http://dx.doi.org/10.1016/j.seps.2019.05.005.

Adjusted indicators of quality and equity for monitoring the education systems over time. Insights on EU15 countries from PISA surveys

Summary. In this work, we investigate how European countries belonging to EU15 are performing in terms of the quality and equity of their educational systems. To do so, we jointly analysed student competencies in mathematics and reading using data collected in four different waves (2006, 2009, 2012, 2015) by the Program for International Student Assessment (PISA) run by the Organisation for Economic Cooperation and Development (OECD). The aim of this analysis is twofold: (i) to assess the associations between students' competencies in mathematics and reading and their socioeconomic status and to investigate how this relationship varies across countries over time; (ii) to present a batch of adjusted indicators relevant to the investigation of educational performance over time in terms of quality (average student competencies) and equity (low association between student achievement and their socioeconomic backgrounds). We fitted a mixed-effect (multilevel) regression model with a bivariate latent structure and random intercepts and slopes to assess the effect of socioeconomic and cultural background on student competencies across countries over time and to assess the performance trajectories of EU15 countries with respect to European Commission benchmarks. We present and discuss our main findings and their implications in terms of the policies of EU15 countries.

Keywords: education, OECD-PISA, multilevel models, adjusted indicators, quality and equity

1. Introduction

The success of a country's educational system implies a better educated population with more skilled and competent workers, which is strictly connected with the success of the educational systems whose main purpose is to enhance the worker skills and to ensure equal opportunities to people with different socioeconomic backgrounds. Thus, education plays a key role in the lives of individuals and promotes the economic development of countries by enhancing productivity, improving social development, and reducing social inequality. We assume the distribution of skills to be a determinant of inequality, and the relationship between individual skills and family background as central to social and intergenerational enhancement (OECD, 2012a).

Given this framework, this paper focuses on the assessment of the determinants of student achievements in mathematics and reading based on an analysis of data on student competencies collected in EU15 countries in four different waves (2006, 2009, 2012, 2015) by the Program for International Student Assessment (PISA)

(OECD, 2009, 2011, 2012b, 2016a), as promoted by the Organisation for Economic Cooperation and Development (OECD). PISA is the main body charged with monitoring how well European countries are progressing with respect to the 2020 European Union education and training benchmarks, which fix as education agenda priorities the enhancement of quality and the improvement in equity of European educational systems (OECD, 2016a). We determined the quality of education systems by monitoring both the average competency levels in the main academic skill areas investigated by the PISA surveys (namely, mathematics, reading and science) and the capabilities of countries and institutions to enhance (or mantain) their average performances over time. Educational equity concerns the capability of offering the same opportunities to students from different demographic, ethnic and socio economic backgrounds. Thus, as argued by Field *et al.* (2007), equity is mainly concerned with fairness and inclusion (OECD, 2012a). A fair education system ensures students of the possibility of achieving their educational potential and removes any kind of penalisation regarding their opportunities, based on their personal or social circumstances. Furthermore, it guarantees to all students the acquisition of a minimum level of skills that is required for successful integration in today's societies. It is widely recognised in the literature (Field *et al.*, 2007; OECD, 2012a) that the best education systems are those that combine high performances in both the quality and equity domains and which are able to boost students' resilience (Agasisti and Longobardi, 2017). Moreover, several recent studies prove as inequalities are often associated to educational policies at country level such as early tracking and grade repetition (Duru-Bellat and Suchaut, 2005; Raitano and Vona, 2016). Other authors highlight the influence of tracking considering also the effect of sorting policies that may influence the allocation of educational opportunities (Raitano and Vona, 2016; Contini *et al.*, 2017).

Moving from this framework, the main aim of this study is twofold: (i) to jointly assess the relationships between student achievements in mathematics and reading and their socioeconomic and cultural backgrounds and to investigate how this relationship varies across countries, time, and within the same country; (ii) to advance a set of indicators at the country level that enable the investigation of student performances over time in terms of quality and equity. In the pursuit of these twofold aim, the PISA data we considered presents a structure with at least three levels of clustering of the observations: students, schools (namely, school-wave combination), and country (namely, country-wave combination). Given this data structure, we fitted a bivariate multilevel model with random effects at the main levels of interest (e.g. country, school, student) to monitor the PISA learning outcomes in the EU15 countries overtime, while taking into account differences across waves in the observed outcomes (i.e., students' performances in maths and reading) (Sulis and Porcu, 2015; Grilli *et al.*, 2016). This approach allows us to shed some light on how the two learning outcomes affect each other at different levels (student, school, country) and over time. Moreover, the joint use of multilevel and bivariate methods can highlight peculiarities in achievements across the investigated countries in these two academic skill areas that are associated with individual, school or country differences (Aitkin and Longford, 1986; Roscigno and Ainsworth-Darnell, 1999; Fuchs and Woessman, 2007). Conducting the analysis over time proved useful in detecting the effect of policies established to enhance quality and equity in education outcomes and in assessing the performance trends of the different EU15 countries.

The paper is organised as follows. In Section 2, we discuss the main general findings arising from the literature in the recent decades, and, more generally, from large-scale international surveys assessing competencies in education. We also present the rationales that motivated this research effort. In Section 3, we provide information on the PISA survey and on the variables considered in the analysis. Section 4 outlines the methodology applied and advances proposals on how best to investigate trends in the education performances of countries over time. In Section 5, we present the main results arising from our data analysis. Sections 6 and 7, discuss main findings and final remarks, respectively.

2. Theoretical Framework

Worldwide, a broad literature in the field of education and economics is aimed at investigating the determinants of student achievement and to assess differences between countries. These studies typically search for empirical support on educational policies to improve students' achievement and learning (Sahlberg, 2007). As schooling is the main way for young people to accumulate human capital, it is basic to understand the relative importance of achievement determinants – student background, household possessions, school quality, community affluence, etc. This is especially important since social inequalities can reproduce themselves through educational inequalities (Lauer, 2003). For example, when in some developing nations or regions school characteristics are found to be more important than the home resources in promoting student achievement, policies to promote learning can be applied differently respect to countries or regions, where family resources have a greater impact on learning than schools and teachers (Heyneman and Loxley, 1983; Hanushek, 1995; Baker *et al.*, 2002). The first major scientific contributions of Mincer (1958), Schultz (1961) and Becker (1964) on the education economics have encouraged studies about the relationship between education and productivity, the distribution of wealth and, more generally, the economic and social development of countries (Romer, 1990). In short, a more educated population (say, a higher stock of human capital) enhances social well-being and ensures economic and social progress (Nelson and Phelps, 1966). The education of a population is highly instrumental and necessary for improving a country opportunities and growth (Schultz, 1971; Sakamota and Powers, 1995; Psacharopoulos and Woodhall, 1997) and a higher stock of skilled workers enhances the adoption of new technologies, increases productivity and generates economic growth and social progress (Mankiw *et al.*, 1992; Harmon *et al.*, 2001). Thus, education is a productive investment that can be considered to be as important as physical capital (Temple, 1999). Higher education is associated with markedly higher earnings, lower unemployment, higher labour force participation, lower criminality and better quality of life; and a highly performing educational system is thought to be fundamental in achieving national economic competitiveness (Hanushek and Luque, 2003; OECD, 2016a). Due to the technological progress, rising skill demands have made qualifications at the upper secondary level of education (general, technical or vocational) a minimum credential for successful labour market entry (Rangvid, 2003). Thus, the role of education in a country's economic and social development is not a trivial issue. If greater competencies denote greater human capital, then the effectiveness and the efficiency of education systems become a basic target, although economic and social contexts where students live and study have the main role (Currie and Moretti, 2003; Lam and Duryea, 1999). Generally, human capital is measured through the participation rate of the population in education and through the number of years of schooling (Barro, 2001; Barro and Lee, 1993), but the simple consideration that a year of study and training may not have the same value in all countries, has brought to consider also measures of human capital quality (Hanushek and Kimko, 2000). In particular, Hanushek and Kimko (2000) underline that the differences of growth among countries are significantly affected by human capital and its quality is influenced by cultural, family and scholastic conditions.

Although the theory is clear, the empirical evidence is less so and sometimes discordant (Krueger and Lindahl, 2001; Cohen and Soto, 2007). The unavailability of comparable statistical evidence and the difficulty in measuring latent variables such as knowledge make everything more uncertain (Barro and Lee, 1993). For this reason, among the useful tools for measuring the performance of an education system, surveys regarding skills, capabilities and competencies seem to be relevant (Tyler *et al.*, 2000). Specifically, surveys are useful in evaluating the education systems based on the school performances of their students (the future workers) (and, in assessing the effect of specific country policies (Hanushek and Wößmann, 2006)), even if the survey results do not exactly reproduce their attitudes and motivations. In any case, empirical analysis provides important evidences regarding the level and quality of school systems in different countries (Afonso and Aubyn, 2005; Hanushek and Wößmann, 2006; Agasisti and Longobardi, 2017) and enables comparisons between them (Ammermüller, 2004; Goldstein, 2004; Beese and Liang, 2010; Martins and Veiga, 2010; Thum *et al.*, 2013) as well as over time (Finnie *et al.*, 2013; Agasisti *et al.*, 2014; Nagy *et al.*, 2016; von Maurice *et al.*, 2017). Currently, many important surveys measure student knowledge, skills and abilities in a number of countries, e.g. the International Adult Literacy Survey (IALS) administered by the OECD and Statistics Canada; the Trends in Maths and Science Study (TIMSS) and the Progress in International Reading Literacy Study (PIRLS), both conducted by the International Association for the Evaluation of Educational Achievement (IEA) through its International Study Centre at Boston College; and, lastly, since 2000, the Program for International Student Assessment (PISA) by the OECD. Today, PISA is considered to be the main data source for monitoring the performance of educational systems in participating countries.

The PISA survey assesses the skills and knowledge in mathematics, reading, and science of students enrolled in private and public schools. The PISA target population comprises students aged between 15 and 16 years at the time of the survey, which is administrated every three years. In 2015, it completed its sixth round. The richness of information provided by the PISA surveys enables the exploration within and especially between countries of the varying and long-term effects of educational policies developed in recent decades.

One of the main *Europe 2020* strategy benchmarks is the reduction of the share of students classified in PISA surveys as low achievers (OECD, 2016a). The PISA 2015 survey confirms that the socioeconomic status has the highest impact on the outcomes at individual, school and country levels. The percentage of students from disadvantaged socioeconomic background is higher among low achievers. This evidence contributes to the ongoing socioeconomic differences in the educational outcomes and thus on the degree of equity in the education systems (OECD, 2012b; Agasisti *et al.*, 2014; OECD, 2016a).

Given this understanding, it is important to analyse the performances of educational systems and to monitor the trajectories of educational outcomes in pursuing benchmarks that combine high quality (student performances) with high degrees of equity (as measured by the variability of results across student groups). In an equitable system, we would expect there to be no divergences in educational outcomes due to individual, family or territorial factors; observed differences in achievement would be fully explained by differences in student attitudes and abilities (Field *et al.*, 2007; OECD, 2012b; Agasisti *et al.*, 2014; Borgona and Contini, 2014; OECD, 2016a; Contini *et al.*, 2017).

The availability of PISA data for different years and for the EU15 countries allows the analysis of student achievement trends between countries over time, the monitoring of progress made in terms of quality (e.g. increasing the average level of student competencies in a country) and equity (e.g. decreasing inequality due to socioeconomic and cultural status) and taking into account the heterogeneity of the populations surveyed.

3. Data

The PISA survey collects information in three areas of competency (mathematics, reading and science) based on test scores, with a focus on one of these three competencies every three years (in the first wave in 2000 the focus was on reading, in 2003 maths and in 2006 sciences). The PISA survey is a self-administered questionnaire that tests student skills and gathers information on several facets of each student's family, home and school background. Students are randomly selected in each sampled school. This study differs from other similar surveys, such as the Progress in International Reading Literacy Study (PIRLS) or Trends in International Mathematics and Science Study (TIMSS), which focus on classes and on the relationship between the taught and learned curriculum. PISA focuses on the effective level of student achievement regardless of the teaching content. To minimise the assessment load on each student and to avoid influencing the scaling of skills by the *booklet effect*, each student is asked to take only a part of the whole assessment following a systematic booklet assembly and rotation procedure. As such, rather than providing a single measure of achievement, the PISA database provides five plausible values (PV) of a student's score for each topic (reading, math and science). The use of PV allows the statistical uncertainty associated with the estimation of each student's achievement to be taken into account (Monseur and Adams, 2009; OECD, 2013) by reproducing the likely distribution of student competencies in each topic. Detailed information on the sampling design and scaling procedures are available in thematic and technical reports available at `http://www.pisa.oecd.org`.

3.1. Dataset description

In this study, we considered the PISA survey data collected in the EU15 countries during the 2006, 2009, 2012, 2015 waves. For any student in each wave, the PISA survey uses five metric indicators (PVs) of student competencies in maths (`MATH`) and reading (`READ`) (five PVs for maths and five for reading). In our analysis, the study data cover a total of 482,004 students in the four waves. Students are clustered in schools belonging to different countries. For each student, we considered the following information available in the four survey waves: country (`COUNTRIES`: Austria - `AUT`, Belgium - `BEL`, Germany - `DEU`, Denmark - `DNK`, Spain - `ESP`, Finland - `FIN`, France - `FRA`, Great Britain - `GBR`, Greece - `GRC`, Ireland - `IRL`, Italy - `ITA`, Luxembourg - `LUX`, Netherland - `NLD`, Portugal - `PRT`, Sweden - `SWE`); gender (`SEX`: 0 =female, 1 =male) and the socioeconomic and cultural status (`ESCS`) index available in the PISA databases†. We used the `ESCS` index as a key analysis indicator for detecting differences across countries in terms of equity and to monitor performance trajectories over time. Furthermore, we used a school-level compositional variable to take into account of average differences in socioeconomic and cultural conditions between schools. Specifically, we calculated the median value of the `ESCS` index calculated at school level (`SCHESCS(P50)`). Tables 1 (pt. A) and 2 (pt. B) in the Appendix present descriptive statistics calculated at the country level for each wave. In the following, we placed the PVs as the bottom level of our data structure. As such, in the analysis, we consider the fourth level of units at the bottom to correspond to the PVs. This data set contains 2,410,020 records and four unit levels: PV at level-1, student at level-2 (♯ 482,004 students), school (♯ 18,054,combination of schools and waves) at level-3 and country (namely ♯ 60 country-wave combinations) at level-4.

†The `ESCS` index is set up (by a Principal Component Analysis) considering three main features: the highest occupational status of parents, their highest educational level, and the index of family possessions. This last index is built up by OECD from students' responses to a wide range of questions linked to the availability at home of culture-related goods (such as works of classic literature, books of poetry, or works of art), educational resources (such as a quiet place to study, a computer for schoolwork, educational software, dictionaries, technical reference books or books useful to help with schoolwork), and ICT devices (educational software, a link to the Internet). The index about family possessions is calculated by using a model-based scaling procedure belonging to the family of Item Response Theory (IRT) applied to dichotomous or Likert-type responses to questionnaire items (OECD, 2013). The items composing the student's family possession change across waves and countries (comparability of the index across waves and countries is guaranteed)(OECD, 2013).

4. Modelling approach and model specifications

We denote with $y_{(pijc)^{(t)}}$ the plausible value p (level-1 unit) of the scores 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, we specified a 4-level bivariate model with random slopes at the school and country levels (Goldstein, 2011; Grilli and Sani, 2011; Leckie and Charlton, 2013; Sulis and Porcu, 2015; Grilli *et al.*, 2016)

$$y^{(d)}_{(pijc)^{(t)}} = \alpha^{(d)}_{pijc} + x'^{(d)}_{ijc}\beta^{(d)} + z'^{(d)}_{jc}\gamma^{(d)} \tag{1}$$

where, d indicates to which skill the score refers to (reading=1 or math=2), $y^{(d=1,2)}_{(pijc)^{(t)}} \sim N(MB, \Pi)$, x refers to a vector of s ($s = 1, \ldots, S$) predictors at the student level and z refers to a vector of compositional variables ($z = 1, \ldots, Z$) at the school level. The intercept α_{pijc} is a function of four random components, as follows:

$$\alpha^{(d)}_{pijc} = \alpha^{(d)}_0 + \theta^{(d)}_{(c)^{(t)}} + \eta^{(d)}_{(jc)^{(t)}} + \zeta^{(d)}_{(ijc)^{(t)}} + \epsilon^{(d)}_{(pijc)^{(t)}}. \tag{2}$$

Specifically, the variability in $y_{(pijc)}$ is split in four components: $\boldsymbol{\theta}_{(c)^t} \sim N(\mathbf{0}, \boldsymbol{\Theta})$ the between-country(-wave) variability at level-4; $\boldsymbol{\eta}_{(jc)^t} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, the between-school (-wave) within-country (-wave) variability at level-3; $\boldsymbol{\zeta}_{(ijc)^t} \sim N(\mathbf{0}, \boldsymbol{\Phi})$, the between-student within-school variability at level-2; $\boldsymbol{\epsilon}_{(pijc)^t} \sim N(\mathbf{0}, \boldsymbol{\Omega})$, the between-PV within-student variability (level-1). Each random term has a bivariate normal distribution with variance-covariance matrix that captures the variances in reading and math competencies at each level (e.g between countries, within schools in the same country, between students) as well as they covariances (Grilli and Sani, 2011; Sulis and Porcu, 2015; Grilli *et al.*, 2016).

We introduced random slopes at level-4 and level-3 to allow the `ESCS` slope to vary across schools, countries and waves. Thus, the random term $\theta^{(d)}_{(c)^{(t)}}$ in equation 2 is specified as a function of a random intercept $\theta^{(d)}_{0(c)^{(t)}}$ and a random slope $\theta_{1(c)^{(t)}}$ of the `ESCS` index, as follows:

$$\theta^{(d)}_{(c)^{(t)}} = \theta^{(d)}_{0(c)^{(t)}} + ESCS_{ijc}\theta^{(d)}_{1(c)^{(t)}}.$$

In a similar way, $\eta^{(d)}_{(jc)^{(t)}}$ in equation (2) is specified as follows:

$$\eta^{(d)}_{(jc)^{(t)}} = \eta^{(d)}_{0(jc)^{(t)}} + ESCS_{ijc}\eta^{(d)}_{1(jc)^{(t)}}.$$

The vector of random terms θ at the country-wave level (level-4) is specified to follow a multivariate normal (MVN) distribution, where

$$\begin{bmatrix} \theta^{(1)}_{0(c)^{(t)}} \\ \theta^{(1)}_{1(c)^{(t)}} \\ \theta^{(2)}_{0(c)^{(t)}} \\ \theta^{(2)}_{1(c)^{(t)}} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{\theta^{(1)}_{0(c)^{(t)}}} & & & \\ \sigma_{\theta^{(1)}_{0(c)^{(t)}},\theta^{(1)}_{1(c)^{(t)}}} & \sigma^2_{\theta^{(1)}_{1(c)^{(t)}}} & & \\ \sigma_{\theta^{(2)}_{0(c)^{(t)}},\theta^{(1)}_{0(c)^{(t)}}} & \sigma_{\theta^{(2)}_{0(c)^{(t)}},\theta^{(1)}_{1(c)^{(t)}}} & \sigma^2_{\theta^{(2)}_{0(c)^{(t)}}} & \\ \sigma_{\theta^{(2)}_{1(c)^{(t)}},\theta^{(1)}_{0(c)^{(t)}}} & \sigma_{\theta^{(2)}_{1(c)^{(t)}},\theta^{(1)}_{1(c)^{(t)}}} & \sigma_{\theta^{(2)}_{1(c)^{(t)}},\theta^{(2)}_{0(c)^{(t)}}} & \sigma^2_{\theta^{(2)}_{1(c)^{(t)}}} \end{bmatrix} \right)$$

$\alpha^{(1)}_0 + \theta^{(1)}_{0(c)^{(t)}}$ is the intercept for country c at time t with respect to reading competencies and $\alpha^{(2)}_0 + \theta^{(2)}_{0(c)^{(t)}}$ is the corresponding intercept for maths. The random component $\theta^{(d)}_{0(c)^{(t)}}$ represents the gap at time t between the EU15 average competencies in math or reading in country c. The random component $\theta^{(d)}_{1(c)^{(t)}}$ allows the `ESCS` slope parameter to vary across countries (for reading and maths). The `ESCS` slope for country c is provided by $\beta^{(d)}_{s^{(t)}} + \theta^{(d)}_{1(c)^{(t)}}$. The magnitude and the sign of the term $\theta^{(d)}_{1(c)^{(t)}}$ represents the gap between the slope parameter estimate in country c and that of the EU15 average. We set similar distributional assumptions to model the association between competencies in maths and reading between schools (level-3) within the same country. The association between quality and equity at country level on the two dimensions and across them (e.g., quality in reading and equity in maths and vice versa) is captured by the four covariances terms in the $\mathbf{\Theta}$ matrix.

We specified the vector of random terms η at level-3 (schools) to follow a MVN distribution, as follows:

$$\begin{bmatrix} \eta^{(1)}_{0(j)^{(t)}} \\ \eta^{(1)}_{1(j)^{(t)}} \\ \eta^{(2)}_{0(j)^{(t)}} \\ \eta^{(2)}_{1(j)^{(t)}} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} \sigma^2_{\eta^{(1)}_{0(j)^{(t)}}} & & & \\ \sigma_{\eta^{(1)}_{1(j)^{(t)}},\eta^{(1)}_{0(j)^{(t)}}} & \sigma^2_{\eta^{(1)}_{1(j)^{(t)}}} & & \\ \sigma_{\eta^{(2)}_{0(j)^{(t)}},\eta^{(1)}_{0(j)^{(t)}}} & \sigma_{\eta^{(2)}_{0(j)^{(t)}},\eta^{(1)}_{1(j)^{(t)}}} & \sigma^2_{\eta^{(2)}_{0(j)^{(t)}}} & \\ \sigma_{\eta^{(2)}_{1(j)^{(t)}},\eta^{(1)}_{0(j)^{(t)}}} & \sigma_{\eta^{(2)}_{1(j)^{(t)}},\eta^{(1)}_{1(j)^{(t)}}} & \sigma_{\eta^{(2)}_{1(j)^{(t)}},\eta^{(1)}_{1(j)^{(t)}}} & \sigma^2_{\eta^{(2)}_{1(j)^{(t)}}} \end{bmatrix} \right).$$

The association between quality and equity at school level on the two dimensions and across them (e.g., quality in reading and equity in maths and vice versa) is captured by the four covariances terms in the matrix $\mathbf{\Sigma}$.

We estimated the models by calling the runMLwiN routine available for STATA (Leckie and Charlton, 2013). In the estimation process, this routine makes use of the Iterative Generalised Least Squares (IGLS), which is equivalent to maximum likelihood whenever a Normality link function is chosen. We estimated the random terms $\nu = (\theta, \eta, \sigma, \zeta)$ using their posterior predictions in the following step: once the fixed parameters $\hat{\alpha}$, $\hat{\boldsymbol{\beta}}$, $\hat{\gamma}$ and the variance-covariance matrices of the random terms are estimated, they are then used to generate the posterior distribution of the residual terms for the intercepts and slopes, namely $\nu = (\theta, \eta, \sigma, \zeta)$, for all random effects specified at a given level (level-4 to level-2) using Bayes theorem.

These posterior estimates are also known as empirical Bayes estimates or the Best Linear Unbiased Predictions (BLUP) of the random effects. We then summarise information regarding $\boldsymbol{\nu} = (\theta_{0(c)}, \theta_{1(c)}, \eta_{0(j)}, \eta_{1(j)})$ using $E(\boldsymbol{\nu})$ and its related $sd(\boldsymbol{\nu})$. So, in summary, $\theta^{(d)}_{0(c)^{(t)}}$ represents an adjusted indicator of a country's quality at time t with respect to the learning outcome d, whereas $\theta^{(d)}_{1(c)^{(t)}}$ represents the related adjusted indicator of a country's equity.

4.1. Assessing differences in equity and quality

We use the expected posterior predictions of the random intercepts and slopes to test divergences in quality (i.e. differences between intercepts) and equity (i.e. differences between slopes) between countries in the four waves (Sulis and Capursi, 2013; Sulis and Porcu, 2015) and to identify countries that perform better or worse than the EU15 average. The posterior predictions of the random component of the country intercepts – $\hat{\theta}^{(1)}_{0(c)^{(t)}}$, $\hat{\theta}^{(2)}_{0(c)^{(t)}}$ – related to the two learning outcomes (reading (1) and maths (2)) account for divergences in the average competency levels in maths and reading between countries and waves, whereas the posterior predictions of the random component of the country slopes related to the two skill areas – $\hat{\theta}^{(1)}_{1(j)^{(t)}}$ and $\hat{\theta}^{(2)}_{1(j)^{(t)}}$ – account for differences in the effects of the `ESCS` across countries and waves.

$\hat{\eta}^{(1)}_{0(j)^{(t)}}$ and $\hat{\eta}^{(2)}_{0(j)^{(t)}}$ allow the average competency of the school j in country c in maths and reading to diverge from the average intercepts of EU15 countries. Thus, they identify the divergences in quality across schools in the same country, whereas $\hat{\eta}^{(1)}_{1(j)^{(t)}}$ and $\hat{\eta}^{(2)}_{0(j)^{(t)}}$ allow the slope of the `ESCS` to vary across schools in the same country; therefore accounting for the gap between the effect of `ESCS` in school j with respect to the average effect of `ESCS` in country c.

The posterior estimates of the random part of the country and school intercepts –$\hat{\theta}^{(d)}_{0(c)^{(t)}}$ and $\hat{\eta}^{(d)}_{0(j)^{(t)}}$– allow for making adjusted comparisons between countries, and between schools within the same country, in terms of quality. Specifically, a country with the random terms $\theta^{(d)}_{0(c)^{(t)}}$ above (or below) the average (i.e., while controlling for the heterogeneity in the composition of its students) is characterised by better (worse) performances and thus its educational system presents better (worse) quality standard.

The sign and the direction of the random components of the `ESCS` slope – $\theta^{(d)}_{1(c)^{(t)}}$ and $\eta^{(d)}_{1(j)^{(t)}}$ – allow us to make comparisons between countries and within schools in the same country in terms of equity. The steeper is the slope at country level, the higher is the effect of `ESCS` on the observed outcomes, while controlling for potential confounding factors that render the results not directly comparable (Draper and Gittoes, 2004).

At the country level, we consider the expected posterior predictions of the two random intercepts in the four waves to be adjusted indicators of country added-value with respect to the (dth) academic skill area (i.e., reading or maths). Namely, they measure the contribution of the country education system with respect to the average competency levels in maths or reading, while controlling for the effect of characteristics at the student and school levels (Aitkin and Longford, 1986; Goldstein and Spiegelhalter, 1996; Rasbash *et al.*, 2010; Goldstein, 2011; Sulis and Porcu, 2015; Grilli *et al.*, 2016).

At the school-level, we used the information provided by the random intercepts and slopes and the variance and covariance matrix of the random terms has been used to investigate the within-country between-schools variability in terms of quality and equity with the main aim to shed some light on the shape of the within country relationship between quality and equity indicators.

5. Model Results

5.1. Explorative analysis

To select relevant predictors, we carried out an explorative analysis comprising three steps. In the first two steps, we conducted separate analyses for maths and reading. In the Step 1, we adopted a model building procedure (see Table 1, Models `M0-3L:M5-4L`) to detect relevant levels of the clustering of the observations and relevant predictors at the student (level-2) and school levels (level-3), starting with a variance component model, which considers random intercepts at both levels (see, `M0-3L` in Table 1) ‡.

Table 1 approximately here

Table 2 approximately here

We then extended Model `M0-3L` to take into account the clustering of schools in countries (see model `M0-4L` Table 1).

The introduction of the students' level predictors `ESCS`, `SEX` in model `M1-4L` shows a decrease in variances of the random terms at student, school, and at country levels. Furthermore, a relevant share of the between-schools residual variability is explained by considering at the school level (see Model `M2-4L` Table 1) the median socioeconomic and cultural conditions of students at the same schools (`SCHESCS(P50)`) (Bratti *et al.*, 2007; Grilli and Sani, 2011; Agasisti *et al.*, 2014; Agasisti, 2014; Sulis and Porcu, 2015; Grilli *et al.*, 2016) (see Table 2).

In Step 2, we assessed the invariance of fixed predictors over time by running a model for each wave. As highlighted by the results shown in Table 2, differences in achievement between males and females (`GENDER`) are narrowed in 2015. Specifically, controlling for the `ESCS`, the magnitude and the sign of the coefficients indicate a reduction in inequalities related to gender. Based on these findings, in model `M3-4`, we introduced the interaction term- `MALEx15`, to take into account this difference in the effect of the predictor across waves.

‡Remember that, at the bottom level we have considered the PV that act as level-1 observations.

Lastly, to characterise the differences across countries and waves with respect to the effect of socioeconomic and cultural conditions on the two competencies, we introduced random slopes at the school and country levels (see `M4-4L` and `M5-4L` in Table 1). In that way, the `ESCS` slope was allowed to vary only at the school level in model `M4-4L` (Table 1) and at school and country level in model `M5-4L` (Table 1).

Comparisons of univariate models (i.e., those considering reading and maths separately) in terms of their goodness of fit statistics (see AIC and BIC in Table 1) indicate that `M5-4L` better fits the data. As such, we selected this model for analysing differences in competencies between countries in the four waves.

We used the main findings from the explorative analysis (see, Steps 1 and 2) in Step 3 to set up a 4-level bivariate model with random slopes at the country and the school levels for `ESCS` (see model `M6-4L` in Table 3). The model `M6-4L` results reveal that males display poorer performances than females in reading (about -30 expected point scores in 2006–2012 and about -17 expected point scores in 2015), whereas males show better performance in mathematics (about $+18$ in 2006–2012 and about $+14$ in 2015) (Matteucci and Mignani, 2011; Sulis and Porcu, 2015; Grilli *et al.*, 2016). However, as indicated by the coefficients of the interaction term for 2015, the gender gap narrowed in both tests in 2015 compared to previous waves (OECD, 2016a). On average the socioeconomic conditions has a significant positive effect. The median `ESCS` of the students in the schools has a significant effect on both skill areas. In summary, regarding the joint effect of socioeconomic and cultural conditions, we see that female students with a disadvantaged `ESCS` (1st quartile of the distribution of `ESCS` index values) who are enrolled in a school with a low `ESCS` index value calculated at the school level (1st quartile of `SCHESCS(P50)`), have an expected reading score that is about -33 point scores under the average (0.35 standard deviations below the average). In contrast, female students with a high `ESCS` (`ESCS` index equals to the 3rd quartile) and who are enrolled in a school with a high `ESCS` level (3rd quartile `SCHESCS(P50)`) have an expected score in reading about 43 point score over the average (0.45 standard deviations above the average). So, we can state that a difference of about 0.78 standard deviation in student test scores in reading is explained by differences in the student characteristics that are beyond the control of the educational systems. The observed gap in mathematics between the two profiles of students in disadvantaged and advantaged conditions is a little bit narrowing.

5.2. Detecting countries' trajectories of average performance using adjusted indexes of quality and equity

The posterior predictions of random intercepts and slopes at the country level provide information about shifts in quality and equity over time. We used these two indicators in a comparative analysis across EU15 countries in the four waves (Aitkin and Longford, 1986; Goldstein and Spiegelhalter, 1996; Bratti *et al.*, 2007; Braga and Checchi, 2010). Looking at the variance and covariances of the `ESCS` estimates at the country and school levels, we see that the size of the slopes varies significantly between schools and countries. The variance of the random intercepts and

`ESCS` slopes between countries and schools both differ significantly from the average, whereas their covariances do not significantly differ from 0. The standard deviation of the distribution of the random slope coefficients for `ESCS` is about 7.2 at the country level and about 7.0 at the school level (see Table 1).

Based on these results, we can assess how quality (average level of competencies) and equity (effect of socioeconomic and cultural conditions) vary across countries and schools in the four PISA waves considered.

Figure 1 shows a summary of the results at the country level for both skill areas. Along the x-axis are the posterior prediction estimates of the random component of the country intercept (labelled in Figure 1 and from here on `INTERCEPT`) and along the y-axis are the posterior predictions of the random component of the country slope (labelled in Figure 1 and from here on `SLOPE`). We set the axes origin at 0 (the EU15 average for both the estimated parameters). We divided the resulting plot into four quadrants to highlight the relative position of those countries that show indicators with averages higher (high intercepts with high slopes [+quality, −equity]: 1st quadrant) or lower ([−quality, +equity]: 3rd quadrant) than the average to those of countries that combine high intercepts with flat slopes ([+quality, +equity]: 4th quadrant) or vice versa ([−quality, −equity]: 2nd quadrant). According to the EU2020 strategy, countries must promote policies to achieve high quality standards combined with a high level of equity (OECD, 2016a; Agasisti *et al.*, 2014; Agasisti, 2014; Agasisti and Longobardi, 2014). Thus, positive trends would be characterised by trajectories that move toward the combination of *intercept higher than the average* with *slope lower than the average*, as identified in the 4th quadrant of the graph characterized by the [+ quality; + equity] pattern.

From Figure 1, we can clearly see the clustering of countries with positions lower or higher than the average intercept and slope (both fixed equal to 0) and the trajectories of these parameters in the four waves. With respect to the reading competencies, Great Britain shows an average relative improvement in terms of equity between 2006 and 2015, whereas no significant changes occurred in the quality dimension. We can see a similar outcome for the equity indicator for Greece. However, the positive performance of the country with respect to the equity benchmark is not combined with an improvement in the quality indicator, as we see that its 2015 and 2012 levels are among the lowest in the EU15 area. Germany and Portugal, despite starting from different quadrants, experienced improvement in both indicators in the four waves. Finally, Italy exhibited an improvement in the quality, which enabled it to enter into the 4th quadrant area in 2015.

In summary, in 2015, Sweden, Denmark, Finland, Ireland and Great Britain were in the region of worst performance (2nd quadrant), which were less than average in both dimensions, whereas Italy, the Netherlands, Germany, Portugal and Spain were in the region of best performance (4th quadrant). In terms of quality Ireland and Belgium are quite close to the average. France maintains its quality level above the average but slightly increases inequality. Spain and Portugal outperform in terms of quality but in terms of equity are close to the EU15 average. With respect to maths performances, many countries exhibited trends similar to those of reading. For instance, Italy, in average improved its performance in terms of quality in 2015 with respect to 2006, with a slight decrease of the equity indicator, whereas Great Britain improved its performance only in terms of equity. In 2015, Luxemburg, Sweden and Denmark were in the region of low quality and high inequality for maths. Great Britain improved its equity indicator and in 2015 reaches a level close to the average value. Spain and Portugal exhibited an improvement in terms of quality and maintained their performance in terms of the equity indicators. We also found the positive reading performance of Germany mirrored in the maths area.

5.2.1. *Considering statistical uncertainty*

To better characterise countries' trajectories and strictly focus only on significant improvements in terms of quality and equity with respect to the EU15 average, Figures 2, 3, 4, and 5 show plots of the estimates of the intercept and slope by country, wave and skill area together with the 95% confidence level. Regarding the quality of reading competencies, Figure 2 shows that in 2015 the performances of Portugal and Spain were significantly higher than the EU15 average; France and Germany, even if in some of the previous waves outperform, in 2015 have the lower bound of their confidence interval just on the EU15 average, whereas Great Britain, Denmark Greece and Sweden under perform (see Figure 2). Negative trends with respect to quality were experienced by Finland, Denmark and Sweden, with the three countries classified in 2015 in the 2nd quadrant (worst performances) in Figure 1. If we look at the equity of the educational systems (i.e. the slopes of `ESCS` index), in 2015 Italy, the Netherland, Austria, Germany and Greece performed well, showing slopes of `ESCS` flatter than the EU15 average, whereas Ireland, Sweden and Finland underperform. Lastly, to detect significant changes between 2006 and 2015, we built up pairwise confidence intervals (Goldstein, 2011) to make comparisons between pairs of parameters related to two years and we used it for comparing the two extreme waves (2015 and 2006) by checking them for any overlap §. The results highlight that in the quality dimension, Italy, Spain and Germany showed a significant improvement between 2006 and 2015 whereas in the equity dimension, Great Britain, Denmark and Greece showed a significant reduction of the inequalities due to differences in student socioeconomic and cultural status.

With respect to mathematics (Figure 4 and 5) we observe similar results. Spain and Germany improve in quality and the countries that significantly reduced its inequalities related to `ESCS` were Germany and Great Britain. It is interesting to note the position of Italy, for which both intercepts are below the average in 2006, as its path revels an increase in quality in both skill areas, thereby preserving its performances attained in terms of equity. The flat slopes of the `ESCS` index in the four waves indicate that Italy, in comparison to most of the other countries, is characterised by a higher level of equity in its students' performances with respect to the influence of the family's `ESCS`. If we combine all the empirical findings, we find that Italy, Portugal, Spain and Germany are the countries that exhibited performances in 2015 that were average in both skill areas considering both criteria (quality and equity) and did not worse the scores obtained in 2006. If we look at the correlation

§Results are available on request.

between country intercepts in 2006 and 2015 on the two domains, the moderate-low correlation coefficients between the two series, namely 0.64 in reading and 0.67 in maths, show a regression toward the mean of countries intercepts after the economic crises. Weaker is the regression toward the mean of inequalities related to socioeconomic and cultural differences as it is shown by the values of the correlation coefficients between the series of the slopes coefficient in the two years (0.88 for reading and 0.80 for maths). Moreover the correlation between intercepts and slopes in 2006 and 2015, even if weaker, shows a change of sign between the two years: moving from 0.14 to -0.20 for reading and from 0.22 to -0.10 for mathematics. Thus we may argue that on average the economic crises narrowed differences in quality between education systems, slightly decreased inequality and the association between quality and equity.

5.3. *Inspecting the within-country variability looking at average differences in the learning outcomes across schools*

Lastly, to better characterise the performances of the EU15 countries, we examined the between-school within-country variability in terms of quality and equity indexes to convey the heterogeneous scenarios within the EU15 education systems. By determining the adjusted indexes of countries' performances, we can assess only differences in the averages across countries. It is well known that two distributions with equal means can be very different in terms of their variability, and thus in terms of their inequality. Figures 6 and 7 show plots of the posterior predictions of the school random component of the intercept (labelled `INTERCEPT`) and of the school random component of the slope (labelled `SLOPE`), respectively, for waves, countries and skills. These plots are shown with the only aim to visually inspect the relationship between quality and equity within countries. The higher is the between-school variability with respect to quality, the higher is the range of variation in the schools' intercept on the x-axis, whereas the greater is the between-school variability in terms of equity, the higher is the range of variation of the slope on the y-axis. Countries with large differences in quality between schools are characterised by a high level of variability along the x-axis, whereas countries with large differences in equity between schools are characterised by a high level of variability along the y-axis. The shapes of the scatter in the diagram clearly depict differences in variabilities along both axes. Both figures show that the Netherlands, Germany, Italy, Belgium, Austria, and France have greater divergences in quality between schools. However, the 2015 data confirm a reduction in the within-country between-school variability in quality for Italy and Germany. We note that a high between-school variability indicates the presence of inequality with respect to school choice in these countries. Thus, it is an indicator of inequality across education institutions providing students with standard levels of competencies. None of the countries exhibit narrower differences in quality and equity between schools for the time span considered.

Figure 2 approximately here

Figure 3 approximately here

Figure 4 approximately here

Figure 5 approximately here

Figure 6 approximately here

Figure 7 approximately here

6. Discussion

As it is well known (OECD, 2016b) that student skills are closely linked to individuals' characteristics, the way resources are allocated in the school systems and the way the school systems assign students to programmes and classes. Most of the descriptive analysis and League Tables (Goldstein and Spiegelhalter, 1996) of countries that share their PISA results are based on unadjusted indexes of student skills (OECD, 2012a, 2016a,b). Therefore, they do not take into account that the observed divergences in skills across countries may be partially related to the heterogeneity of students, programmes, classes and schools with respect to key characteristics (Goldstein and Spiegelhalter, 1996; Leckie and Goldstein, 2009). Moreover, the reading of unadjusted indexes can lead to misleading conclusions when making international comparisons across countries. The main findings suggest a persistence (on average) of advantages related to sociodemographic and cultural conditions (e.g., gender, and socioeconomic conditions) but also confirm that inequalities related to gender narrowed in the last wave. Moreover, the median of the `ESCS` of the students within the schools has a significant effect only with respect to their reading performances. The main evidences for this is found by comparing the countries using the model-based adjusted indexes of quality and equity, as follows: (i) if we consider only the point estimates in 2015, the countries demonstrating performances higher than the average in both competencies and according to both criteria are Portugal, Germany and Spain. (ii) If we consider the uncertainty in the point estimates only Germany satisfies both conditions on both indicators whereas Italy performs better than the EU15 average in the two indexes of equity and on the average with respect to both indexes of quality. By combining reading in the main findings with the uncertainty measures, we see that in 2015 the countries that performed at least on the average level in both skill areas according to both criteria (quality and equity) and performed no worse than they did in 2006 are Italy, Portugal, Spain, Germany, Netherland and Belgium. (iii) Spain and Portugal also exhibit an improvement in terms of quality and maintain their performance in terms of equity indicators between 2006 and 2015. (iv) Great Britain is the only country that significantly reduced its inequalities related to `ESCS`. However, since the adjusted quality and equity indicators at the country level provide information on the average performances of a country with respect to the two domains, the findings do not capture differences in the variability of their distributions. Nonetheless, a joint reading of the between- and within-countries variability of the quality indexes show that the Netherlands and Italy also exhibited the highest within-country divergences in skills between schools. This highlights the fact that in countries with high between-school variability, the presence of inequality in student opportunities seems to be more related to school choice (e.g., effectiveness of school management, track field, qualification of teaching staff, school leadership) and geographical factors (urban vs. rural areas) than to the disadvantaged conditions of student families.

These findings are based on an analysis that show clusters of EU15 countries that have been able to combine quality and equity and those that have been able to reform their educational systems to improve their standards for overcoming initial disadvantages (see, for instance, Portugal). Although the recent recession had an impact on the investment in education, most EU15 countries have continued to implement policies aimed to improve student achievement and reduce the gap related to disadvantaged students. See for instance the policy implemented in primary schools in the Netherlands and in the United Kingdom to provide additional funding to schools that host disadvantaged students (OECD, 2012b). If we look at the countries that, based on these results, have maintained high performance over years, or that have shown improvements in their performances, we see a framework of education policies that contributed to boosting these education systems toward attaining standards above the average. If we analyse the characteristics of the education system in the Netherlands (a high performing country), we find a framework characterized by centralised policies combined with a high level of school autonomy OECD (2014b). Over the last decade in the Netherland, educational policies have been established to (i) increase the percentage of students from disadvantaged backgrounds in early childhood education programmes, (ii) prevent disparity in school composition by setting rules at local level with an open school choice system and (iii) provide guidance and support to students in transition into higher education levels. Moreover, the Netherland fixed as priorities the recruitment and training of high quality teachers and the strengthening of school autonomy (OECD, 2014b). These factors have led to an education system which is among the higher performing in OECD countries, and which has maintained its performances in terms of quality and equity across the four waves. Germany has linked its improvement across the four waves to policies addressed to (i) the integration of students with disadvantaged and immigrant backgrounds, (ii) the provision of increased autonomy to school leaders and (iii) the implementation of external school evaluation policies (OECD, 2014a). Spain increased its performance in terms of quality despite the economic crises and the related budget cuts. This result has been achieved in a centralized system via the enactment of a new law that set the following priorities (i) more autonomy to schools, (ii) new diagnostic tests and exit exams and (iii) the introduction of more vocational programs in the last year of lower secondary education (OECD, 2014d). In Great Britain the practice of monitoring schools (http://www.ofsted.gov.uk) via governmental agencies inspections helped policymakers to implement policies designed to support poorly performing schools by removing barriers related to their ability to address unstable leadership, high staff turnover and high rates of disadvantaged students. To do so, considerable attention has been paid to enhance practices in schools located in deprived areas. Also, we found the successful performances of Portugal to be closely linked to its extensive programme of primary school reforms undertaken in the education system between 2011–2015, in which the main aims were to change learning outcomes and teaching materials, introduce specialised teaching methods and strengthen teacher qualifications (OECD, 2014c, 2016a). Another important aspect arising from our analysis is the persistence of a high level of between-school within-country variability in countries such as Italy and the Netherlands, even after adjusting for socioeconomic and cultural disadvantages. Recent studies of the Italian system (see Grilli and Sani (2011); Agasisti and Vittadini (2012); Sulis and Porcu (2015); Giambona and Porcu (2018)) have highlighted this finding, which suggests that unobserved factors related to geographical component (type of economy, local policies, etc.) and school characteristics (dimension, student–teacher ratio, digital resources, training of teaching staff, teacher turnover rate, etc.) may play a key role in these differences across institutions. For instance, in Italy, this variability is mainly due to regional differences, with Northern regions showing the lowest level of between-school and between-class variabilities compared with southern regions. Recent reforms implemented in Italy to address inefficiencies in its education system (e.g., early school-leaving rates, high rate of grade repetition, etc.) prioritised expanding school autonomy, promoting digital innovation and promoting teacher recruitment. Moreover, resources have been also allocated to establish a more competitive system (OECD, 2017).

7. Conclusions

In this paper, we presented our joint analysis of PISA data for EU15 countries in the last four waves to identify similarities and differences across education systems in the pursuit of quality and equity in education. To this end, we used four model-based indexes to measure the average maths and reading skills of students and investigated the equity with respect to the influence of socioeconomic and cultural status of their students. In addition, to investigate the capability of the education system to raise both student skills and equity, we determined the trajectories of each country's performances in maths and reading with respect to the EU15 average and between the two extreme waves (2006 versus 2015). Specifically, we used a bivariate mixed-effect model to analyse data over time, which enabled a better understanding of the differences between EU15 countries with respect to student skills. We took into account the heterogeneity in the composition of the surveyed population between countries and waves due to student characteristics and other factors related to schools' socioeconomic compositions. To do so, we specified random intercepts and random slopes at the country and school level to set a framework of adjusted indicators related to quality and equity for analysing between- and within-country performances in the four waves. Our purpose was twofold: (i) to make a comparative evaluation across countries and identify performance trajectories for both evaluation criteria (between-country variability) and (ii) to investigate the within-country between-school variability in the two skills areas with respect to the quality and equity criteria. The main value-added component of the model-based indexes we adopted, as compared with other findings (OECD, 2012a, 2016a,b)), is that they allowed us to make comparative evaluations across countries while removing the effect of confounding factors due to the clustering of observations (e.g. country, school, wave) and social composition of the surveyed populations (e.g., socioeconomic background). We highlight the fact that the complexity and diversity of the education systems of the EU countries considered in the analysis, with respect to their history and heterogeneity of population composition, make it challenging to establish a simple framework for pointing out outperforming countries with respect to multiple criteria and determine the real impact of the policies and best-practices implemented. Moreover, the uncertainty of the results requires that caution be exercised in making pairwise comparisons and in detecting departures from the average. Despite this inherent complexity, in this paper, we have attempted to shed some light on the limitations of analysing EU15 countries' performances and improvements in educational outcomes based only on the average level of student skills, while ignoring other important factors such as the degree of equity of their education systems, the heterogeneity of subgroups with different socioeconomic conditions and geographical factors and the between-school and within-country variability. In addition, our analysis provides tools for monitoring the equity and quality of EU15 countries over the last decade while considering the uncertainty associated with point measures to assess differences between countries.

Appendix

Table 1a approximately here

Table 2b approximately here

References

Afonso, A. and Aubyn, M. (2005). Nonparametric approaches to education and health efficiency in oecd countries. *Journal of Applied Economics*, **7**, 227–246.

Agasisti, T. (2014). The efficiency of public spending on education: an empirical comparison of EU countries. *European Journal of Education*, **49**(4), 543–557.

Agasisti, T. and Longobardi, S. (2014). Inequality in education: can Italian disadvantaged students close the gap? *Journal of Behavioral and Experimental Economics*, **52**(1), 8–20.

Agasisti, T. and Longobardi, S. (2017). Equality of educational opportunities, schools' characteristics and resilient students: An empirical study of eu-15 countries using oecd-pisa 2009 data. *Social Indicators Research*, **134**(3), 917–953.

Agasisti, T. and Vittadini, G. (2012). Regional Economic Disparities as Determinants of Students' Achievement in Italy. *Research in Applied Economics*, **4**, 33–54.

Agasisti, T., Longobardi, S., and Regoli, A. (2014). Does public spending improve educational resilience? A longitudinal analysis of OECD-PISA data. *SIEP Working Paper*, **669**, 1–27.

Aitkin, M. and Longford, N. (1986). Statistical Modeling Issues in School Effectiveness Studies. *Journal of the Royal Statistical Society A*, **149**(1), 1–443.

Ammermüller, A. (2004). What Makes the Difference? Explaining the Gap in PISA Test Scores Between Finland and Germany. Mannheim.

Baker, D., Goesling, B., and LeTendre, G. (2002). Socioeconomic status, school quality, and national economic development: a cross-national analysis of the heynemanan loxley effect on mathematics and science achievement. *Comp. Educ. Rev.*, **46**, 291–312.

Barro, R. (2001). Human capital and growth. *American Economic Review*, **91**(2), 12–17.

Barro, R. and Lee, J.-W. (1993). International comparisons of educational attainment. *Journal of Monetary Economics*, **32**(3), 363–394.

Becker, G. (1964). *Human capital: A theoretical and empirical analysis with special reference to education.* University of Cicago Press, Chicago.

Beese, J. and Liang, X. (2010). Do resources matter? PISA science achievement comparison between students in the United States Canada and Finland. *Improving Schools*, **13**(3), 266–279.

Borgona, C. and Contini, D. (2014). Migrant achievement penalties in western europe: do educational systems matter ? *European Sociological Review*, **30**(5), 670–683.

Braga, M. and Checchi, D. (2010). Sistemi scolastici regionali e capacità di sviluppo delle competenze. Il divario delle indagini PIRLS e PISA. *Italian Journal of Social Policy*, **3**, 1–24.

Bratti, M., Checchi, D., and Filippin, A. (2007). Geographical Differences in Italian Students' Mathematical Competencies: Evidence from PISA 2003. *Giornale degli Economisti e Annali di Economia*, **66**, 299–333.

Cohen, D. and Soto, M. (2007). Growth and human capital: Good data, good results. *Journal of Economic Growth*, **12**, 51–76.

Contini, D., Tommaso, M. L. D., and Mendolia, S. (2017). The gender gap in mathematics achievement: Evidence from italian data. *Economics of Education Review*, **58**, 32–42.

Currie, J. and Moretti, E. (2003). Mother's education and the intergenerational transmission of human capital: Evidence from college openings. *Quarterly Journal of Economics*, **118**, 1495–1532.

Draper, D. and Gittoes, M. (2004). Statistical analysis of performance indicators in UK higher education. *J. R. Statist. Soc. A*, **167**(3), 449–474.

Duru-Bellat, M. and Suchaut, B. (2005). Organisation and context, efficiency and equity of educational systems: what pisa tells us. *European Educational Research Journal*, **4**(3), 181–194.

Field, S., Kuczera, M., and Pont, B. (2007). *Education and Training Policy No More Failures Ten Steps to Equity in Education: Ten Steps to Equity in Education*. Education and training policy. OECD Publishing, Paris.

Finnie, R., Mueller, R., and Wismer, A. (2013). Access and barriers to post-secondary education in canada: evidence from a longitudinal PISA dataset. SSRN eLibrary.

Fuchs, T. and Woessman, L. (2007). What accounts for international differences in student performance? a re-examination using oecd-pisa data. *Empirical Economics*, **32**(2), 433–464.

Giambona, F. and Porcu, M. (2018). School size and students' achievement. Empirical evidences from PISA survey data. *Socio-Economic Planning Sciences*.

Goldstein, H. (2004). Measuring educational standards. News. Significance.

Goldstein, H. (2011). *Multilevel Statistical Models*. Wiley series in probability and statistics. Wiley & Sons, 4 edition.

Goldstein, H. and Spiegelhalter, D. (1996). League Tables and Their Limitations: Statistical Issues in Comparisons of Istitutional Performance. *Journal of the Royal Statistical Society A*, **159**, 385–443.

Grilli, L. and 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.

Grilli, L., Pennoni, F., Rampichini, C., and Romeo, I. (2016). Exploiting TIMSS and PIRLS combined data: multivariate multilevel modelling of student achievement. *The Annals of Applied Statistics*, **10**(4), 2405–2426.

Hanushek, E. (1995). Interpreting recent research on schooling in developing countries. *World Bank Res. Obs.*, **10**(2), 227–246.

Hanushek, E. and Luque, J. (2003). Efficiency and equity in schools around the world. *Economics of Education Review*, **22**, 481–502.

Hanushek, E. A. and Kimko, D. D. (2000). Schooling, labor-force quality, and the growth of nations. *American Economic Review*, **90**, 1184–1208.

Hanushek, E. A. and Wößmann, L. (2006). Does educational tracking affect performance and inequality? differences- in-differences evidence across countries*. *The Economic Journal*, **116**(510), C63–C76.

Harmon, C., Walker, I., and Westergaard-Nielsen, N., editors (2001). *Education and earnings in Europe: A cross analysis of the returns to education*. Elgar, Cheltenham.

Heyneman, S. and Loxley, W. (1983). The effect of primary-school quality on academic achievement across twenty-nine high- and low-income countries. *Am. J. Sociol.*, **88**(6), 1162–1194.

Krueger, A. and Lindahl, M. (2001). Education for growth: Why and for whom? *Journal of Economic Literature*, **39**, 1101–1136.

Lam, D. and Duryea, S. (1999). Effects of schooling on fertility, labour supply, and investments in children with evidence from brazil. *Journal of Human Resources*, **34**, 160–192.

Lauer, C. (2003). Family background, cohort and education: a french–german comparison based on a multivariate ordered probit model of educational attainment. *Labour Economics*, **10**, 231–151.

Leckie, G. and Charlton, C. (2013). runmlwin - A Program to Run the MLwiN Multilevel Modelling Software from within Stata. *Journal of Statistical Software*, **52**(11), 1–40.

Leckie, G. and Goldstein, H. (2009). The Limitation of Using School League Tables to Inform School Choice. *Journal of the Royal Statistical Society A*, **172**(4), 835–851.

Mankiw, N., Romer, D., and Weil, D. (1992). A contribution to the empirics of economic growth. *Quarterly Journal of Economics*, **107**, 408–437.

Martins, L. and Veiga, P. (2010). Do inequalities in parents' education play an important role in PISA students' mathematics achievement test score disparities? *Economics of Education Review*, **29**, 1016–1033.

Matteucci, M. and Mignani, S. (2011). Gender differences in performance in mathematics at the end of secondary school in Italy. *Learning and Individual Differences*, **21**, 543–548.

Mincer, J. (1958). Investment in human capital and personal income distribution. *Journal of Political Economy*, **66**, 281–302.

Monseur, C. and Adams, R. (2009). Plausible values: how to deal with their limitations. *Journal of Applied Measurement*, **10**(3), 320–334.

Nagy, G., Lüdtke, O., and Köller, O. (2016). Modeling test context effects in longitudinal achievement data: Examining position effects in the longitudinal German PISA 2012 assessment. *Psycological Test and Assessment Modelling*, **58**(4), 641–670.

Nelson, R. and Phelps, E. (1966). Investment in humans technological diffusion, and economic growth. *American Economic Review*, **56**, 69–75.

OECD (2009). *PISA 2006 Technical Report.* OECD Publishing, Paris.

OECD (2011). *Education at a Glance 2009.* OECD Publishing, Paris.

OECD (2012a). *Equity and Quality in Education: Supporting Disadvantaged Students and Schools.* OECD Publishing, Paris.

OECD (2012b). *PISA 2009 Technical Report.* OECD Publishing, Paris.

OECD (2013). *PISA 2012 Assessment and analytical framework.* OECD Publishing, Paris.

OECD (2014a). *Education Policy Outlook Germany.* OECD Publishing, Paris.

OECD (2014b). *Education Policy Outlook Netherlands.* OECD Publishing, Paris.

OECD (2014c). *Education Policy Outlook Portugal.* OECD Publishing, Paris.

OECD (2014d). *Education Policy Outlook Spain.* OECD Publishing, Paris.

OECD (2016a). *PISA 2015 EU performance and initial conclusions regarding education policies in Europe.* OECD Publishing, Paris.

OECD (2016b). *PISA 2015 Results (Volume II): Policies and Practices for Successful Schools.* OECD Publishing, Paris.

OECD (2017). *Education Policy Outlook Italy.* OECD Publishing, Paris.

Psacharopoulos, G. and Woodhall, M. (1997). *Education for development: An analysis of investment choice.* Oxford University Press, New York.

Raitano, M. and Vona, F. (2016). Assessing students' equality of opportunity in oecd countries: the role of national- and school-level policies. *Applied Economics*, **48**(33), 3148–3163.

Rangvid, B. (2003). Do Schools Matter? The Influence of School Inputs on Student Performance and Outcomes. Mimeo.

Rasbash, J., Leckie, G., and Jenkins, J. (2010). Children's educational progress: partitioning family, school and area effects. *Journal of the Royal Statistical Society A*, **173**(3), 657–682.

Romer, P. (1990). Endogenous technological change. *Journal of Political Economy*, **98**, 71–102.

Roscigno, V. and Ainsworth-Darnell, J. (1999). Race, cultural capital, and educational resources: Persistent inequalities and achievement returns. *Sociology of Education*, **72**(3), 158–178.

Sahlberg, P. (2007). Education policies for raising student learning: the finnish approach. *Journal of Educational Policy*, **22**(2), 147–171.

Sakamota, A. and Powers, P. (1995). Education and the dual labour market for-Japanese men. *American Sociological Review*, **2**, 222–246.

Schultz, T. (1961). Investment in human capital. *American Economic Review*, **51**, 1–17.

Schultz, T. (1971). *Investment in human capital*. The Free Press, New York.

Sulis, I. and Capursi, V. (2013). Building up adjusted indicators of students' evaluation of university courses using generalized item response models. *Journal of Applied Statistics*, **40**(1), 88–102.

Sulis, I. and Porcu, M. (2015). Assessing divergences in mathematics and reading achievement in italian primary schools: A proposal of adjusted indicators of school effectiveness, social indicators research. *Social Indicators Research*, **122**(2), 607–634.

Temple, J. (1999). The new growth evidence. *Journal of Economic Literature*, **37**, 112–156.

Thum, A., Beblavỳ, M., and Teteryatnikova, M. (2013). Expansion of higher education and declining quality of degrees. NEUJOBS Working Paper No. 4.4.2B.

Tyler, J., Murnane, R., and Willett, J. (2000). Do the cognitive skills of school dropouts matter in the labor market? *Journal of Human Resources*, **35**, 748–754.

von Maurice, J., Zinn, S., and Wolter, I. (2017). Large scale assessments: potential and challenges in longitudinal designs. *Psycological Test and Assessment Modelling*, **59**(1), 35–54.

Submission of the manuscript March 2018
Submission of the revised manuscript May 2019

Adjusted indicators of quality and equity for monitoring the education systems over time. Insights on EU15 countries from PISA surveys

Isabella Sulis[1], Francesca Giambona[2], Mariano Porcu[1]

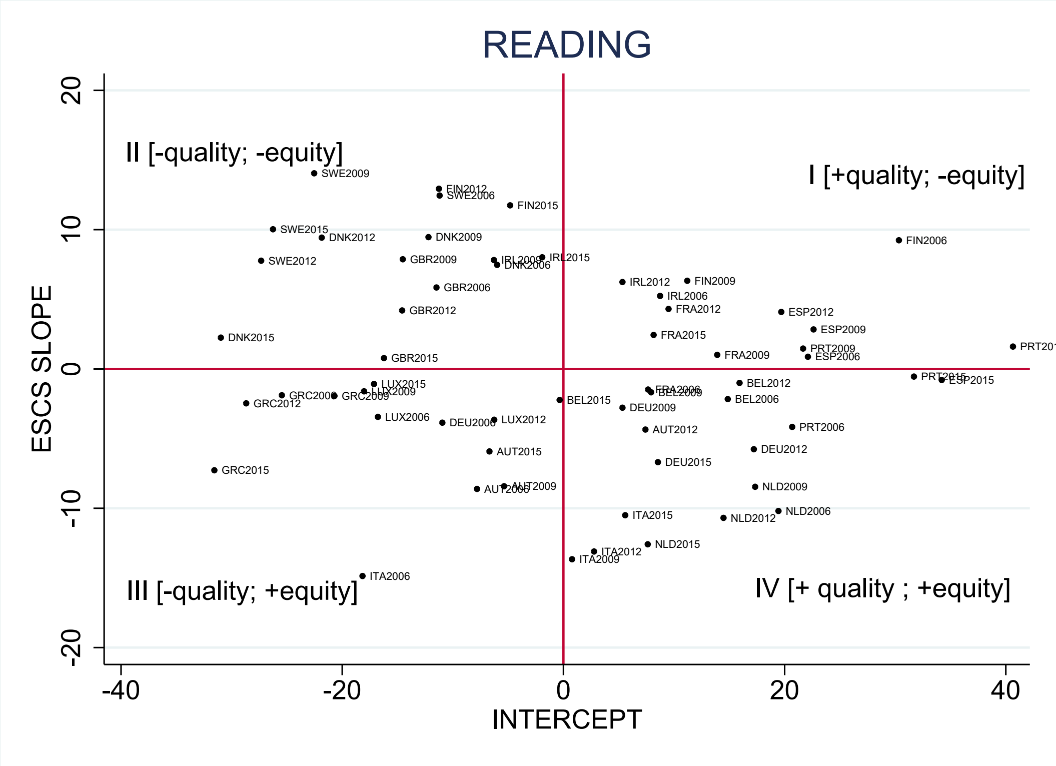
[1] Dipartimento di Scienze Politiche e Sociali, Università di Cagliari
Viale S.Ignazio 78, 09123 Cagliari, Italy

[2] Dipartimento di Statistica, Informatica e Applicazioni "G" Parenti", Università di Firenze
Viale Morgagni, 59, 50134 Firenze, Italy

Corresponding author: Isabella Sulis, isulis@unica.it

Abstract: In this work, we investigate how European countries belonging to EU15 are performing in terms of the quality and equity of their educational systems. To do so, we jointly analysed student competencies in mathematics and reading using data collected in four different waves (2006, 2009, 2012, 2015) by the Program for International Student Assessment (PISA) run by the Organisation for Economic Cooperation and Development (OECD). The aim of this analysis is twofold: (i) to assess the associations between students' competencies inmathematics and reading and their socioeconomic status and to investigate how this relationship varies across countries over time; (ii) to present a batch of adjusted indicators relevant to the investigation of educational performance over time in terms of quality (average student competencies) and equity (low association between student achievement and their socioeconomic backgrounds). We fitted a mixed-effect (multilevel) regression model with a bivariate latent structure and random intercepts and slopes to assess the effect of socioeconomic and cultural background on student competencies across countries over time and to assess the performance trajectories of EU15 countries with respect to European Commission benchmarks. We present and discuss our main findings and their implications in terms of the policies of EU15 countries.

keywords: education, OECD-PISA, multilevel models, adjusted indicators, quality and equity



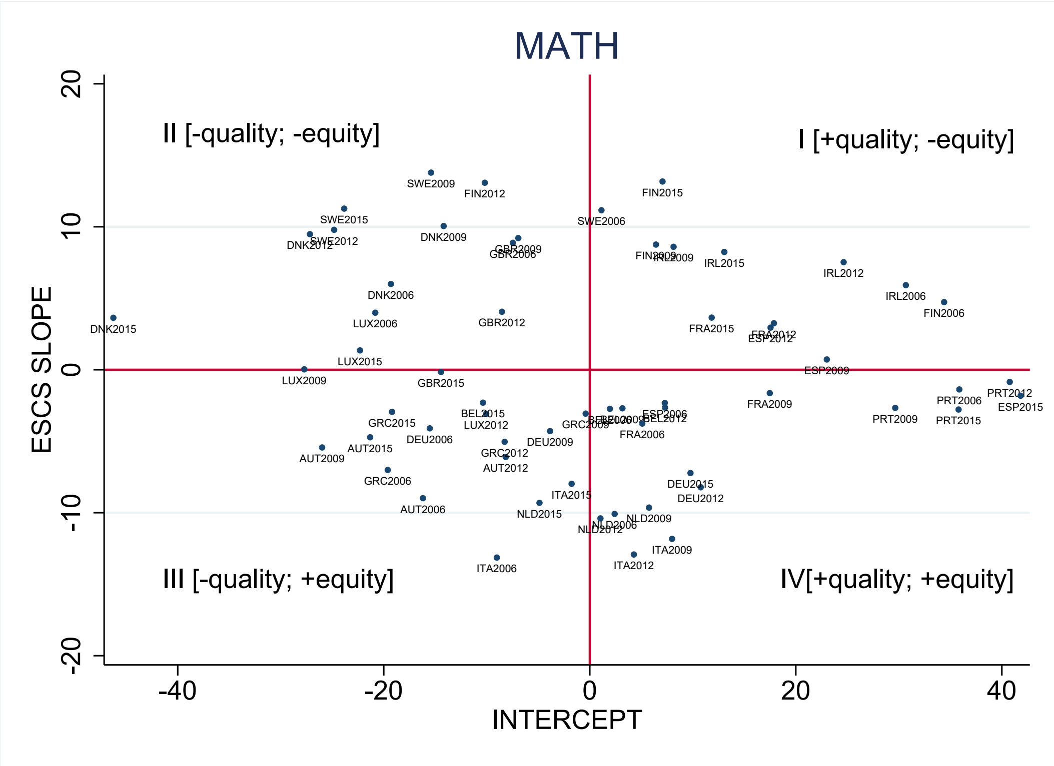


Figure 1: Country Intercept $\hat{\theta}_{0(c)}$ and ESCS Slope $\theta_{1(c)}$

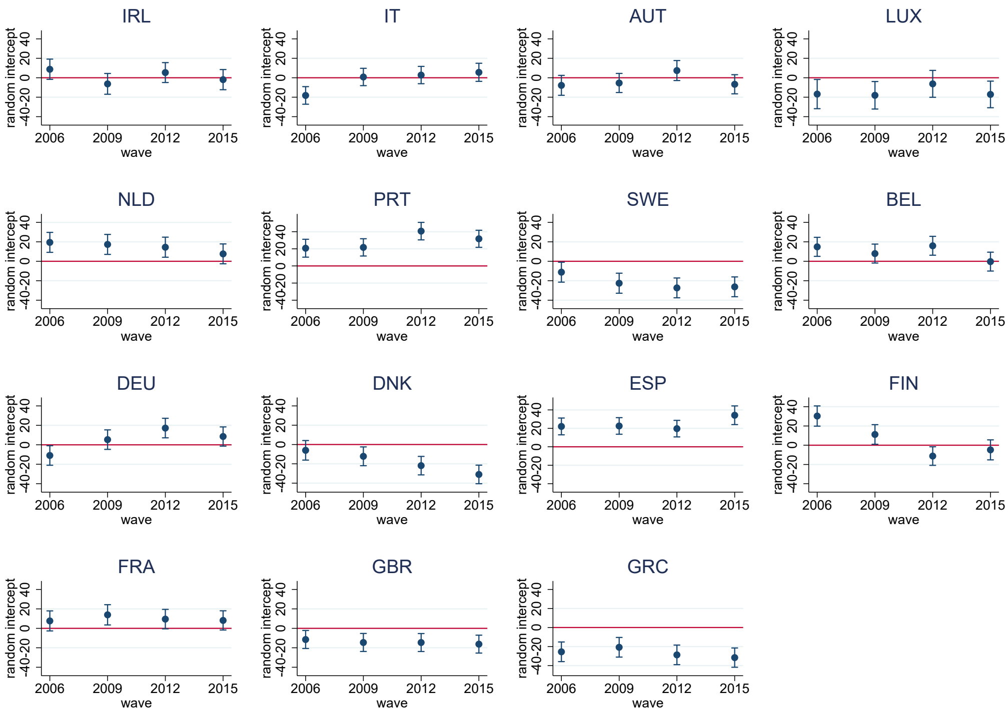


Figure 2: Reading: Country Intercept residual $\hat{\theta}_{0(c)}$ and 95% confidence intervals

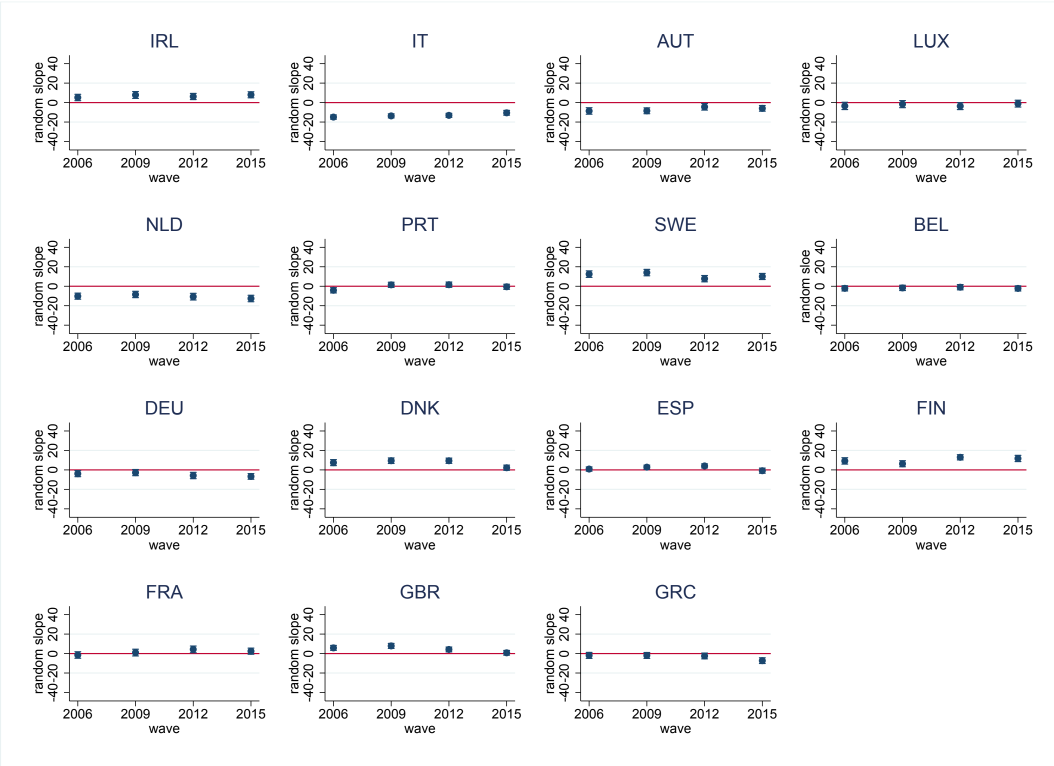


Figure 3: Reading: Country ESCS slope residual $\hat{\theta}_{1(c)}$ and 95% confidence intervals

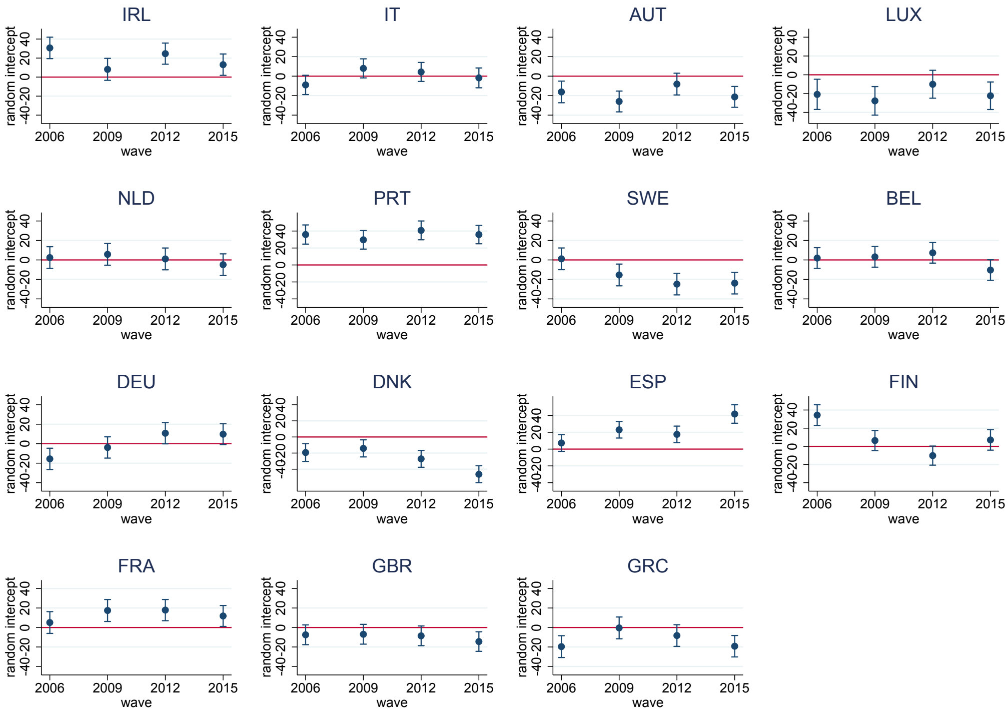


Figure 4: Mathematics: Country Intercept residual $\hat{\theta}_{0(c)}$ and 95% confidence intervals

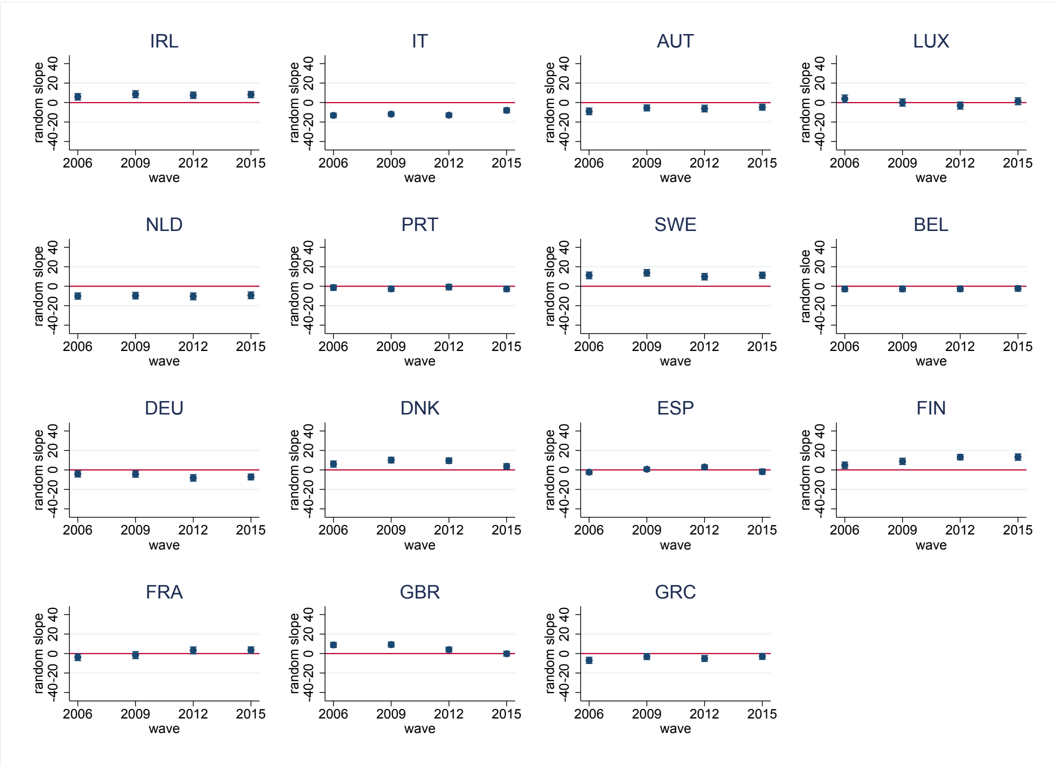


Figure 5: Mathematics: Country ESCS slope residual $\hat{\theta}_{1(c)}$ and 95% confidence intervals

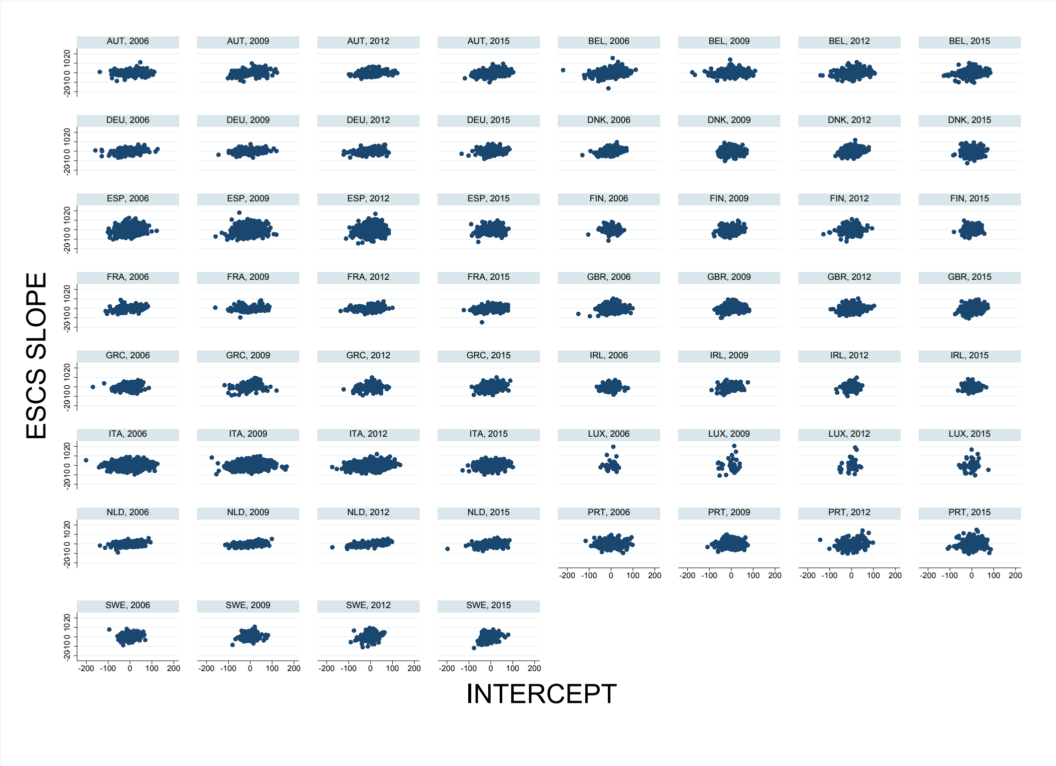


Figure 6: Reading: School Intercept $\hat{\theta}_{0(s)}$ and ESCS Slope $\hat{\theta}_{1(s)}$ by country and wave

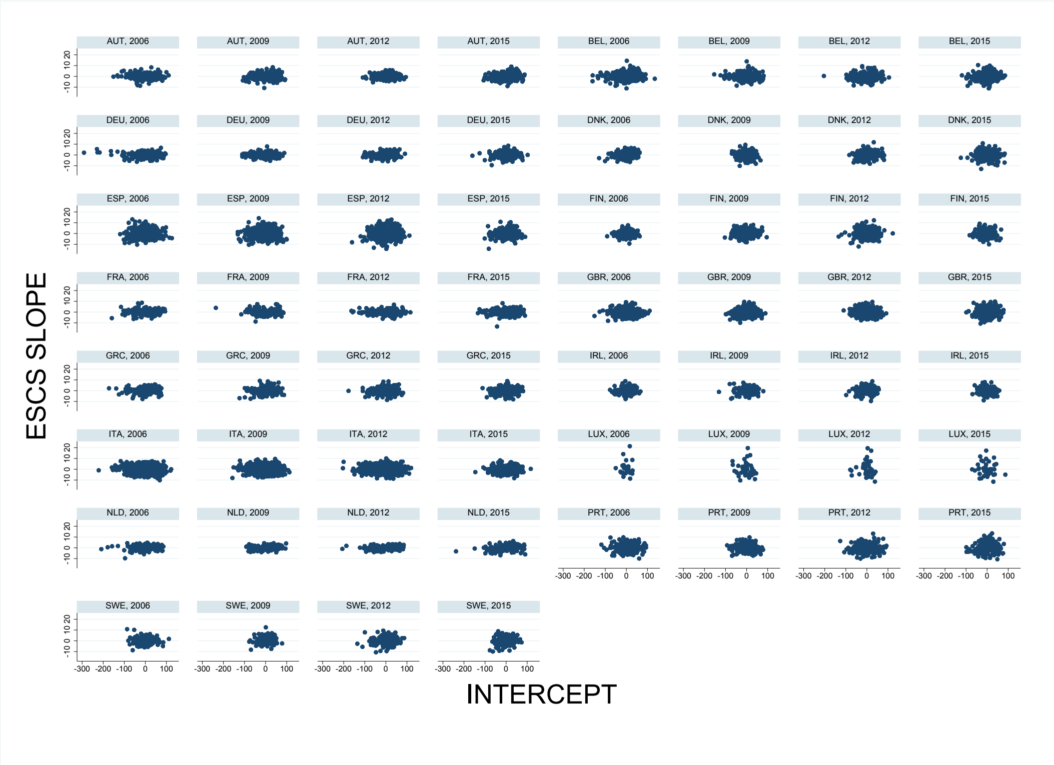


Figure 7: Math: School Intercept $\hat{\theta}_{0(s)}$ and ESCS Slope $\hat{\theta}_{1(s)}$ by country and wave

Table 1: Model Comparisons: 4-Level univariate regression models

Variables	M0 -3L		M0-4L		M1-4L:		M2-4L		M3-4L		M4-4L		M5-4L	
	read	math	read	math	read	math	read	math	read	math	read	math	read	math
CONS	490.8***	492.7***	493.5***	495.3***	505.7***	490.4***	503.7***	488.5***	505.2***	488.1***	505.3***	487.9***	504.3***	487.0***
	(0.476)	(0.449)	(2.304)	(2.526)	(4.091)	(4.380)	(4.901)	(4.455)	(5.071)	(4.455)	(4.952)	(4.544)	(5.059)	(4.637)
ESCS					18.97***	19.43***	17.62***	18.10***	17.64***	18.10***	17.48***	17.99***	19.11***	19.51***
					(0.121)	(0.117)	(0.122)	(0.118)	(1.487)	(0.118)	(0.149)	(0.147)	(0.956)	(0.953)
SEX (=M)					-27.76***	16.27***	-27.68***	16.47***	-30.59***	17.23***	-30.44***	17.38***	-30.19***	17.61***
					(0.211)	(0.205)	(0.211)	(0.204)	(1.140)	(0.231)	(0.238)	(0.230)	(0.237)	(0.229)
SCHESCS(P50)							59.07***	55.41***	59.04***	55.42***	59.83***	56.27***	60.39***	56.59***
							(0.599)	(0.573)	(5.245)	(0.573)	(0.606)	(0.582)	(0.610)	(0.585)
WAVE09					-0.983	-3.074	-2.452	-4.397	-2.462	-4.394	-2.605	-4.474	-2.493	-3.835
					(5.775)	(6.184)	(6.924)	(6.293)	(6.435)	(6.293)	(6.995)	(6.418)	(7.142)	(6.527)
WAVE12					2.080	-6.958	0.340	-8.578	0.343	-8.579	0.138	-8.773	0.264	-8.113
					(5.769)	(6.179)	(6.920)	(6.289)	(6.653)	(6.289)	(6.992)	(6.415)	(7.138)	(6.524)
WAVE15					3.038	-6.235	1.905	-7.316	-4.833	-5.563	-5.383	-6.000	-5.058	-5.832
					(5.773)	(6.182)	(6.922)	(6.291)	(7.737)	(6.296)	(6.999)	(6.422)	(7.145)	(6.531)
MALE15									13.36***	-3.476***	13.32***	-3.470***	13.18***	-3.559***
									(2.546)	(0.494)	(0.509)	(0.492)	(0.504)	(0.491)
Leve-4 var(θ_0)			298.3***	364.9***	233.8***	272.0***	348.8***	287.2***	348.5***	287.1***	356.1***	299.0***	373.4***	318.4***
Level-3 var(η_0)	3,832***	3,405***	3,610***	3,127***	2,745***	2,464***	1,694***	1,546***	1,691***	1,546***	1,697***	1,555***	1,745***	1,598***
Level-2 var(ζ)	4,792***	4,450***	4,792***	4,450***	4,444***	4,162***	4,444***	4,161***	4,437***	4,161***	4,364***	4,084***	4,357***	4,078***
Level-1 var(ϵ)	908.1***	856.2***	908.1***	856.2***	908.1***	856.2***	908.1***	856.2***	908.1***	856.2***	908.1***	856.2***	908.1***	856.2***
Leve-4 cov($\theta_0, \theta_1(escs)$)													-10.62	-22.29
Leve-4 var($\theta_1(escs)$)													53.23***	52.93***
Leve-3 cov($\eta_0, \eta_1(escs)$)											-15.86**	9.294	2.356	21.24***
Leve-3 var($\eta_1(escs)$)											104.2***	109.49	52.28***	52.49***
♯ Level-2	482004	482004	482004	482004	482004	482004	482004	482004	482004	482004	482004	482004	482004	482004
♯ Level-3	18,054	18,054	18,054	18,054	18,054	18,054	18,054	18,054	18,054	18,054	18,054	18,054	18,054	18,054
♯ Level-4			60	60	60	60	60	60	60	60	60	60	60	60
observations	2,410,020	2,410,020	2,410,020	2,410,020	2,410,020	2,410,020	2,410,020	2,410,020	2,410,020	2,410,020	2,410,020	2,410,020	2,410,020	2,410,020
number of groups	18,054	18,054	60	60	60	60	60	60	60	60	60	60	60	60
deviance	24903882	24754270	24903008	24753006	24864715	24719116	24856857	24711493	24856170	24711444	24854657	24709704	24851985	24706633
♯ levels	3	3	4	4			4	4	4	4	4	4	4	4
♯ param	4	4	5	5	11	10	11	11	12	12	14	14	16	16
AIC	24903886	24754274	24903013	24753011	24864726	24719126	24856868	24711504	24856182	24711456	24854671	24709718	**24852001**	**24706649**
BIC	24903941	24754329	24903081	24753079	24864877	24719263	24857019	24711655	24856346	24711620	24854863	24709910	**24852220**	**24706868**

Table 2: Model Comparisons and results: univariate multilevel regression models by wave

Variables	M2-4 L							
	read				math			
	2006	2009	2012	2015	2006	2009	2012	2015
CONS	505.2***	502.6***	505.6***	500.6***	488.7***	482.9***	479.7***	482.7***
	(5.022)	(4.073)	(4.442)	(5.828)	(4.456)	(3.688)	(4.741)	(4.790)
ESCS	16.59***	17.43***	17.90***	18.70***	17.11***	17.99***	18.84***	18.36***
	(0.255)	(0.225)	(0.229)	(0.273)	(0.243)	(0.220)	(0.229)	(0.254)
SEX (=M)	-30.99***	-30.02***	-30.66***	-17.40***	15.86***	18.92***	16.78***	13.71***
	(0.446)	(0.393)	(0.394)	(0.465)	(0.425)	(0.385)	(0.394)	(0.434)
SCHESCS (p50)	62.37***	58.32***	58.18***	57.70***	57.07***	54.37***	56.02***	54.16***
	(1.393)	(1.114)	(1.112)	(1.183)	(1.257)	(1.161)	(1.060)	(1.098)
Level-4 var(θ_0)	363.7***	238.9***	285.7***	500.2***	285.9***	193.3***	327.9***	336.2***
Level-3 var(η_0)	2,120***	1,550***	1,706***	1,388***	1,696***	1,723***	1,516***	1,195***
Level-2 var(ζ)	4,517***	4,255***	4,353***	4,683***	4,111***	4,011***	4,420***	4,068***
Level-1 var(ϵ)	954.3***	673.8***	881.6***	1,185***	837.7***	965.9***	602.1***	1,064***
# Level-2	112743	130739	133805	104717	112743	130739	133805	104717
# Level-3	4169	4863	5182	3840	4169	4863	5182	3840
#Level-4	15	15	15	15	15	15	15	15
Observations	563,715	653,695	669,025	523,585	563,715	653,695	669,025	523,585
Number of groups	15	15	15	15	15	15	15	15
time	54.04	63.16	66.22	46.57	52.05	63.38	61.13	46.97
deviance	5839229.7	6579083.4	6881973.2	5517108.9	5769226.5	6762422.1	6677644.6	5457436.9
numlevels	4	4	4	4	4	4	4	4

Standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 3: Results 4-Level Bivariate Models with random slope at Level-4 and Level-3

BIVARIATE M6-4L		
VARIABLES	READING	MATH
	STUDENT PREDICTORS	
CONS	503.7***	486.0***
	(4.838)	(4.391)
ESCS	19.11***	19.49***
	(0.958)	(0.956)
SEX (M=1)	-30.20***	17.65***
	(0.237)	(0.229)
MALEx15	13.18***	-3.574***
	(0.508)	(0.491)
WAVES		
$wave\_2009$	-2.413	-3.295
	(6.716)	(6.043)
$wave\_2012$	-0.886	-8.953
	(6.712)	(6.039)
$wave\_2015$	-1.397	-1.827
	(6.719)	(6.045)
SCHESCS (P50)	SCHOOL PREDICTORS	
	60.00***	56.41***
	(0.609)	(0.585)

RANDOM TERMS	Level-4 θ	Level-3 η	Level-2 ζ	Level-1 ϵ
var(cons_d1)	375.6***	1,745***	4,357***	908.1***
	(70.49)	(20.77)	(9.564)	(0.925)
cov(cons_d1\escs_d1)	-10.60	1.989		
	(18.86)	(5.794)		
var(escs_d1)	53.43***	52.06***		
	(10.73)	(3.096)		
cov(cons_d1\cons_d2)	293.1***	1,428***	3,606***	316.7***
	(60.43)	(18.51)	(8.527)	(0.675)
cov(escs_d1\cons_d2)	-36.50**	3.270		
	(18.07)	(5.579)		
var(cons_d2)	322.2***	1,599***	4,077***	856.2***
	(60.59)	(19.08)	(8.954)	(0.872)
cov(cons_d1\escs_d2)	5.874	26.01***		
	(18.80)	(5.677)		
cov(escs_d1\escs_d2)	50.74***	47.25***		
	(9.766)	(2.810)		
cov(cons_d2\escs_d2)	-20.83	21.96***		
	(17.62)	(5.412)		
var(escs_d2)	53.31***	52.43***		
	(10.01)	(2.949)		
deviance	48711903			
ll	-24355951			
#Levels	4			

Standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 1: Descriptive statistics (pt. A)

Waves	Country								**Total**
-Variables	**AUT**	**BEL**	**DEU**	**DNK**	**ESP**	**FIN**	**FRA**	**GBR**	**(EU15)**
Wave 2006									
-n	4859	8612	4596	4450	19415	4683	4588	12599	113385
-F (%)	49.89	48.14	49.65	51.51	50.14	50.78	51.55	50.72	50.13
-ESCS μ (σ)	0.20 (0.82)	0.20 (0.90)	0.31 (0.93)	0.31 (0.89)	-0.16 (1.03)	0.26 (0.79)	-0.08 (0.86)	0.19 (0.80)	0.03 (0.97)
-SCHESCS μ	0.12	0.15	0.25	0.30	-0.20	0.26	-0.12	0.16	-0.01
Wave 2009									
-n	6332	8238	4502	5703	25464	5744	4216	11780	131268
-F (%)	51.26	49.22	49.96	51.50	49.37	51.15	51.76	50.70	50.16
-ESCS μ (σ)	0.09 (0.83)	0.22 (0.92)	0.18 (0.91)	0.14 (0.93)	-0.25 (1.06)	0.41 (0.78)	-0.12 (0.84)	0.19 (0.78)	0.01 (0.98)
-SCHESCS μ	-0.03	0.12	0.14	0.12	-0.27	0.41	-0.17	0.17	-0.03
Wave 2012									
-n	4647	8209	4052	7216	24906	8564	4455	12012	134179
-F (%)	50.14	50.36	50.49	50.90	50.33	50.21	52.14	50.21	50.08
-ESCS μ (σ)	0.11 (0.83)	0.18 (0.91)	0.20 (0.93)	0.28 (0.91)	-0.11 (1.00)	0.35 (0.83)	-0.02 (0.80)	0.24 (0.81)	0.06 (0.95)
-SCHESCS μ	0.01	0.09	0.15	0.25	-0.15	0.34	-0.08	0.23	0.01
Wave 2015									
-n	6864	9299	5488	6907	6514	5758	5867	12969	104585
-F (%)	49.61	49.61	50.64	50.63	50.48	48.92	51.32	49.78	50.09
-ESCS μ (σ)	0.10 (0.85)	0.19 (0.89)	0.15 (0.94)	0.45 (0.95)	-0.44 (1.18)	0.26 (0.75)	-0.11 (0.79)	0.23 (0.86)	0.07 (0.96)
-SCHESCS μ	0.05	0.12	0.07	0.43	-0.49	0.24	-0.18	0.20	0.03

Table 2: Descriptive statistics (pt. B)

Waves	Country							**Total**
-Variables	**GRC**	**IRL**	**ITA**	**LUX**	**NLD**	**PRT**	**SWE**	**(EU15)**
Wave 2006								
-n	4810	4463	21595	4456	4815	5081	4363	113385
-F (%)	50.27	50.68	49.94	49.93	48.68	52.53	48.84	50.13
-ESCS μ (σ)	-0.11 (0.97)	0.00 (0.86)	-0.08 (0.96)	0.09 (1.10)	0.29 (0.88)	-0.58 (1.28)	0.23 (0.78)	0.03 (0.97)
-SCHESCS μ	-0.27	-0.02	-0.15	0.14	0.27	-0.61	0.24	-0.01
Wave 2009								
-n	4921	3711	30734	4527	4681	6243	4472	131268
-F (%)	51.43	50.28	49.33	50.17	50.91	52.22	49.78	50.16
-ESCS μ (σ)	0.03 (0.99)	0.06 (0.85)	-0.10 (0.98)	0.22 (1.09)	0.31 (0.85)	-0.30 (1.17)	0.34 (0.81)	0.01 (0.98)
-SCHESCS μ	-0.10	0.03	-0.15	0.23	0.29	-0.36	0.35	-0.03
Wave 2012								
-n	4991	4864	30760	5092	4311	5558	4542	134179
-F (%)	50.43	51.09	49.16	49.43	48.34	50.11	50.90	50.08
-ESCS μ (σ)	-0.04 (0.99)	0.14 (0.85)	-0.03 (0.95)	0.08 (1.10)	0.22 (0.78)	-0.48 (1.17)	0.29 (0.81)	0.06 (0.95)
-SCHESCS μ	-0.18	0.10	-0.12	0.14	0.21	-0.52	0.30	0.01
Wave 2015								
-n	5354	5516	11309	5140	5282	7176	5142	104585
-F (%)	49.22	49.84	50.27	50.84	50.40	49.86	50.62	50.09
-ESCS μ (σ)	-0.01 (0.95)	0.16 (0.84)	-0.04 (0.92)	0.08 (1.10)	0.17 (0.76)	-0.55 (1.15)	0.35 (0.82)	0.07 (0.96)
-SCHESCS μ	-0.13	0.13	-0.11	0.16	0.14	-0.63	0.34	0.03

Isabella Sulis is Associate Professor in Social Statistics at the University of Cagliari. Her research interests include multilevel analysis, item response theory, and measurement models to build-up adjusted indicators of performance, with main focus toward the evaluation of educational institutions.

Francesca Giambona is Researcher in Economic Statistics at the University of Florence. The focus of her academic researches is mainly related on risk modeling in economic field, school system evaluation and students' achievement, analysis and measurement of well-being.

Mariano Porcu is Professor of Social Statistics at the University of Cagliari. His research activities are mainly related to the analysis of student achievement and on the construction of composite indicators to evaluate the performance of educational institutions.