

# USING BIVARIATE MIXED EFFECT MODELS FOR ASSESSING EQUITY AND EFFICACY IN EDUCATION. AN ANALYSIS OVER TIME USING EU15 PISA DATA

Isabella Sulis<sup>1</sup>, Francesca Giambona<sup>1</sup> and Mariano Porcu<sup>1</sup>

<sup>1</sup> Department of Social Sciences and Institutions, University of Cagliari, (e-mail: [isulis@unica.it](mailto:isulis@unica.it), [fgiambona78@gmail.com](mailto:fgiambona78@gmail.com), [mrporcu@unica.it](mailto:mrporcu@unica.it))

**ABSTRACT:** This work investigates upon the determinants of students' literacy in mathematics and reading in Italy by analysing data collected in EU15 countries in three different waves (2006, 2009, 2012) by the Program for International Student Assessment ran by OECD. The main aim of our analysis is: (i) to assess the associations between students' skills in mathematics and reading and their socio-economics and cultural background and to investigate how these associations vary across countries and waves; (ii) to suggest value-added measures of performance at the different levels of analysis in the time span considered; (iii) to advance a system of indicators at country level which allows to investigate the performance of different countries in terms of efficacy (students' achievement) and equity (low variability in students' achievement) in the three waves. A 4-level multilevel regression model with a bivariate latent structure and random slopes at country level and school level has been adopted. Implications in terms of trends of efficacy and equity in comparisons across countries are discussed.

**KEYWORDS:** "multilevel models", "value-added measures", "pseudo-panel", "PISA", "education".

## 1 Introduction

Since the year 2000, the OECD carries out its Program for International Student Assessment (PISA). It is administered every three years to provide comparisons of students' achievement among the participating countries; it has completed in 2015 its fifth round. PISA collects information on all three areas of competencies (mathematics, reading and science) in terms of test scores, with a focus on one of the three competencies every three years. PISA test is administered to a group of students who attend the school sampled without taking into account the class group; in such a way, the PISA survey focuses more attention on the effective level of achievement gained by the students re-

regardless the content of teaching received. PISA survey targets students aged between 15.25 and 16.25 years at the time of the survey regardless the type of school where they are enrolled in. Detailed information on sampling design and procedures are available in a series of thematic and technical reports at <http://www.pisa.oecd.org>.

The aim of this work is to investigate upon the determinants of students' literacy in mathematics and reading in Italy by analysing data collected in EU15 countries in three different waves (2006, 2009, 2012) of the PISA survey. To keep minimal the assessment burden on each student and to avoid that the scaling of achievement would be influenced by the 'booklet effect' each student is asked to handle only a part of the assessment following a systematic booklet assembly and rotation procedure; for that reason PISA data provides five Plausible Values (PVs) of achievement both for math and reading for each student. The PVs represent the likely distribution of a student's proficiency and are provided to take into account the uncertainty associated with the estimates (Monseur & Adams, 2009; OECD, 2013a).

Four levels of analysis are detectable in the data structure: at level-1 students' PVs (5 for math and 5 for reading); at level-2 students, at level-3 schools (school-wave combination), and at level-4 countries (country-wave combination). The main aim of our analysis is: (i) to jointly assess the relationships between students' achievement in mathematics and reading and their socio-economics and cultural background and to investigate how this relationship varies across countries and waves; (ii) to suggest value-added measures of performance at the different levels of analysis in the time span considered; (iii) to advance a system of indicators at country level which allows to investigate the performance of different countries in terms of efficacy and equity in the three waves. A 4-levels Multilevel Regression (ML) model with a bivariate latent structure and random slopes at country level and school level has been adopted to single out the factors which seem to account for the greatest variability in students' achievement as well as to jointly model the unobserved heterogeneity across schools and countries and to detect divergences in the effects of socio-economics and cultural predictors (Goldestein, 2011; Grilli & Sani 2012; Sulis & Porcu, 2015).

## **2 Data description**

The PISA survey data collected in the EU15 countries for 2006, 2009 and 2012 waves are considered. Data refer to 1,869,065 students (111,857 in 2006, 129,726 in 2009, 132,211 in 2012), tested in 14,212 schools [school-wave

combination] (4168 in 2006, 4863 in 2009, 5181 in 2012) in the 15 countries in three waves of PISA survey.

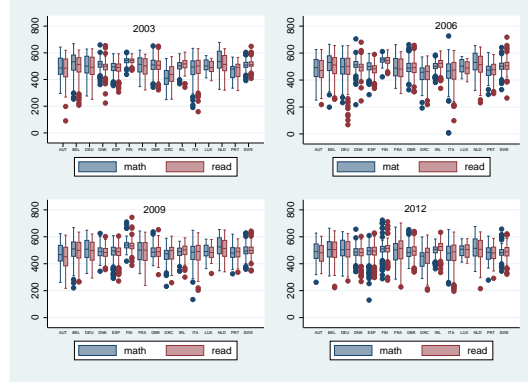
For each wave the metrical indicators of students' performance in mathematics and reading have been considered, namely students' performance scores in maths (MATH) and reading (READ). For each student we consider 5 PVs for maths and reading. Moreover each wave contains information on the following covariates: gender (SEX: 0 = female, 1 = male); non native (NONATIVE: 0 = native, 1 = non native); the country in which the survey has been administered (COUNTRIES: AUT = Austria, BEL = Belgium, DNK = Denmark, FIN = Finland, FRA = France, DEU = Germany, GRC = Greece, IRL = Ireland, ITA = Italy, LUX = Luxembourg, NLD = Netherlands, PRT = Portugal, ESP = Spain, SWE = Sweden, GBR = United Kingdom); the index of economic social and cultural status of the family (ESCS); the index of family possession of cultural resources (CULTPOSS); the index of family possession of educational resources (HEDRES); the index of home possessions (HOMEPOSS).

In order to depict differences across countries in the test results, Figure 1 plots the average scores at school level ( $\bar{y}_j$ ) in math and reading for the waves 2003, 2006, 2009, 2012 for the EU15 countries (OECD, 2009, 2012, 2013). The analysis highlights that the main differences across countries are in divergences in the variability between schools, with a polarisation between countries such as Italy, France, Netherlands, Austria, Germany and Belgium that are characterised by a high between school variability in the results and countries such as Ireland, Sweden, Finland and Denmark which are characterised by a low between school variability. This first result would suggest the presence of relevant differences in efficacy across the schools of the former group of countries. In order to investigate changes over waves at country and school levels and their relationship with the determinants of the achievement a multivariate-multilevel approach is carried out in the following.

## Modeling approach

We denote with  $y_{(pijc)(t)}$  the plausible value  $p$  (level-1 unit) of the score in maths or reading of student  $i$  (level-2) in school  $j$  (level-3) and country  $c$  (level-4) at time  $t$ . To jointly model the effect of predictors on both reading and maths a 4-level ML bivariate model has been specified with random slopes at school and country level (Leckie & Charlton, 2013; Sulis & Porcu, 2015):

$$y_{(pijc)(t)}^{(d)} = \alpha^{(d)} + x_{ijc}^{(d)} \beta^{(d)} + \theta_{0(c)(t)}^{(d)} + \eta_{0(jc)(t)}^{(d)} + \zeta_{0(ijc)(t)}^{(d)} + \epsilon_{(pijc)(t)}^{(d)} \quad (1)$$



**Figure 1.** Average score at school level  $\bar{y}_j$  by country and wave

where,  $d$  indicates which test the score refers to (math or reading) and  $t$  stands for the wave.

The variability in  $y_{(pijc)}$  is split in four components:  $\theta_{(c)}^t \sim N(\mathbf{0}, \Theta)$  the between-country variability at level-4;  $\eta_{(jc)}^t \sim N(\mathbf{0}, \Sigma)$  the between-school within-country variability at Level-3;  $\zeta_{(ijc)}^t \sim N(\mathbf{0}, \Phi)$  the between-student within-school variability at level-2 and the  $\epsilon_{(pijc)}^t \sim N(\mathbf{0}, \Omega)$  the between-plausible value within-student variability (level-1). In a second step, random slopes are introduced at level-3 and level-4 to allow the slope of ESCS to vary across schools and countries in each wave, adding to the predictors of equation 1 the two terms (Leckie & Charlton, 2013)

$$x_{ijc} \theta_{1(c)}^{(d)t} + x_{ijc} \eta_{1(jc)}^{(d)t}. \quad (2)$$

The posterior estimates at country level of the random intercepts  $\theta_{0(c)}^{(d)t}$  and slopes of ESCS  $\theta_{1(c)}^{(d)t}$  allow to make a comparative assessment in terms of efficacy and equity.

## Model Results

Different 4-level univariate ML Models have been run to select relevant covariates (demographic, socio-economic, cultural), levels of analysis, and the effect of relevant predictors (see Table 1, Models  $M0 - M4$ ). From model comparison (see Table 1) arises that among the univariate (i.e., modeling reading and maths

separately) models, (i) Model *M4* is the championed for analysing differences in achievement scores; (ii) the effect of socio-economics and cultural covariates does not change significantly in the three waves (model results related to single wave are not listed here, but are available on request); (iii) the availability of material resources conditional upon to other covariates has a negative effect on both test scores, with a greater penalisation on reading scores; (v) the availability of cultural resources in the family has a positive influence even taking into account differences in the *ESCS* values; (vi) males have better performances than females in maths whereas female performance are higher in reading; (vii) non native students perform worst in both tests. Looking at the effect of the *ESCS* arises that the size of the slopes significantly varies between schools and countries. The variance of random intercepts and *ESCS* slopes at country-level (level-4) and school-level (level-3) are both statistically significant. The results allow to asses as efficacy and equity vary across countries and schools in the three PISA waves.

On the basis of this first explorative results a bivariate level-4 models (see Model *M5* Table 1) has been fitted allowing intercepts and slopes to vary across countries and schools. The posterior predictions of the random slopes and random intercepts at country level provide information of change in efficacy and equity across countries and waves and are used for a comparative analysis. Results are summarised at country level in Figure 2. Results show the position of the country system in the three waves. From Figure 2 clearly arises the clustering of the countries with respect to their position under or over the average intercept and slope (both fixed at 0) and the trajectories of these parameters in the three waves. It is interesting to highlight the position of Italy, which is characterized by intercepts under the average in the three waves and a positive path in their values between 2006 and 2009-2012. The flat slopes of the *ESCS* index in the three waves denote that Italy with respect to most of the other countries (13 out 14) is characterized by a higher level of equity of students' performances with respect to the influence of the family's social economic conditions.

### 3 Conclusion

The use of Bivariate Mixed Effect Models for analysing PISA data allowed to assess differences in countries' performance in mathematics and reading in terms of efficacy and equity. The works is in still in progress. Specifically further researchers aim to better shape the analysis of countries trajectories of performance with respect the two criteria.

**Table 1. Model Comparisons and results**

PREDICTORS	UNIVARIATE MODEL					BIVARIATE MODEL						
	M0	M1	M2 Reading	M3	M4	M0	M1	M2 Math	M3	M4	M5 Math	M5 Reading
CONS	493.7	510.7	510.5	511.1	510.7	496.7	493.0	492.6	492.9	492.5	492.6	*
ESCS		15.0	16.1	15.2	16.1		14.6	15.8	14.8	15.7	15.6	*
CULTPOSS		11.2	10.9	11.1	10.9		8.0	7.7	7.9	7.7	7.6	*
HOMEPoS		-9.1	-9.1	-9.1	-9.1		-4.6	-4.5	-4.5	-4.5	-4.4	*
HEDRES		7.2	6.9	7.0	6.9		6.9	6.6	6.8	6.6	6.6	*
NONNATIVE		-31.6	-31.8	-31.6	-31.7		-30.4	-30.6	-30.5	-30.6	-30.6	*
SEX		-28.3	-28.0	-28.2	-28.0		18.9	19.2	19.0	19.2	19.2	*
WAVE_2009		<b>1.4</b>	<b>0.3</b>	<b>1.2</b>	<b>0.3</b>		-0.8	-1.0	-0.9	-1.1	-1.1	0.2
WAVE_2012		<b>4.7</b>	<b>4.6</b>	<b>4.4</b>	<b>4.4</b>		-4.4	-4.3	-4.5	-4.4	-4.7	4.2
Random Components												
Level-4 countries												
var(cons)	324.2	275.3	271.7	285.7	270.2	418.5	351.7	355.8	350.3	356.1	355.5	*
cov(cons\escs)			<b>37.2</b>		<b>37.0</b>			2.4		3.6	4.4	*
var(escs)			53.0		52.7			51.2		52.4	53.0	*
cov(cons.d1\d2)												259.8*
cov(cons.d1\escs.d2)												1.2*
cov(cons.d2\escs.d1)												37.6
cov(escs.d1.escs.d2)												50.1*
level-3 school												
var(cons)	3668.0	2617.0	2718.2	2626.4	2709.7	3182.0	2368.1	2626.4	2453.9	2455.0	2458.9	*
cov(cons\escs)				-62.8	-25.9				1.7	24.8	28.2	*
var(escs)				89.5	41.6				95.5	42.3	41.4	*
cov(cons.d1\d2)												2310.1*
cov(cons.d1\escs.d2)												23.6*
cov(cons.d2\escs.d1)												31.7*
cov(escs.d1.escs.d2)												35.9*
level-2 stud												
var(cons)	4697.6	4169.6	4128.7	4107.2	4099.7	4451.6	4033.8	3988.9	3964.7	3958.9	4000.8	*
cov(cons.d1, cons.d2)												4136.5
level-1 p-val												3467.17*
var(cons)	830.6	830.6	830.6	830.6	830.6	798.4	798.4	798.4	798.4	798.4	798.7	830.8
Statistics												317.04*
N	1869065	1868970	1868970	1868970	1868970	1869065	1868970	1868970	1868970	1868970	1868970	1869065.0
deviance	19172504	19125610	19122862	19124616	19122540	19092316	19053212	19049974	19052038	19049630	19049630	37478644.0
#params	5	13	15	15	17	5	13	15	15	17	17	44
BIC	19172576	19125798	19123079	19124833	<b>19122785</b>	19092388	19053400	19050191	19052255	<b>19049875</b>	19052255	<b>37479279</b>
AIC	19172514	19125636	19122892	19124646	19122574	19092326	19053238	19050004	19052068	19049664	19052068	37478732
#level 2	373794	373794	373794	373794	373794	373794	373794	373794	373794	373794	373794	373794
#level 3	14212	14212	14212	14212	14212	14212	14212	14212	14212	14212	14212	14212
#level 4	45	45	45	45	45	45	45	45	45	45	45	45

\*p-value < 0.01; non-significant parameters in M1-M4 have been highlighted in bold.



**Figure 2.** Country Intercept  $\theta_{0(c)}$  and ESCS Slope  $\theta_{1(c)}$

## References

- GOLDSTEIN, H. 2011. *Multilevel Statistical Models - 4th ed.* Wiley series in probability and statistics. Wiley & Sons.
- GRILLI, L., & SANI, C. 2011. Differential Variability of Test Scores Among Schools: a Multilevel Analysis of the Fifth-grade INVALSI Test Using Heteroscedastic Random Effects. *Journal of Applied Quantitative Methods*, **53**(6), 88–99.
- LECKIE, G., & CHARLTON, C. 2013. runmlwin - A Program to Run the MLwiN Multilevel Modelling Software from within Stata. *Journal of Statistical Software*, **forthcoming**.
- MONSEUR, C., & ADAMS, R. 2009. Plausible values: how to deal with their limitations. *Journal of Applied Measurement*, **10**(3), 320–334.
- OECD. 2009. *PISA 2006 Technical Report*. Paris: OECD Publishing.
- OECD. 2012. *PISA 2012 Technical Report*. Paris: OECD Publishing.
- OECD. 2013. *PISA 2012 Assessment and analytical framework*. Paris: OECD Publishing.