

The determinants of students' achievement: a multilevel analysis of new Italian data

Tommaso Agasisti

Abstract. The international comparisons about students' achievements and competencies (OECD-PISA, TIMSS, PIRLS, etc.) showed a clear evidence about the poor performance of Italian students. In 2009, for the first time, the National Committee for the Evaluation of the Italian Educational System (Invalsi) released a new dataset, containing student scores obtained in a standardized test promoted at the end of the lower-secondary education (scholastic year: 2007/08). In this paper, I use a multilevel approach to study the relationship between math Scores and individual-level and school-level factors. As the Invalsi dataset does not include reliable information about schools, it was merged with the TIMSS2007 dataset, which instead include many relevant indicators. The results tell that individual characteristics related to performance are, mainly, the disabled status, the foreign nationality and the repetition of one or more years (late): all these factors are associated to lower scores in the test. When looking at the school-level variables, it turns out that composition of student body (in terms of socio-economic background) matters more than schools' resources. In the last part of the paper, policy implications and methodological suggestions are formulated.

1 Motivation and objectives

In recent years, there is a growing interest in analysing the determinants of Italian students' performances. This interest lead to an improvement of the evaluation procedures of schools' characteristics and results in Italy. At present, a national Evaluation Committee (Invalsi – *Istituto Nazionale per la Valutazione del Sistema di Istruzione e Formazione*) has been established with the specific aim to evaluate the Italian schools. Among other tasks, a recent law prescribed that the Invalsi should carry out an analysis of students' achievement at different levels of education: second and fifth year of the primary school (age 7 and 10, respectively), first and third of the lower-secondary (age 11 and 13), second and fifth of the upper-secondary (age 15 and 18). This is the first time that a law imposes a national evaluation by using standardized tests. In the year 2008/09, the evaluations covered the primary-education level, and the third year of the lower-secondary; in the next two years, the evaluation system will be fully implemented.

The present paper employs the data concerning the national final examination at the end of the lower secondary education (third year of the lower secondary schools, when the regular students are 13 years old) in the year 2007/08. This final examination, for the first time in Italy, has been conducted through a standardized test identical for

¹ Tommaso Agasisti, Politecnico di Milano, Department of Management, Economics and Industrial Engineering, +39.02.2399.3963. tommaso.agasisti@polimi.it

all the students involved in the exam. About 560,000 students in 5,923 schools have compiled the test (Invalsi, 2008).

The present paper is noteworthy because three main reasons.

First, it uses for the first time the new dataset developed by Invalsi. This is a very important innovation, as previous studies only dealt with OECD-PISA data; while important, they suffer some major limitations, which impose to interpret the results cautiously (e.g. Vittadini, 2007). National tests (e.g. those provided by Invalsi) are usually better for within country analyses, because they have two advantages: (i) they are focused on competencies specifically taught at school, and (ii) all the students involved in testing procedures attend the same schooling level. The innovation of this paper is that it uses, for the first time, the first wave of available data derived from a national standardized test in Italy (reference year: 2007/08).

A major limitation of available Invalsi data is that they do not include information about (i) students' socio-economic conditions and (ii) schools' characteristics and resources. This probably explains why this dataset has not been used before for studies about the determinants of students' achievement: indeed, since early studies in the field (Hanushek, 1986) it was demonstrated that these are crucial factors for determining students' achievement. This is the second innovation of this paper, as I propose a methodology to overcome this problem (solving the lack of such important information) by merging the Invalsi data with those obtained through the TIMSS international questionnaires. This procedure was possible because Invalsi knows the schools' identification codes (details about merging procedure and variables' selection are provided in the section §3).

The third innovation is that this paper uses a multilevel approach: to the best of the author's knowledge, only Longobardi *et al.* (2009) used this method before for analysing Italian students' data. On the contrary, the use of multilevel modelling to explore the determinants of students' performances is quite diffuse in the international literature (among others, see Lamb & Fullarton, 2002; Fekjaer & Birkelund, 2007; Dronkers & Robert, 2008). It is important to remark this methodological choice, that is not diffuse in the Italian literature. Indeed, the structure of the data requires a multilevel modelling: students (level-1 data) are sorted into schools (level-2), and schools are sorted into geographical macro-areas characterized by different socio-economic characteristics likely to affect students' results, because previous studies demonstrated major differences in students' performances across the different areas in Italy (Bratti *et al.*, 2007; Longobardi *et al.*, 2009).

The main objective of this paper is to describe the main determinants of the students' achievement (as measured by the Invalsi test) in Math and Reading, with reference to the year 2007/08.

2 Background

There are few studies analysing the determinants of Italian students' achievement and performance specifically.

The most extensive effort, in the economics field, has been put by researchers at the Milan University, which were collected in a book (Bratti *et al.* 2007) with the objective to provide a wide-range set of explanations for achievement differentials among Italian students. This research used OECD-PISA2003 data, and the analysis was

carried out at student-level. The methodological approach was to estimate an Educational Production Function (EPF), by using student achievement (OECD-PISA scores) as outputs, and several students and schools' characteristics as inputs. Moreover, the analysis controlled also for some major factors affecting student performance, like school location, the type of schools, etc. Given the richness of their results, it is impossible to summarize them in few rows. However, their main result is that factors affecting students' achievements are related to (i) the socio-economic status (SES), (ii) the macro-area of the country (Northern Italy schools perform much better than their Central, Southern and Isles counterparts), and (iii) the type of school (academic schools – *Licei* – and technical schools perform better than vocational schools).

Another study in this field is that of Longobardi *et al.* (2009). They employ a multilevel modelling to analyse the effects of different covariates on students' performances, by using OECD-PISA2006 data (their study, too, was conducted with students as the unit of observation). Their statistical analysis (that is more robust than previous ones as it explicitly accounts for the hierarchical nature of data) pointed out results, which are consistent with Bratti *et al.* (2007): (i) a gap between Northern and Southern part of Italy, (ii) a strong differences among schools' types (namely, with *Licei* schools outperforming technical and vocational schools), (iii) a major impact of socio-economic background in predicting students' results. Through their results, the authors concluded that a high-centralized system as the Italian was not finally able to guarantee homogenous students' performances.

I also contribute myself to the debate with two papers (Agasisti, 2008; 2009). Both contributions rely upon OECD-PISA2006 data, but they differ with regard to the methodological approach: the first specifies an Educational Production Function, while the second realises a Data Envelopment Analysis. In both cases, data are aggregated at school-level. The results are pretty similar to those provided by Bratti *et al.* (2007) and Longobardi *et al.* (2009), and more specifically: the role of geographical gaps, the type of school, the socio-economic conditions of students. Moreover, in both cases a measure of competition has been added among the explanatory set of covariates, to investigate whether competition actually fosters schools' performance (the empirical results partially confirm this hypothesis).

To sum up the previous literature about Italy: while some evidence seems to emerge, it is based exclusively on international tests, which suffer some important limitations – described above. Moreover, the methodological approaches adopted in previous work did not explicitly account for the multilevel nature of data (Longobardi *et al.*, 2009, is a notable exception). The present paper starts filling these gaps, (i) by using, for the first time, the unique national standardized tests, administered by Invalsi, and (ii) by modelling the relationship between achievement and schools and students' characteristics in a multilevel setting.

3 Methodology and data

In the analysis, I employed two levels of variables (to account for the hierarchical structure of data): student-level (level 1) and school-level (level 2). The methodological approach is a traditional multilevel model with a random intercept.

Table 1 contains the descriptive statistics of the dataset. All the data at the individual level come from the Invalsi dataset, which refer to the final examination at the end of the lower-secondary education (reference year: 2007/08). I used two alternative output indicators: the test scores in reading (Reading_Score) and Math (Math_Score)¹. The scores have been standardized into a range [0;100], that represents the percentage of right answers to the questions of the test.

Table 1: Descriptive statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Math Score	5,355	54.05	19.04	0.00	100
Reading Score	5,355	61.41	19.11	0.00	100
Late	5,320	0.09	0.28	0.00	1
Early	5,320	0.04	0.20	0.00	1
Disabled	5,364	0.01	0.07	0.00	1
Foreign	5,355	0.08	0.27	0.00	1
Female	5,355	0.49	0.50	0.00	1
Students	5,355	469.79	221.84	1.01	940
Private	5,355	0.01	0.12	0.00	1
Community	5,364	3.62	1.38	1.00	6
Disadvantaged	4,984	1.95	0.93	1.00	4
Istr_Days	5,364	203.67	3.05	200.00	216
Short_instr	5,364	1.68	0.78	1.00	4
Internet	5,364	1.59	0.83	1.00	4
South	5,355	0.42	0.49	0.00	1
North	5,355	0.39	0.49	0.00	1
Central	5,355	0.18	0.38	0.00	1

As covariates at individual level, I employed several students' characteristics: gender (dummy: Female), citizenship (a dummy – Foreign – for students who are not Italian), disabled status (dummy: Disabled). Two variables have been added to control for the age of students. A student who is in time for the final examination should be born in 1994; however, some students were enrolled a year before (early), and some students were not admitted to the next grade during their past academic career (late).

Unfortunately, the dataset does not include information on the individual student's socio-economic status (SES), so I cannot control for this important characteristic. This point is strongly important here, and it must be borne in mind when interpreting the results – much of the variance at individual level is not explained because the lack of this important information.

When turning to school-level variables, the source of data is twofold. Part of the variables comes from the same Invalsi dataset (final examination of the lower secondary education, year 2007/08). Another important source was the TIMSS 2007 dataset, which refer to the year 2006/07: however, as I only matched variables recorded at school-level, I assume that variations in two adjacent years are very low indeed. An indicator is included to define if the school is public or private (dummy: Private). The number of students is considered to control for size effects (students). The proportion of students coming from disadvantaged families has been included to control for low socio-economic conditions of the students population (disadvantaged): this variable takes value 1 if the proportion is in the range [0;10], 2 if [11;25], 3 if [26;50], and 4 if [>50]. An indicator of the environment in which the school is located has been

¹ In the remainder of the paper, almost all results concern Math Scores, as no interesting patterns emerged for Reading.

controlled for, by including an ordinal variable considering the dimension of the city/town (community): the value is 1 if citizens are $[>500,000]$, 2 if $[100,000;500,000]$, 3 if $[50,000;100,000]$, 4 if $[15,000;50,000]$, 5 if $[3,000;15,000]$, and 6 if $[<3,000]$. We also considered how many instructional days (per year) have been administered in the school (*instr_days*). Finally, we control for the intensity of resource availability, by including two indicators: (i) the “shortage” of instructional materials (*short_instr*), and (ii) the diffusion of internet connections (*internet*). The former variable is recorded on a four-tiers scale as follows: (1=none, 2=a little, 3=some, 4=a lot), while the latter's scale is: (1=all, 2=most, 3=some, 4=a lot).

Finally, I considered differences according to the macro-area in which the school is located. Indeed, previous literature on the achievement of Italian students demonstrated that there are relevant differences across the different areas of the country, with schools located in the Central part of Italy performing worse than those in the North and better than those in the South (Bratti *et al.* 2007, Agasisti, 2009).

4 Results

The results of the multilevel analysis have been reported in the table 2.

The first column illustrates the coefficients estimated for the intercept, that is the average math score for all students in all schools. What is interesting here is the analysis of variance, which confirms how both between and within schools differences exist. The between-schools variance (in our model, τ^2) is 80.25, and within-schools variance (in our model, σ^2) is 284.57; thus, the latter is higher than the former. The most part of the variance is at student-level more than at school-level, even though also the latter plays a significant role – suggesting the existence of some degree of segmentation among schools.

The second column shows what happens where individual-level characteristics have been added. It is important to note that all the individual variables are statistically significant. Female students perform worse than their male counterparts (-1.89); disabled and foreign students have lower performance (10.81 and 4.44 points, respectively). Early students outperform the “regular” (born in 1994) ones (+4.44), while students who repeated one or more years suffer a disadvantage (-8.69).

When adding individual-level characteristics, within-schools variance decreases. The calculation on the part of the variance explained by such characteristics is as follows: $(284.57 - 271.74)/284.57 = 4.7\%$. As expected, these variables are not contributing too much in reducing within-schools variance, as we did not include individual-level socio-economic status.

Finally, the third column considers the model where school-level variables are included. Only statistically significant variables are reported.

All the individual-level variables remains significant and with similar coefficients. The number of students is positively associated with Math Score, suggesting the existence of a scale effect; the coefficient is very small, nevertheless it is probably due to the scale of this variable (the number of students). Finally, there is a positive effect associated to lower proportions of disadvantaged students in the school – and this effect is stronger for schools where this proportion is $<10\%$, lower (but positive) where the proportion is between 11% and 25%, and finally again lower (but still positive) where

this proportion is between 26% and 50% (the reference group is the school where the proportion is >50%).

When adding school-level variables, the between-schools variance decreases substantially $[(80.25-72.65)/80.25 = 9.4\%]$, suggesting that the socio-economic composition of the student body strongly matters for achievement.

Table 2: Results (math scores)

	Model 1 "Empty"	Model 2 "Individual level characteristics"	Model 3 "School and individual level characteristics"
Coefficients			
Intercept	53.5062 <i>0.0000</i>	55.4008 <i>0.0000</i>	46.3753 <i>0.0000</i>
Female		-1.8940 <i>0.0000</i>	-1.7647 <i>0.0002</i>
Foreign		-4.4435 <i>0.0000</i>	-4.5067 <i>0.0000</i>
Disabled		-10.8079 <i>0.0006</i>	-9.9355 <i>0.0024</i>
Early		4.4457 <i>0.0003</i>	5.0055 <i>0.0001</i>
Late		-8.6999 <i>0.0000</i>	-8.6900 <i>0.0000</i>
Private			4.0611 <i>0.4161</i>
Students			0.0084 <i>0.0147</i>
Disadvantaged 1			6.4861 <i>0.0153</i>
Disadvantaged 2			5.4362 <i>0.0453</i>
Disadvantaged 3			4.9317 <i>0.0891</i>
Random effects (variance)			
Between-schools variance	80.2533 <i>0.0000</i>	76.5023 <i>0.0000</i>	72.6594 <i>0.0000</i>
Within-schools variance	284.5740 <i>0.0000</i>	271.7446 <i>0.0000</i>	270.3104 <i>0.0000</i>

5 Discussion

A discussion about the differences between best and worst performers of the sample is proposed. The basic idea is to look at differences in covariates that can help to identify the major determinants of good results (considering schools as the unit of analysis).

The selection of the two groups ("high-performers" and "low-performers") has been obtained by analysing the residuals of the multilevel model, which express the effectiveness "adjusted for" of each school – that is, after considering the impact of the explanatory variables. Indeed, the residual term (U_{0j}) represents the effectiveness level of each school j_{th} , given that only the intercept has been assumed to vary among the

schools. Those schools for which U_{oj} is positive, and statistically different from 0, have been included in the “high-performance” group; those for which U_{oj} is <0 , are included in the group of “low-performance” schools. At the end, the sample of the “high-performers” comprises 35 schools, while that of “low-performers” includes 34 institutions. The picture that emerges is interesting¹. First, there are factors that do not discriminate the belonging to the “high-performers” and “low-performers” groups: the proportion of students enrolled one year before the regular track (early), the proportion of disabled students, the proportion of female students, the proportion of computers connected to the web at school.

On the other side, there are some characteristics that look associated with higher performances. On average, the schools of the “high-performers” group are characterized by:

- lower proportion of students who repeated one or more years (late);
- lower proportion of foreign students;
- slightly lower dimension (479 students vs 492);
- lower proportion of students coming from disadvantaged families;
- much lower problems with shortage of instructional materials.

It is also interesting to note that being a “high-performance” group school is also associated to a number of instructional days (*per year*) of 203 (instead of 204).

Overall, however, what determines the “high-performance” group status is much more linked to the composition of student body than to school's characteristics.

The Invalsi data do not include information about the socio-economic status of individual students. This is a strong limitation for the analysis, and it is evident because the other individual characteristics account for just 5% of the variance between students' scores. This problem raises a question about the future evaluations promoted by Invalsi; collecting data on this dimension is probably costly and time-consuming (as it involves parents) but can provide much more robustness to the results of the analyses. A possible solution that should be suggested to Invalsi is to define a sample of schools and students to involve in a more complete exercise, in which collect more high-quality information.

It is very weird that no differences between macroareas (Northern, Central and Southern Italy) emerge. Actually, it is a very counterintuitive result; as it was shown in the section §2, previous studies on this topic showed huge differences among the three macro-areas. The resulting absence of differences has been widely discussed in the descriptive report provided by Invalsi (2008), that suggests three potential hypotheses:

- (i) higher motivation for students in the South at the moment of a national (standardized) test, that also is important for their academic career (indeed, it is the final exam at the end of lower secondary education, and not an experimental test – as typically are the international ones);
- (ii) the items proposed in the test are more focused on the specific competences provided by Italian schools than others proposed elsewhere (e.g. OECD-PISA, TIMSS, Pirls, etc.), so differences are lower here than in other cases because the specific questions included in the national test are more specifically taught homogeneously in the different geographical areas (while competencies as measured by OECD-PISA etc. are more related to socio-economic conditions);

¹ Specific results are not presented here due to reasons of space constraints.

- (iii) there was an opportunistic behaviour by teachers in the schools of the South (e.g. cheating).

The analysis proposed by Invalsi for the year 2007/08 does not test these different hypotheses; nevertheless, in the report about the test in 2008/09 (Invalsi, 2009) they consider the latter as the most credible. Indeed, in the second national examination (2008/09) the final results have been corrected to take “cheating” into account, through a complex procedure based on scores obtained by the students during the previous years (details in Invalsi, 2009): such procedure generates a new set of scores, from which the North/South gap emerges again.

Thus, Invalsi supports the view that most likely there is a “cheating” opportunistic behaviour in some schools in the South in the year 2007/08. If it is the case, it is reasonable to assume that this specific result can partly change, when taking into account this effect. In terms of future analyses, it is important that Invalsi put more attention on this phenomenon, and adopts procedure to limit this problematic behaviour. However, future research will address this point also with available data. Indeed, a straightforward extension of the present paper is to use Invalsi’s 2008/09 “cheating-modified” data, matching them with TIMSS 2007. Nevertheless, in this case the hypothesis of the stability of schools’ characteristics would seem less credible.

References

1. Agasisti, T., The effect of competition on schools’ performance: preliminary evidence from Italy through OECD-PISA data, mimeo, Politecnico di Milano (2008).
2. Agasisti, T., The efficiency of Italian secondary schools and the potential role of competition. A Data Envelopment Analysis using OECD-PISA2006 data, mimeo, Politecnico di Milano (2009).
3. Bratti, M., Checchi, D., Filippin, A., Da dove vengono le competenze degli studenti? (*Which are the determinants of students’ competencies?*), Bologna, Il Mulino (2007).
4. Dronkers, J., Robert, P., Differences in scholastic achievement of public, private government-dependent, and private independent schools, *Educ. Policy*, 22(4), 541-577 (2008).
5. Fekjaer, S.N., Birkelund, G.E., Does the ethnic composition of upper secondary schools influence educational achievement and attainment? A multilevel analysis of the Norwegian case, *Eur.Sociol.Rev.*, 23(3), 309-323, (2007).
6. Hanushek, E., The economics of schooling, *J.Econ.Lit.*, 24, 1141-1177 (1986).
7. Invalsi, La prova nazionale al termine del primo ciclo, 2007/08 (*The results of the national examination at the end of lower secondary education, 2007/08*), Istituto Nazionale per la Valutazione del Sistema di Istruzione e Formazione, Roma (2008).
8. Invalsi, La prova nazionale al termine del primo ciclo, 2008/09 (*The results of the national examination at the end of lower secondary education, 2008/09*), Istituto Nazionale per la Valutazione del Sistema di Istruzione e Formazione, Roma (2009).
9. Lamb, S., Fullarton, S., Classroom and school factors affecting mathematics achievement: a comparative study of the US and Australia using TIMSS, *Aust.J.Educ.*, 46(2), 154-171 (2002).
10. Longobardi, S., Quintano, C., Castellano, R., The influence of social and economic factors on the competencies of Italian students. A multilevel analysis of data OECD-PISA2006, *Rivista di Economia e Statistica del Territorio*, 2, 109-149 (2009).
11. Vittadini, G., Buono scuola e sistema “misto” nell’istruzione: una nuova ipotesi di lavoro (*School vouchers and a “mixed” educational system: a new theoretical hypothesis*), in Violini, L. et al., Politiche sussidiarie nel settore dell’istruzione: il caso del buono scuola in Lombardia, Milano, Giuffrè editore (2007).