this paper (vii)

- Model 7: VA multilevel models with grade 2 (prior achievement) and student's characteristics as control, as well as class-average characteristics

$$y_{isc} = \beta t_i + \delta X_i + \partial \bar{X}_j + \gamma_c + schRE + \varepsilon_{isc}$$

  - y_{isc} is test score at grade V
  - t_i is the student prior achievement
  - X_i is a vector of student's characteristics (gender, immigrant status, socioeconomic background)
  - $\bar{X}_j$ is a vector with class-level characteristics (including class size and average test score at grade 2)
  - γ_c is the cohort fixed effect
  - $schRE$ is the school (random) effect

Results (i)

- High correlations of school-effects' estimates across models

Italian

	M1	M2	M3	M4	M5	M6	M7
M1	1.000	0.000					
M2	0.807 *	1.000					
M3	0.758 *	0.961 *	1.000				
M4	0.735 *	0.937 *	0.981 *	1.000			
M5	0.801 *	0.899 *	0.974 *	0.957 *	1.000		
M6	0.744 *	0.965 *	0.999 *	0.979 *	0.964 *	1.000	
M7	0.795 *	0.913 *	0.980 *	0.962 *	0.996 *	0.974 *	1.000

Mathematics

	M1	M2	M3	M4	M5	M6	M7
M1	1.000						
M2	0.762 *	1.000					
M3	0.754 *	0.979 *	1.000				
M4	0.741 *	0.961 *	0.986 *	1.000			
M5	0.840 *	0.905 *	0.960 *	0.951 *	1.000		
M6	0.734 *	0.982 *	0.999 *	0.985 *	0.950 *	1.000	
M7	0.844 *	0.935 *	0.973 *	0.962 *	0.993 *	0.965 *	1.000

Results (ii)

- Despite high correlations, there are still schools ranked in different quartiles depending upon the model used
  - But the inclusion of prior achievement controls for extreme quartiles

Italian: Model 7 vs. Model 1

	M7 Quartiles			
M1 Quartiles	Top	Q2	Q3	Bottom
Top	71%	23%	5%	1%
Q2	20%	46%	28%	7%
Q3	6%	22%	45%	26%
Bottom	3%	9%	22%	66%

Italian: Model 7 vs. Model 3

	M7 Quartiles			
M1 Quartiles	Top	Q2	Q3	Bottom
Top	89%	10%	0%	0%
Q2	11%	76%	13%	0%
Q3	0%	14%	76%	11%
Bottom	0%	0%	11%	89%

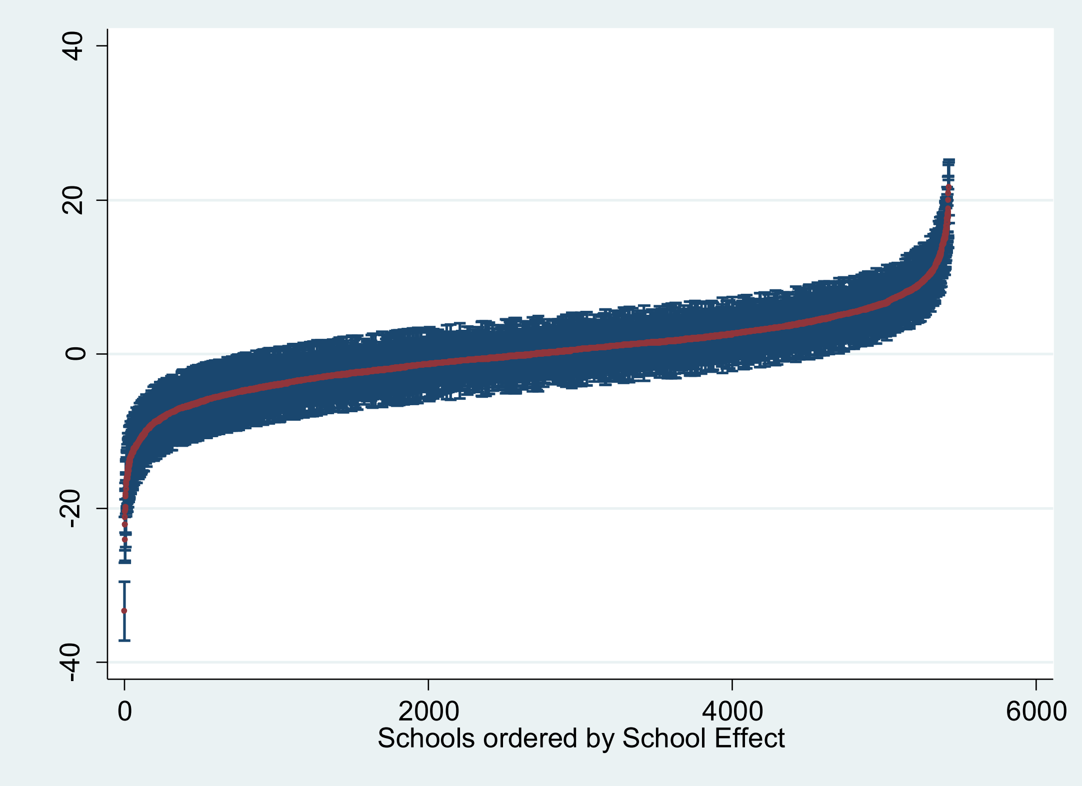
Mathematics: Model 7 vs. Model 1

	M7 Quartiles			
M1 Quartiles	Top	Q2	Q3	Bottom
Top	73%	23%	3%	0%
Q2	20%	48%	29%	3%
Q3	6%	23%	48%	23%
Bottom	1%	6%	20%	73%

Mathematics: Model 7 vs. Model 3

	M7 Quartiles			
M3 Quartiles	Top	Q2	Q3	Bottom
Top	87%	12%	1%	0%
Q2	13%	72%	14%	1%
Q3	0%	16%	73%	11%
Bottom	0%	0%	12%	88%

- For many schools, the estimated school effect is not statistically different from zero, and many schools have a statistically identical school effect



95% confidence intervals

- Represent the precision in estimating the school effect
- Based on model 7

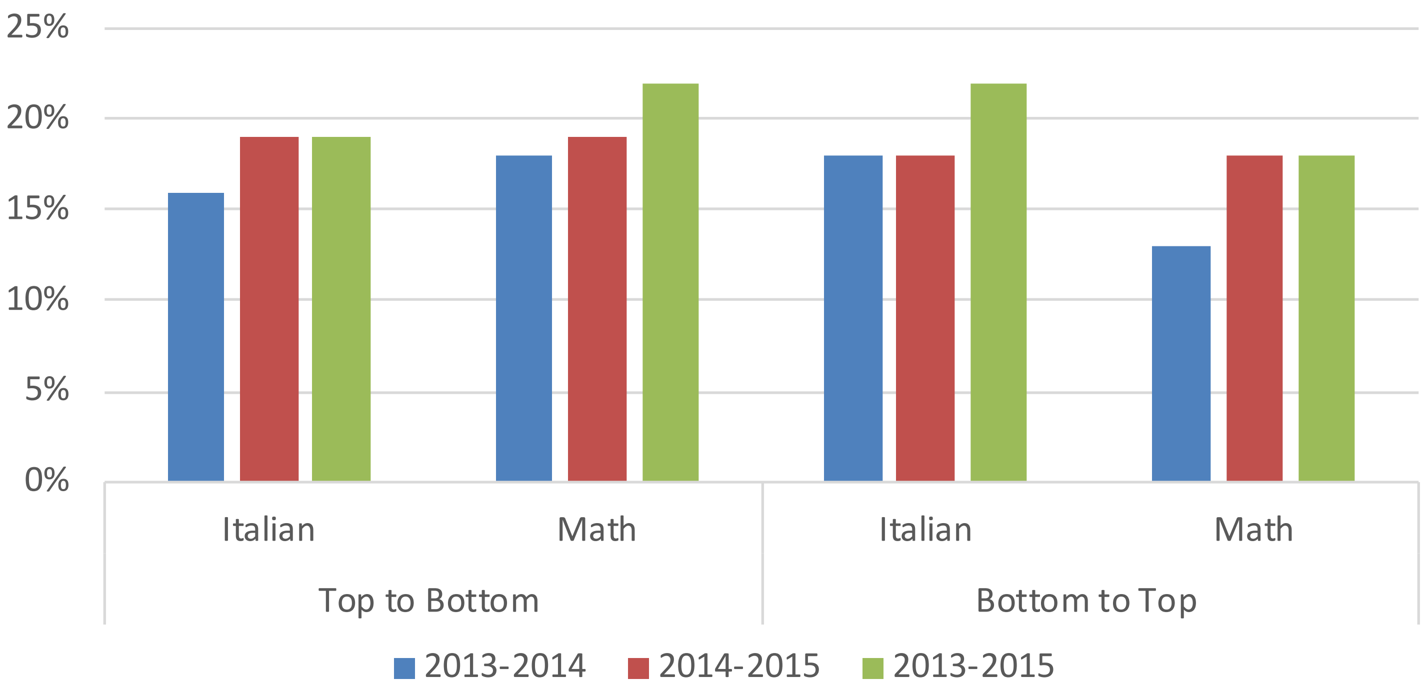
Results (iv)

- Correlations are low for school-effects estimated across cohorts
  - Low utility for predicting future school VA using present school VA

	M1		M2		M3		M4		M5		M6		M7	
Italian														
Corr(2013, 2014)	0.424	*	0.242	*	0.202	*	0.226	*	0.273	*	0.223	*	0.231	*
Corr(2014, 2015)	0.432	*	0.293	*	0.247	*	0.257	*	0.000		0.260	*	0.264	*
Corr(2013, 2015)	0.404	*	0.149	*	0.110	*	0.146	*	-0.079	*	0.139	*	0.147	*
Mathematics														
Corr(2013, 2014)	0.424	*	0.346	*	0.317	*	0.307	*	0.308	*	0.289	*	0.306	*
Corr(2014, 2015)	0.368	*	0.202	*	0.194	*	0.226	*	0.170	*	0.227	*	0.226	*
Corr(2013, 2015)	0.322	*	0.119	*	0.110	*	0.152	*	0.109	*	0.150	*	0.152	*

Results (v)

- Correlations are low for school-effects estimated across cohorts
  - Around 35% for math and Italian move from bottom to top and from top to bottom across cohorts



Results (vi)

- Correlations are low for school-effects estimated with models that use different output variables
  - Teachers' grades are very unrelated with INVALSI test scores

	Mathematics			Italian		
Method	**Corr(INV, WLE)**	**Corr(INV, Voto)**	**Corr(WLE, Voto)**	**Corr(INV, WLE)**	**Corr(INV, Voto)**	**Corr(WLE, Voto)**
M1	0.998	0.196	0.187	0.999	0.244	0.245
M2	0.997	0.182	0.174	0.998	0.169	0.170
M3	0.997	0.116	0.100	0.998	0.037	0.038
M4	0.997	0.030	0.022	0.998	-0.007	-0.007
M5	0.997	0.146	0.128	0.998	0.057	0.058
M6	0.997	0.044	0.039	0.998	-0.004	-0.004
M7	0.997	0.033	0.026	0.998	-0.011	-0.011

Results (vii)

- Correlations are low for school-effects estimated with models that use different output variables, even over time across cohorts
  - When considering two years apart, the correlations are even lower

	MODEL 1						MODEL 7					
	INV		**WLE**		**Voto**		**INV**		**WLE**		**Voto**	
Math												
Corr(2013, 2014)	0.416	*	0.418	*	0.569	*	0.303	*	0.286	*	0.520	*
Corr(2013, 2015)	0.317	*	0.334	*	0.571	*	0.146	*	0.157	*	0.510	*
Corr(2014, 2015)	0.361	*	0.361	*	0.553	*	0.223	*	0.223	*	0.486	*
Italian												
Corr(2013, 2014)	0.411	*	0.412	*	0.574	*	0.232	*	0.228	*	0.517	*
Corr(2013, 2015)	0.397	*	0.400	*	0.552	*	0.153	*	0.154	*	0.488	*
Corr(2014, 2015)	0.420	*	0.420	*	0.575	*	0.270	*	0.270	*	0.511	*

Implications from the longitudinal analysis (i)

- School effects' estimates are quite consistent across models
  - The policy debate should not be too much focused on the specification of the VA model, but more in understanding the different assumptions behind each model

- The estimated school VA is much different across cohorts
  - The idea of using them for promoting school choice must be regarded with caution → unintended consequence is misleading message
  - Understanding the determinants of cohort-specific determinants → the role of teachers' quality and school management

Implications from the longitudinal analysis (ii)

- School effects are <u>extremely unstable when using different outcome metrics</u>
  - Which ones would be more interesting (useful!) for school choice?
    - High-stake (teachers' grades) vs low-stake (INVALSI) measures
    - School choice (grades) vs school evaluation (INVALSI)

  - School effects estimated through using teacher assessments are more stable across cohorts than when using standardized tests scores

  - How would estimates based on long-term outcomes (persistence, success in HE, earnings, etc.) look like?