How does schools' efficiency look like across Europe? An empirical analysis of Germany, Spain, France, Italy and UK using OECD PISA2012 data

Tommaso Agasisti
Politecnico di Milano School of Management

Department of Management, Economics and Industrial Engineering
e. tommaso.agasisti@polimi.it
t. +39 02 2399 3963

*** Version #2 ***

Abstract.
This research conducts a comparison of secondary schools' efficiency in an international perspective, focusing on five economies in the European Union (UK, Germany, France, Italy and Spain) and employing an institution-level dataset built through data from the 2012 edition of the OECD's Programme for International Student Assessment (PISA). Overall, around 2,700 schools from these five countries are included in the empirical analysis; it uses a bootstrap version of Data Envelopment Analysis (DEA), and a common (international) frontier of efficient schools is assumed. The production process is modelled in a very simple way, by including human and capital resources, together with students' socioeconomic background, among inputs; and average scores in reading and mathematics, as outputs. Although within-country dispersion of efficiency scores is much wider than between-countries differences, some average between-countries efficiency differentials can be observed. A second-stage tobit regression reveals that some factors are statistically associated with schools' efficiency, as for example the indexes for the quality of educational resources and teachers' morale. Conversely, the efficiency scores are inversely correlated with the proportion of students who perform below proficiency level 2, suggesting that there is complementarity between efficiency and equity. All these evidences can stimulate interesting reflections for national and European-based policy-makers.

Keywords.
Schools' efficiency, equity, OECD-PISA2012, bootstrap DEA, cross-country comparison

JEL Codes.
I21, I28, C14, H52

Acknowledgments.
This research is funded under the TJ Alexander Fellowship program of OECD; I am grateful to the Directorate for Education (EDU) for hospitality during the work for this draft (Paris, Summer 2014). At OECD, I also received great comments and suggestions, during several and fruitful discussions; I am particularly grateful to F. Avvisati, F. Borgonovi, G. Montt, P Santiago, D. Nutsche, C. Shewbridge, P. Zoido. J. Wolszczak-Derlacz provided useful indications for conducting the double-bootstrap procedure. All eventual errors are, obviously, my solely responsibility.

1. Introduction and objectives

Among the key debates in the field of educational policy and politics, the one about the efficiency of public spending stands as a priority. A well-know stream of the academic literature criticises the effectiveness of the so-called input-oriented policies, claiming that there is no evidence about the relationships between higher investments and better academic results (Hanushek, 1986; 1989; 1997; 2003). This literature also highlights that, when comparing students' performances across countries, institutions (such as accountability, autonomy, etc.) matter much more than spending in explaining performances' differentials (Hanushek & Wößmann, 2010). Therefore, other scholars argue that it is important to study the determinants of different efficiency levels obtained by various countries. Indeed, some educational systems are able to "produce" better results for each dollar spent, and this needs to be somehow explained (in this field, the main contributions are those by Afonso & St. Aubyn, 2005 and 2006; but see also Clements, 2002 and Agasisti, 2014). In addition, taking efficiency of spending into account is a specific, main objective of the European Commission, as confirmed since its recommendation in 2006 (EU Commission, 2006).

In all Europe, there is a recurrent debate about the importance of spending more (public) money on education; the evidence that the level of human capital is strictly correlated with economic growth and social development reinforces the arguments of policy-makers, analysts, teachers, etc. for providing more investments in this field. The financial problems raised by the persistent financial and economic crisis exacerbate the necessity of make every dollar the most value for "producing" good quality education. In this perspective, there is still wide variation across European countries in the ability of obtaining good average results (as measured by standardized test scores, in international exercises like PISA, TIMSS and PIRLS[1]), given the level of resources spent in the educational system. The Figure 1 shows, with illustrative purposes, the graphical relationship between the average country's score in mathematics (as resulting from PISA 2012), in the vertical axis, and the expenditure per student in secondary schools (in Purchasing Power Parity dollars, 2011), on the horizontal axis. Although a general positive relationship can be detected, there are heterogeneous situations that deserve more attention and claim for a better understanding of the factors behind efficiency, defined as the ratio between output and resources. For instance, Finland and UK have similar spending level per pupil, but the students in the former country perform on average much better than their counterparts in the latter. A similar situation exists for Germany, France and Sweden – all the countries with similar expenditure per student, and different average test scores (514, 495 and 478 respectively).

[Figure 1] around here

In this paper, I estimate the efficiency scores for a sample of around 2,500 schools in the five largest European countries (i.e. Italy, Germany, Spain, UK and France), with the aim of comparing each school's efficiency with an international benchmark. Efficiency is defined in its technical sense, which is the ability of transforming inputs (students' starting background

[1] These are the acronyms for the three widest and most famous analyses of achievement scores in an international perspective: PISA is the Programme for International Student Assessment, managed by OECD;

characteristics, human and capital resources) in outputs (achievement scores, as measured by PISA test scores). This concept resembles the one introduced by Farrell (1957) and contextualized in the field of education by Levin (1976). The empirical analysis uses advanced robust versions of the technique called Data Envelopment Analysis (hereafter), by employing the bootstrapping procedures suggested by Simar & Wilson (2000). The dataset is built for having the single school as the unit of analysis; all data refer to the PISA 2012 edition. A battery of robustness checks is conducted for testing the sensitivity of results to different specifications of the efficiency model. After having estimated the efficiency scores for each school, they are regressed against a set of school-level variables, to investigate whether schools' characteristics (i.e. resources, practices and processes, etc.) are statistically correlated with their efficiency. This would also allow describing the main factors that determine the international frontier of efficient schools, and this can in turn suggest policy and managerial reflections.

The remainder of the paper is organised as follows. Section 2 briefly reviews some relevant contributions from the academic literature. Section 3 is devoted to describing the methodological approach and the dataset. Section 4 contains the main results about schools' efficiency and its determinants. Section 5 presents the robustness checks and the corresponding results. Section 6 hosts concluding remarks.

2. Existing literature

Reviewing the vast literature about the efficiency of schools is not the primary objective of this paper, and it would require a specific research focus – the interested reader can refer to Worthington (2001) and Johnes (2004) for early reviews. From this rich group of studies, however, some general lessons were drawn and contributed to some relevant aspects of my research. First, there is a widespread, growing consensus about the use of achievement test scores as outputs; while early papers used indirect information (for instance, success rates, or grades assigned by teachers, etc.) the more recent studies use quantitative indicators deriving from test scores. When discussing the inputs, all the studies try to include some measures of schools' resources (when available, the expenditure per student, eventually articulated in subcategories; otherwise, students:teachers ratio, facilities, etc.). Instead, the inclusion of students' socioeconomic status (SES) among inputs is more discussed. While the (huge) role of SES in explaining the students and schools' educational results is all but doubtful, the decision of considering it as a school input has been criticized. Some authors proposed to include measures of SES as non-discretionary inputs (see Ruggiero & Vitaliano, 1999); in the present paper, I include SES as a traditional input, allowing for the schools selecting (formally or not) their student population, and having them however the role of making the most with this input. The existing literature also evidenced the opportunity of finding indicators that can explain efficiency's differentials between schools; it also suggests some of these indicators, such as the quality of labour force (especially teaching staff), the degree of accountability, competition or autonomy, the private or public governance, etc. In the choice of variables to correlate with efficiency scores, however, I refer not only to the efficiency literature, but also to the one that investigate the so-called school effectiveness (Teddlie & Reynolds, 2000); the empirical exercises in this field are intended to explore the determinants

of students' results, taking advantage of different school climates, cultures, processes, educational and pedagogical approaches used in different settings[2].

However, among the various existing papers, then, there are few contributions that, more than others, have influenced my work. A priority reference is to Sutherland *et al.* (2009). To the best of my knowledge, it is the only study that undertook a comparison of schools' performances and efficiency across different countries; it does so by using PISA 2003 data. The empirical analysis considers schools in 30 countries/economies; the main findings point at demonstrating that indicators for higher quality of teaching resources appear to be correlated with higher schools' efficiency. However, despite the use of a school-level dataset, the main focus of the paper is on the macro-characteristics of countries' educational systems. Among the features that negatively affect country-level average schools' efficiency, there are small school size, streaming (that is tracking in the sense of Brunello & Checchi, 2007) and residence-based selection. Therefore, the attention of my paper is much more on school-level factors; in other words, I am interested to investigate more profoundly the between-schools differences in efficiency, complementing this analysis with an estimation of cross-country variation – but leaving the specific study of its determinants for future research.

Some contributions used PISA data for analysing the efficiency of schools in a specific country. Among them, Mancebon *et al.* (2012) focuses on Spanish school's results in the PISA 2006 edition. The authors propose an innovative approach, using a Hierarchical Linear Model (HLM) to choose the variables to be included in the model as inputs and outputs. In the subsequent analysis for estimating efficiency (through a DEA approach), the authors compare the efficiency of public and private schools – net of the role of students' socioeconomic background, for which they control for by including the indicator ESCS (Economic, Social and Cultural Status) among inputs. In so doing, their empirical analysis disentangles the part of efficiency related to each school's activity and that attributable to being in a specific subsector (public or private). Such decomposition takes inspiration by the work of Portela & Thanassoulis (2001); their original model, indeed, allows considering separately three components of efficiency, labelled as pupil, school and school-type efficiencies, respectively. In this paper, I use this intuition to decomposing the frontier of efficient schools in two parts, one that constitutes each school's benchmark within country, and one representing the "international" efficiency frontier.

Kim *et al.* (2006) for the first time explicitly indicated the potential trade-offs between efficiency and equity. Their paper deals with the estimation of the efficiency of primary schools in NYC, and poses the research question about whether the most efficient schools are those able to better serve the disadvantaged student populations. The empirical analysis does not reach any conclusive evidence, and it does not provide a list of characteristics that a school should possess to obtain both high efficiency score and being considered as "equitable" in helping disadvantaged students. Building on this idea, the present paper undertakes a specific investigation of the statistical relationships between efficiency scores and some indicators that are considered as proxies for equity or equality, such as (i) the variability of test scores within a school and (ii) the proportion of students who get scores

[2] In addition, the framework proposed by OECD's project "School Resources Review" (SRR) was helpful; it is however limited to focusing on the amount and use of resources, while here my analysis extends to school processes, governance and students' characteristics beyond their background.

below a given (very low) proficiency level. The empirical exercise conducted by Alexander *et al.* (2010) is the most similar to the one presented here, and is based on the double-bootstrap procedure proposed by Simar & Wilson (2007). The authors study the efficiency of New Zealand secondary schools, and find that the teachers' quality – more precisely, their qualification and experience – have a strong, positive impact on schools' efficiency.
Summarizing: the existing literature offers many suggestions about the best way to conduct efficiency analyses with school-level data, and reached a pretty stable consensus about the inputs (human and capital resources, students' background) and outputs (test scores) to be used for this purpose. The development of the methodology for estimating efficiency scores has developed rapidly, and is statistically more robust today. At the same time, the state-of-the-art if characterized by a lack of studies that compare schools' efficiency across countries; the present paper starts filling this gap.

3. Methodology and data

3.1. Bootstrap Data Envelopment Analysis and second-stage regression

In this paper, I use a well-known methodology, called Data Envelopment Analysis (DEA), for calculating the efficiency scores of the schools operating in Germany, France, Italy, Spain and UK. DEA has been extensively used in the literature about school efficiency (Worthington, 2001), and this methodology is well accepted by scholars in the field (the interested reader can find excellent manuals for details about the basics, as well as for recent developments (Cooper *et al.*, 2007 and Thanassoulis *et al.*, 2008).
In this brief section, I only report the simple version of the mathematical formulation for the DEA model estimated here. When considering the generic j-th school, which is included in the sample of n schools (such as $j = 1, \dots, n$), it has a input-output levels (x_0, y_o) – the model allows to consider s outputs r $(r = 1, \cdots, s)$ and m inputs i $(i = 1, \cdots, m)$. The technical efficiency for the j-th school can be mathematically expressed as the ratio of its outputs over the inputs, such as:

$$Max \left\{ h_0 = \frac{\sum_{r=1}^{s} u_r y_{r0} + w}{\sum_{i=i}^{m} v_i x_{i0}} \middle| \begin{array}{c} \frac{\sum_{r=1}^{s} u_r y_{rj} + w}{\sum_{i=1}^{m} v_i x_{ij}} \leq 1, \\ j = 1, \dots, n \quad u_r, v_i \geq 0 \quad w \; is \; free \end{array} \right\} \tag{1}$$

where u_r and v_i are the weights assigned by each DMU to the various inputs and outputs, with the aim of maximizing h_0. By construction, the efficiency is maximized, because the efficiency score of each school j_0 is computed as $1/h_0$.
The computation of h_0 is practically done, through DEA, by solving the dual problem specified in (1) by means of linear programming:

$$Max \left\{ h_0 = \sum_{r=1}^{s} u_r y_{r0} + w \middle| \begin{array}{c} -\sum_{i=1}^{m} v_i x_{ij} + \sum_{r=1}^{s} u_r y_{rj} + w \leq 0 \quad j = 1, \dots, n \\ \sum_{i=1}^{m} v_i x_{i0} \quad u_r, v_i \geq 0 \quad w \; is \; free \end{array} \right\} \tag{2}$$

As explained later in the next section, I opted for including three inputs and two outputs. The specific DEA model used for the empirical part of the paper is output-oriented (i.e. it is assumed that the schools use their human and financial resources as given, and try to

maximize students' achievement using them). Also, the model assumes variable returns to scale (VRS), in so doing making the hypothesis that there can be positive or negative scale effects due to the scale of operations, and allowing each school to benchmark its efficiency against a set of other schools with similar levels of inputs. Some authors criticized the properties of efficiency scores as derived through DEA, because of its deterministic nature; then, they suggested a method – based on bootstrapping techniques – for estimating bias-corrected efficiency scores, as well as confidence intervals around them (Simar & Wilson, 2000). In this paper, I calculated the efficiency scores in this more robust way, by using the package "Benchmarking" in R (Bogetoft & Otto, 2010).

After having estimated the efficiency scores for each school, in a second stage I regress them against a set of potential factors that can influence them. The aim of this multi-stage procedure is not to incorporate non-discretionary inputs into efficiency estimates (as suggested by Cordero-Ferrera *et al.*, 2008). Instead, I want to explore which are those factors – which can be both controllable or not by the school – that are statistically correlated with higher/lower levels of efficiency, and thus can explain variability between schools[3].

I used two types of regressions for this purpose. The first is a tobit regression, which considers the censored nature of efficiency scores (where the maximum=1, although in the case of bootstrap DEA scores none unit has really a value=1). The use of Tobit for the second stage is widespread in the literature, and has been used in similar contexts by Kirjavainen & Loikkanen (1998), Ruggiero & Vitaliano (1999), Bradley *et al.*, (2001), Borge & Naper (2006), and recently by Burney *et al.* (2013) and Chakraborty & Blackburn (2013), among others. The tobit regression here follows the mathematical formulation:

$$y_i \begin{cases} y_i^* \; if \; y_i^* < y_u \\ y_u \; if \; y_i^* > y_u \end{cases} \tag{3}$$

where y_i is the observed efficiency score, y_i^* is the latent (unobservable) variable that measures correctly the efficiency of the school i-th; it is assumed that $y_i = y_i^*$ if it is below a limit y_u that censors the distribution of efficiency scores (in traditional two-stage DEA studies $y_u = 1$, while here – where bootstrap DEA scores are never equal to 1, $y_u = y_{max}$. After having specified the output y_i in this way, the simple following equation can be estimated *via* the maximum likelihood estimator proposed by Tobin (1958) (see Wooldridge, 2012 for details):

$$y_i = \beta_0 + \beta_i \bar{X}_i + u_i, u_i \sim N(0, \sigma^2) \tag{4}$$

where $\bar{X}_i$ is a vector of schools characteristics that are (potentially) statistically correlated with the efficiency scores.

[3] The adoption of a two-stage procedure raised concerns as well, based on the fact that efficiency scores are serially correlated; for solving this problem Simar & Wilson (2007) propose a method based on a double-bootstrap algorithm, that allows obtaining meaningful estimates of coefficients in second stage regressions. In this paper, I use this method as a sensitivity test for the naïve estimation based on Tobit and quantile regressions.

As I am interested to check whether those factors associated with the schools' efficiency scores are the same along the entire distribution of the scores themselves, I also estimated the equation (4) through a quantile regression approach (Koenker & Hallock, 2001). The idea of such regression type is to estimate the statistical correlations between the dependent variable y_i (in this case, the efficiency score) and the set of covariates $\bar{X}_i$ (here, the school level factors associated with efficiency) at different points of the y distribution; mathematically:

$$min_\beta \sum \theta |y_i - \beta \bar{X}_i| + \sum (1-\theta)|y_i - \beta \bar{X}_i| \quad (5)$$

where y_i is the efficiency score of the i-th school, $\bar{X}_i$ is the vector of the school factors that are potentially statistically associated with efficiency, and θ is the quantile of the distribution at which estimating the relationship – in this paper, I consider the 25th, 50th (median) and 75th percentile of the distribution.

The list of variables, which have been used for populating $\bar{X}_i$ in both second-stage regressions (4)-(5), is reported and discussed in the next sections.

3.2. Data: building the school-level variables and the selection of inputs and outputs

All the school-level variables come from the OECD PISA2012 dataset; some of them are directly imputed at school level as they are from the school principal's questionnaire, while others are school averages of student-level information[4].

As the aim of the paper is providing an analysis of schools' performances, the first empirical effort consisted in selecting only schools that are institutionally comparable each other. For this purpose, two complimentary steps have been conducted. First, the modal grade attended by students in each country has been observed, by looking at the distribution of the variable STQ0101 (grade attended) in the PISA2012 dataset. Second, the information provided by OECD's Education GPS helped in describing the structure of each country's educational system. The aim of these two steps is to illustrate how the schools that are subsequently compared in the empirical analyses are somehow different because (i) they serve students at different stages of their scholastic pattern, and (ii) students who attend these schools were educated in those schools for different lengths of time.

The Annex 1 reports a table where the distribution of the students across grades in the original dataset is reported. The case for selecting the modal grade is pretty clear in Spain (model grade: 10), Italy (10) and UK (11), and the fifteen years-old students who are in lower grades are likely to be repeaters, while the (relatively few) ones who are in higher grades are those who started the school a year before the normal age. In Spain, all these students are at ISCED-2 level (enrolled in the last year of this level), while in Italy and UK they are attending ISCED-3 schools (from 1 and 5 years, respectively). The situation is a bit more complicated for France, where there is a significant proportion of students who were not passed to the first year of their new school (*Lycée*); to maintain the comparability between schools that are classified as upper-secondary (ISCED 3), I included only students at grade 10. The most articulated case is that of Germany. First of all, most part of fifteen years-old students are at ISCED-2 level (albeit at the last year of it), and can be enrolled at grade 9 in

[4] When aggregating the information at school level, students' final weights (W_FSTUDWT) have been used.

the tracks that bring directly to ISCED-3 or to the 10th grade of different tracks. The best choice to maintain a number of schools high enough to be compared with other countries was to include schools with students of both grades, and I control for this heterogeneity by including in the second-stage regression a dummy for the school type – I built a variable, called *progn_germany1*, that includes the most academically oriented tracks in the secondary education. Also, for considering the differences in tracks and grades, I added two further variables in the second stage: (i) *orgen*, that is a dummy that indicates if the school has an explicit academic/general orientation (for details about how it was built, see the codebooks provided in PISA), and (ii) *isced2*, a dummy indicating if the school is at that grade (compared with isced3)[5].

When considering the choice of variables for inputs and outputs, I ended up with three inputs and two outputs. The (inverse of) students/teachers ratio (*StRatio*) measures the intensity of human resources used by the school for educational purposes. The number of computers per student (*Computer_n*) is a proxy for the infrastructural dotation of schools. The index for the Economic, Social and Cultural Status of students (*ESCS*) is the proxy for the school's average socioeconomic composition of students[6]. For outputs, I chose the average test score in mathematics and reading (*pv1math* and *pv1read*)[7] for allowing differences across subjects. Based on these variables, a second selection of schools has been done, by excluding those for which some inputs were not available (in particular, *StRatio* and, for only two cases, average *ESCS*). After this selection, the final dataset comprises 194 German, 841 Spanish, 193 French, 447 British and 1,044 Italian schools (total: 2,719); in these schools, are enrolled around 90% of the students who have the characteristics described above to be included in the sample for the empirical analysis (details about these numbers are in the Annex 1).

The descriptive statistics about the inputs and outputs for the selected schools are reported in the Table 1 The school-average *ESCS* (mean value in the sample: -0.05) is highest in UK, followed by Germany and France, while Spanish and Italian schools are those in which the average sociocultural composition of students is more disadvantaged. The (inverse) of students:teachers ratio (*StRatio*) illustrates that Italy is the country where schools have more teachers per student (0.12), while UK is at the opposite extreme (0.07) – the mean in the sample being 0.1. Turning to the last input measure (*Computer_n*), on average the schools selected have 0.68 computers for each student, with this distribution ranging between 0.98 (UK) and 0.54 (Italy). The picture that emerges from the analysis of outputs *(pv1math* and *pv1read*) mirrors more or less the one showed in the Figure 1 (net of missing students from missing data schools). The variability of test scores across schools is lower in UK and Spain, and the highest heterogeneity is found in Germany, Italy and France – notably, three

[5] Also in the Annex 1, there is a table that illustrates how many years are these students, considered in the empirical analyses, were enrolled in the school when tested in PISA 2012. The schools considered here are mainly ISCED-3 in Italy, France and UK, while they are mainly ISCED-2 in Spain and Germany. On average, students were enrolled in these schools for 3.5 years (ranging from 1 year in France to 6 years in Germany) out of 5 years (on average) of the legal duration of the corresponding school level.

[6] OECD calculates ESCS through merging information about parents' education and occupation, as well as their home possession. Conventionally, it is set to have mean=0 and s.d.=1 on an international scale.

[7] As baseline results, I use only one (the first) out of the five plausible values. Later on in the paper, I check the robustness with other values – the results are unchanged.

educational systems characterized by practising tracking, i.e. by streaming students across different types of schools.

[Table 1] around here

Another aspect that is worth to be mentioned here is the degree of correlation between inputs and outputs, as well as correlations among inputs themselves and output themselves. On one side, a well-known critic to the "efficiency analysis" approach resides in the fact that quantitative measures of inputs are rarely related with measures of output (see, for instance, the arguments in Costrell *et al.*, 2008). On the other side, an excessive degree of correlation among inputs themselves can lead to problems of multicollinearity when estimating the efficiency scores (i.e. if different inputs captures collinear resources' dimensions). The Annex 2 contains a table illustrating the pairwise correlation across inputs (I) and outputs (O), for each country. The only I which is systematically correlated with O is the students' average *ESCS*, and this evidence corroborates the choice of include this variable among inputs. The other indexes of inputs are often uncorrelated with the outputs; sometimes they are even negatively correlated – i.e. poorer schools obtaining better results on test scores. I decided to maintain these indicators among inputs because they are relevant on a policy and managerial ground: even if they do not "explain" statistically the outputs, schools' managers are deemed as responsible for the use of the educational resources (teachers, facilities, etc.), so the policy-maker expects that better-equipped schools should produce better educational results. Lastly, the outputs (*pv1math* and *pv1read*) are always well statistically correlated, and this indicates the two measures' concordance in signalling the better/worse performing schools. Nevertheless, the non perfect coincidence opens the space for interesting analyses of the performance in the two subjects' space of output production – in other words, some schools can well have obtained relatively high results in mathematics, and relatively low in reading, and *vice versa*.

3.3. The variables used for second-stage regression

The selection of variables to be included in the second-stage regression has been driven by both theoretical and practical considerations. On the theoretical side, the literature and the conceptual frameworks about school effectiveness suggest that the factors, which can have an effect on efficiency scores, can be somehow grouped in three main categories:

- School's structural characteristics (i.e. private/public ownership, class and school size, diversity between classes in achievement and students' socioeconomic characteristics, etc.) and resources;
- Students' factors, such as the proportion of female or immigrant students, their socioeconomic background, attitudes and aspirations, etc.;
- School's practices and processes, such as the degree of competition, accountability procedures, school climate, selection and/or grouping, etc.

Two methodological remarks are important at this point. First, some variables that can be classified under one or more of those groups, are actually inserted in the model for computing efficiency scores as inputs; as a consequence, the variables selected from the categories listed

above must include information beyond the one captured by *ESCS*, *StRatio* and *Computer_n*. The second note is that I am using this information at school level, to see whether variation in these indicators of accountability, competition, governance, etc. are able to explain the variation in efficiency scores (obviously, country dummies are included to take country fixed-effects into account). In this sense, my approach is substantially different from the one adopted by those studies that aim at demonstrating the role of autonomy, accountability, competition etc. in shaping the overall educational system's performance (see Wößmann, 2003; Hanushek & Wößmann, 2010). Later on in this paper, I estimate how the efficiency scores of schools vary between different countries because of structural differences across them, and the role of "institutional" characteristics are likely to act at this level – but I do not further investigate also this aspect here.

The list of variables for the second stage is reported in the Table 2, which also contains a brief description, the mean value (or % for dichotomous variables) and the number of observations. Some specific explanations are useful to adequately interpret the table. In addition to *orgen* and *isced2*, mentioned above, the dummy *progn_Germany1* aims at considering the fact that the variable *orgen* does not discriminate enough between school types in Germany (and the variable prong_germany1 identifies only those schools that are most academically oriented) – and it is likely that the students attending the first track have systematically different characteristics than their counterparts in other tracks. The variable *ESCSsd* measures how different is the ESCS of students attending a specific school, that is to say if the school population is homogenous (low standard deviation) or heterogeneous. Two indicators are intended to capture the relationship of efficiency with equity or equality: (i) the standard deviation of scores within school (*pv1mathsd*; the higher the value the lower in equality, as it indicates wide dispersion of students scores around the mean), and (ii) the proportion of students who have a score below the proficiency level 2 (*pv1_belowprof2*), signalling that the school is not able to raise the guarantee to a certain proportion of students to obtain a minimum level of proficiency[8]. This way, equity is characterized in its sense of "inclusion", that is "(…) ensuring that all students reach at least a basic minimum level of skills. Equitable educational systems are fair and inclusive (…)" (Schleicher, 2014; p. 19).

[Table 2] around here

4. Results

4.1. Estimates based on the international frontier

The results of the efficiency analyses are presented as summary statistics by country, in the Table 3; efficiency scores, albeit derived in an output-oriented setting (where 1 is the maximum efficiency and values >1 represent inefficiency), were converted as a reciprocal to present efficiency scores that range in the [0;1] interval, where 1 is the maximum efficiency and (1 – efficiency score) is the degree of inefficiency. At a first glance, it can be noticed that, on average, German, Spanish and French schools have very similar average scores, while the British are the least efficient ones. If considering the standard deviation of

[8] The choice of proficiency level 2 as a threshold is justified under the assumption that "Level 2 is considered the baseline level of mathematical proficiency that is required to participate fully in modern society" (OECD 2014; p. 68).

efficiency scores as a measure of internal homogeneity, the Spanish and British educational system are those where the differences between schools' efficiency scores is lower, while Italy, France and Germany show a higher degree of internal heterogeneity.
An interesting way to analyse the characteristics of schools' efficiency in a cross-country comparison is to observe how many schools are deemed as statistically "more efficient than the average"; to be defined as such, the confidence interval around the bias-corrected (bootstrap) efficiency score must not include the across-country average efficiency score (in the specific case, 1.279, computed with the output-oriented measure, where 1 is the maximum efficiency). Conversely, a school, which score is statistically lower than 1.279 (its confidence interval does not cross the critical value), is considered inefficient. The results of this analysis is reported in the Table 4 and reveals that almost 50% of British schools are statistically more inefficient than the average, while the same proportion is lower than 30% in Spain; on the other side, more than 60% of French schools are more efficient than the average. The Spanish case is also characterized by a high proportion of efficient schools and a low proportion of inefficient ones (58.9% and 25.6%, respectively), while the Italian educational system presents an "equilibrated" situation, with around 41% of inefficient and 47% efficient schools.

[Table 3 and Table 4] around here

A graphical illustration (Figure 2) allows considerations that are beyond the comparison of means, and that take instead into account the whole distribution of (bootstrap, bias-corrected) efficiency scores (and their confidence intervals). The clear picture that emerges is that the distribution of Spanish schools is narrow, and for the most part well below the line that indicates the overall average efficiency score. On the contrary, the distribution of schools' efficiency scores in Germany, France, UK and Italy is much wider, with the presence of very inefficient and very efficient organisations. The case of UK is particularly interesting: albeit the most part of the distribution is above the red line (confirming the presence of a high proportion of schools "more inefficient than the average"), there are some schools with an "excellent" efficiency score (that is, an efficiency score >95th percentile of the international distribution) – even if they are only 1.5% of the total, while they are more than 11% in France and Germany. Also worth of mention is the case of Italy, where a number of very inefficient schools populates the lower tail of distribution.

[Figure 2] around here

It is useful to recall how the schools' efficiency scores derived here are conceptually different from school's average, raw performance. In the Figure 3, which is articulated by country, it can be seen that in all of them there are schools that, despite showing very different levels of performance obtain the same efficiency score, because of taking inputs availability into account. Analogously, many schools with very similar performance levels get different efficiency scores (for instance among schools with an average math score between 450 and 550, there is wide dispersion in efficiency scores, from 0.5 to 0.9). The patterns of correlation

between efficiency and performance are different across countries, thus. In particular, it appears clear how efficient schools tend to be more frequently those high performing ones in Germany, France and UK; while in Italy and especially in Spain, the performance level is rarely predictive of good efficiency scores. Another perspective through which look at the differences between efficiency and performance is to analyse the descriptive statistics of very efficient and very inefficient schools (those which efficiency score is >95th and <5th percentile, respectively). Such information is contained in the Table 5. Statistics reveal, for example, that in Spain the students' background is quite similar in the two groups of schools (average *ESCS*=-0.5/-0.6), but the performance is much higher among most efficient schools – in this sense, these schools are not the most efficient because of their high performance, but because their high performance obtained despite the (low) students' socioeconomic background. Again, in Germany the most inefficient schools are not only those with low performance levels, but those that have a high availability of facilities (*Computer_n* is 2.24 on average) but cannot transform them into achievement. The Italian schools in the group of best efficient are so despite their relatively high teachers:students ratio, because of the relatively low students' socioeconomic status; and so on and so forth. Overall, across countries*, StRatio* and *Computer_n* are not strikingly different between highly efficient and inefficient schools; instead, the latter group has (on average) a more disadvantaged student population to serve, and obtain an average score in mathematics and reading which is 190 and 200 points below the schools in the former (such huge difference is due to the fact that I am describing the very most efficient and inefficient units).

[Figure 3 and Table 5] around here

A factor that can play a relevant role is shaping the distribution of efficiency scores is the presence of different school types (i.e. the operating "tracking" system). With the aim of taking such differences into account, I will estimate later a model when including the two variables described in the section §3: *isced2*, and *orgen*. This model (estimated through a proper econometric model, see next subsections) will illustrate the statistical differences of the efficiency scores between tracks (grades, types).

A final relevant question is whether the scale of operations is affecting output levels, and efficiency. With the aim of providing some first impressions about this point, I calculated efficiency scores of each school under the different assumptions of variable returns to scale (VRS, as described above) and of constant returns to scale (CRS)[9]. This way, "scale efficiency" (Eff_{SCALE}) can be calculated as the ratio $Eff_{SCALE} = Eff_{CRS}/Eff_{VRS}$, and this can be interpreted as "(…) a measure of the distance between the CRS and VRS boundaries at the scale size of a unit" (Thanassoulis *et al.*, 2008; p. 290). Clearly, here the "scale of operations" (or scale size) does not refer to the school's size (as it is not included among inputs) but to its *ESCS*, *Computer_n* and *StRatio*. The results of these calculations are presented in the Table 6, and point at highlighting a high degree of scale inefficiency. These results should suggest that the average values of the three inputs must be reconsidered (and

[9] These calculations are made with the package "Benchmarking" in R; they are based on VRS and CRS traditional efficiency scores (not bootstrap ones; the correlation with them is >0.98).

rescaled) for benefiting from the unexhausted scale economies. For example, a possible action could be on the schools' socioeconomic composition: the Figure 4 seems indicating that there is a more than proportional return of (average) achievement to the (average) students' indicator of socioeconomic background (*ESCS)*.
When looking at another side of this story, that is the eventual relationship between schools size and performance/efficiency, a graphical analysis (see Figure 5) indicates that there is not detectable correlation neither between school size and test score in mathematics (*pv1math*) nor between size itself and (VRS) efficiency scores. I will check more closely this correlation (size/efficiency) later in the second-stage regression (section §4.3).

[Table 6, Figure 4, and Figure 5] around here

4.2. Estimates based on country-specific frontier, and disentanglement of efficiency scores' components

I re-estimated the efficiency scores of the schools by considering each country separately. In this case, the assumption is that each country's educational system is characterized by a different frontier of efficient schools; so, each school's efficiency score is calculated with respect to its country's frontier. Once obtained the new set of efficiency scores, I report the results in the same order as the previous subsection: descriptive statistics by country, number and proportion of efficient and inefficient schools, results by school type. In general terms, the previous results are mirrored in this set of results, with the Spanish and French schools obtaining the highest (relative to their own country's frontier) efficiency scores, and also presenting the lowest degree of internal heterogeneity (Table 7). On the other extreme, Italian schools are, on average when compared each against others, those with the lowest average efficiency score and the lowest within-country homogeneity. The estimated efficiency scores are on average higher in all the countries, and the overall mean is 0.86 instead of 0.80 – meaning that, on average, schools' efficiency is more similar within national boundaries than when considering the international frontier. In other words, these findings reveal that the distributions of schools' efficiency scores is more disperse (between and within-countries) when assuming a common international frontier, than when benchmarking the schools' efficiency rates within countries. The comparison of **Figure 6** (that represents the distribution of efficiency scores when considering the country-specific frontier) with the Figure 2 also confirms this point. Moreover, it shows that the confidence intervals estimated around bias-corrected (bootstrap) efficiency scores are much wider than before, especially in the tail of distribution with the most efficient schools. A consequence of this more imprecise estimation of bias-corrected efficiency scores is that the number and proportion of schools "more efficient than the average" is smaller than in the case of assuming international frontier (Table 8) conversely, almost 66% of schools (even 84% in the case of UK) are judged as less efficient than the average (it is important to recall here that the "average" is artificially calculated as a cross-country average, despite the country-specific nature of efficiency scores' estimates).

[Table 7, Table 8 and **Figure 6**] around here

With the aim of measuring the difference between the efficiency scores when estimated with respect to international or country-specific frontier, I calculate the variable $diff_{jc}$ that refers to each j-th school in the c-th country; it is mathematically computed as:

$$diff_{jc} = eff_{jc(C)} - eff_{jc(INT)} \quad (6)$$

where $\text{eff}_{\text{jc(C)}}$ is the efficiency score calculated through assuming the country-specific frontier as the reference, and $\text{eff}_{\text{jc(INT)}}$ the one when the schools are all pooled together for the empirical analysis. Given its mathematical formulation, the interpretation is that this index reflects how better a school is judged in terms of efficiency because it is compared with its own country's counterparts, instead that against the international benchmark. I report the descriptive statistics of this indicator in the Table 9 , and they show that, on average, the efficiency of the schools is overvalued of 5.8% when the within-country comparison is conducted. Such country-specific effect is the highest in UK, and the lowest in Italy – meaning that the schools in the latter result less efficient than schools in other countries not because their own technical inefficiency, but because there is a wider (structural) difference between the country's frontier and the international one. The distribution of diff_{jc}, by country, is graphically reported in the Figure 7, and shows how the tails are denser for the Spanish, German and British case. Lastly, in the Annex 3, I report the main statistics about the "overall" efficiency of the schools, defined (following Mancèbon *et al.*, 2012) as:

$$overall(eff_{jc}) = eff_{jc(c)} \times eff_{jc(INT)} \quad (7)$$

[Figure 7 and Table 9] around here

4.3. The factors associated with efficiency scores

I report the results of the tobit regression of the various indicators of schools' activities on efficiency (baseline, bootstrap) scores in the Table 10. The coefficients reported are the marginal effects, computed at the average of the dependent variable. The results are organised in three columns: in the first, the model contains only structural schools' characteristics, in the second model (school-average) students' features were added, and the third (complete) model includes all the remaining variables about the school's processes and practices. The method of running stepwise regressions, and showing the corresponding results accordingly, is not new in educational research (see, for instance Dronkers & Robert, 2008) and its use is opportune to show if (and how) the estimated coefficients change, when inserting new explanatory variables into the model.

The sign of *orgen* and *isced2* have the expected sign (positive and negative, respectively), even though the former is not statistically significant. The schools belonging to the more academic segment of Germany show higher efficiency scores than others. Higher dispersion of scores within-school (*pv1mathsd*) is slightly associated with higher efficiency, although the magnitude of this relationship is very small and overall is not statistically significant. Conversely, there are two indicators, which are strongly negatively correlated with efficiency: the diversity of students' socioeconomic background within school (as measured

by *ESCSsd*) and the proportion of students who report a test score below the proficiency level 2 (*pv1_belowprof2*). In this perspective, it seems that a school is more efficient when its student population is more homogenous (even if this association is contrasted by the positive sign of *immig_1*); therefore, most efficient schools seem those that are more able to reduce the proportion of students who perform below that proficiency level considered as "inadequate". All else equal, private schools have lower efficiency scores than their public counterparts. School size is negatively related with efficiency, but the coefficient of such relationship is small (even negligible, especially when considering that 1% change in size of a school is only 7 students, on average). Schools where the class size is typically small (<15 students) are less efficient than their counterparts where class size is bigger; even if the effect is small, it raises questions about the real "efficiency" behind the choice of keeping class smaller[10]. When concentrating on the average characteristics at school level, higher proportion of females is associated with higher efficiency scores. The same holds for the proportion of (first generation) immigrant students (*immig_1*), that seems to suggest a more diverse school population being associated with higher efficiency, partly contrary to what suggested by reading *ESCSsd* alone. The number of hours studied outside school (*out_hours*) is negatively related to efficiency; this is probably a proxy for less academic ability of students, so the sign of this relationship is meaningful indeed.

Among the processes conducted at school level, one has a positive statistical correlation with their efficiency: the role of principal in monitoring the activities and results of teachers (*monitor_p*). The variables about the school climate appear with the expected sign: the proportion of students reporting to have skipped school days for truancy (*st_truancy*) is negatively correlated with efficiency; instead, the principal's feeling that teachers' morale is high (*t_morale*) is a factor that contributes to efficiency. There are two indicators of school practices that emerge as not related to efficiency, even though they are designed explicitly for this: (i) using selectivity criteria for admitting students (*select_1*) and (ii) *grouping*. The sign is negatively related with efficiency, however; maybe suggesting that schools with a more selected group of students do not perform well enough to "compensate" their relative advantaged, and that grouping is not benefit test scores enough to make the school more efficient than others (it is also important to remember that previous literature suggested that grouping is likely to affect negatively both equity and absolute performance – see, for instance, Oakes, 2005). The index for the quality of educational resources (*sc_matedu*) turns out as statistically and positively associated with efficiency, suggesting that better equipped schools use their resources more efficiently. The absence of statistically significance for the proportion of certified teacher (*prop_cert*) is difficult to explain. One possible reason is that the variation between schools is too small to capture any meaningful correlation, at least in these five countries where the proportion of certified teachers is very high (>90%); if it is the case, then the estimated coefficient should be discarded. Another possible cause is that the schools with more certified teachers are also (more than proportionally) richer than others, and such gap is not compensated by better results. Indeed, it can also be the case that the indicator is not a good one to measure differential quality of teachers' characteristics between

[10] It must be kept in mind that no causation is implied here – i.e. small classes causing less efficiency. Also, the relationship deals with efficiency (that is to say, the ability of maximizing achievement, given the available resources) and not effectiveness (that is to say, if students in smaller classes have better achievement scores).

schools. The schools that report to compete with at least two other schools for students (*competition*) are more efficient, all else equal, albeit the relationship is statistically weak. A possible direction of this relationship is that these schools are more stimulated to use their resources in a more competitive way, in so doing increasing efficiency; the data at-hand do not allow to investigate if it is really the case, therefore, in similar exercises this is the interpretation in Bradley *et al.* (2001), Millimet & Collier (2008) and Agasisti (2013). Lastly, the indicator about how much hours are devoted to mathematics in the curricular programs (mminutes) indicates that this "teaching intensity" is beneficial to efficiency, although the magnitude of this effect seems very low.
After controlling for the all the variables now discussed, Spanish, Italian and French schools have efficiency scores that are statistically higher than those of schools in Germany; the contrary being true for UK.

[Table 10] around here

It is however interesting to explore if these correlations among efficiency and other school variables are heterogeneous. For this purpose, I show and comment the results of the quantile regression presented in the section §3.1, which estimates the statistical correlations at different point of the efficiency scores' distribution, namely 25th, 50th (median) and 75th percentile (see Table 11; only results for the complete model are showed)[11]. The overall results are not dramatically different from that reported above (especially at the median, when compared to the mean), but some noteworthy specificities emerge. For Germany, France and Italy, academic schools' higher efficiency scores are particularly evident in the right tail of distribution, among the highly efficient ones. The negative effect of the diversity of students' background (*ESCSsd*) is statistically relevant only at the 75th percentile of scores' distribution. When analysing the students' features, the proportion of immigrant students is positive only for less efficient schools; instead, the proportion of females is more positively associated with higher efficiency in all the distribution of efficiency scores. Then, when turning the attention to schools' practices and processes, the practice of monitoring teachers by principal (*monitor_p*) is more important for less efficient schools; while the teachers' morale (*t_morale*) is a factor that plays a major role in the more efficient schools. These findings should help policy makers and school managers to reflect about the various leverages that can be activated to improve the use of available resources, even when scarce. A variable that have its highest coefficient's magnitude among the most efficient schools is the index for the quality of educational resources (*scmatedu*), suggesting that having availability of resources is a key circumstance for being able of realising very efficient processes. Similar evidence exists for the indicator of *competition*; the sign for this variable is positive when considering the schools in the high tail of efficiency distribution – and the magnitude is substantial. Net of all these variables, at the 75th percentile, Italian schools perform statistically better than others, followed by French and German ones.

[11] Another potential source of heterogeneity is the different extent of correlations between school variables and efficiency scores, separately by country. Annex 4 is devoted to presenting the results about this point.

[Table 11] around here

4.4. More on efficiency and equity

An issue that deserves a particular attention is that of the relationship between efficiency and equity. In the previous section, it has been showed how efficiency scores are negatively correlated with an indicator of inequality, namely the proportion of students below proficiency level 2 (*pv1_belowprof2*) and positively with a measure of dispersion of scores around the mean, within school (*pv1mathsd*). A more profound analysis of these relationships can shed a clearer light on the potential trade-offs between efficiency (i.e. a school being able to make the most with its resources) and equity (i.e. minimizing the number of students left behind). I report the pairwise correlations between the estimated efficiency scores and the two indicators of "equity" defined in the previous sections in the Table 12. Two interesting elements must be discussed and commented here. First, on average there is not a real correlation between average efficiency and test scores' dispersion. The second aspect is that in all countries there is a strong, negative correlation between the proportion of students below proficiency level 2 in a school and its efficiency score. The magnitude of this correlation is similar between Italy, France and Germany (around 0.8), much weaker in Spain (0.5) and slightly lower in UK (0.7). This negative correlation is partly driven by the definition of efficiency itself; however, the different magnitude of the statistical correlation (well<1) suggests that there are efficient schools where there is still a high proportion of students below proficiency level 2, as well as inefficient schools where few students have this very low degree of proficiency. The difference between the arithmetical relationship between performance (test score) and equity and the analytical one efficiency/equity can be observed graphically. The left panel of the Figure 8 shows the plot of scores (horizontal axis) against the proportion of students below proficiency level 2; the right panel substitutes the x axis with the efficiency scores. As can be easily seen, albeit a clear negative correlation persists, in the panel where efficiency scores are considered, there is a considerable amount of dispersion, with a relevant number of schools where (i) efficiency scores is high (ii) the proportion of students with low scores is also high (see the right-up side of the panel), and another group of schools that, despite having few students below proficiency level 2, are not gaining high efficiency scores (see the left-down part of the panel).

[Table 12 and Figure 8] around here

Another way to investigate the relationship between efficiency and equity measures is to see if are there differences in the statistics about equity in the different groups of schools at various points of the efficiency distribution. This is what I report in the Table 13, where the following partitioning of the distribution has been decided: $<25^{th}$ percentile, between 25^{th} and the median, between the median and 75^{th}, $>75^{th}$. The statistics highlight that the main difference regards the proportion of students below proficiency level 2, and appears at the median of efficiency scores. Specifically, the schools with an efficiency score lower than the mean have more than double the proportion of these low-performing students; the schools at the lower end of efficiency distribution ($<25^{th}$ percentile) report even more than six times the proportion of these students than most efficient counterparts. No significant variation is found

for the measure of scores' dispersion within school (*pv1mathsd*); while this indicator ranges between 43 (5th percentile) and 89 (95th percentile), the differences between the four groups of schools by efficiency range between 68 and 73.
A further, last point to be underlined is that the characteristics of schools to be both efficient and equal (inclusive) is quite different across the five countries considered here. I calculated the number of schools that have (i) an efficiency score above the median *and* (ii) the indicator for the number of students below proficiency level 2 below the median (in both cases, the median is calculated with respect to the international distribution). The results (available on request from the author) indicate that first of all, these schools are not half of the sample, but only 39% – that is to say, for 11% of the schools one of this indicator is positive, and the other is not. Secondly, in France 54% of the schools satisfy both conditions, while about 25% in UK. Simply put, it is much easier to find a school that is more-than-average efficient *and* equal in France than in UK.

[Table 13] around here

Summarizing, the efficiency/equity relationship goes in the direction of complementarities: the higher the efficiency score, the lower the proportion of students performing below proficiency level 2 – which is a good proxy for equity. The efficiency scores are calculated by using students' average *ESCS* as an input; thus, this relationship with equity is net of the role exerted by differences in the socioeconomic status of student population across schools. In this perspective, finding methods and activities that can stimulate efficiency – i.e. the ability of making the most with the available resources – can also help improvements in the equality of education, raising the results for all students. In other words, my results show no evident trade-off between efficiency and equity. No causation is investigated here – it is unknown if efficiency is causing equity, or *viceversa*, or it is a simple correlations; nevertheless, the existence of such statistical relationship would provide interesting suggestions to policy-makers and school principals.

5. Robustness checks

In this section, I perform a battery of robustness checks to test how sensitive are the estimates of the schools' efficiency scores. In a policy perspective, indeed, it is definitively necessary to show that the results do not change dramatically when varying some assumptions behind the estimation of efficiency. Otherwise, the instability of the scores themselves would make them unusable in practical situations – for instance, when it must be decided which schools' factors should be promoted or discouraged to pursue higher levels of efficiency.

5.1. Stochastic frontier analysis

The most important sensitivity test conducted here is using a completely different methodological approach to estimate the schools' efficiency scores, and then see if the results obtained through it are consistent with those produced by DEA. While the idea of comparing

different methods is not new in the literature[12], the most part of existent articles limit themselves to the choice of a single method – whereas it would be extremely important to verify that the efficiency scores' estimates are robust across different techniques. Therefore, there is not *a priori* reason to be sure that different methods lead to similar results. Indeed, the various techniques are based on sets of completely different assumptions, and reconciling them is not always straightforward.

In this subsection, I use Stochastic Frontier Analysis (SFA) for deriving a new set of efficiency scores for the schools under analysis, and then I analyse the correlation with the DEA-based scores reported in the previous sections. While discussing the relative advantages and weaknesses of DEA and SFA is not pertinent here – a methodological proper discussion is in Fried *et al.* (2008) – it is important to recall here that SFA allows the estimation of a statistical error around the efficiency score (i.e. it is not a deterministic method). This advantage comes at the cost of (i) imposing a functional form to the relationship between output and inputs, and (ii) limiting the production function at the use of a single output.

With the aim of comparing the results with those obtained through DEA, I specify a translog production function with the same inputs used in that case. Mathematically, I estimated the following:

$$\ln(Math) = \beta_0 + \sum_i \beta_i lnX_i + \sum_i \beta_i lnX_i{}^2 + \frac{1}{2}\sum_i \sum_j \beta_i\beta_j \, lnX_i lnX_j + \delta_c + \varepsilon \qquad (8)$$

where X_i and X_j are two vectors of inputs that are defined as in the DEA model, $X_i \neq X_j$ and δ_c is a vector of country fixed effects. In the SFA context, the error term for the i-th school does not represent the traditional random noise; instead: $\varepsilon_i = v_i - u_i$ where v_i represents the idiosyncratic error variance, and u_i (strictly non-negative) is the one sided error that measures inefficiency in production. A functional form of u_i must be decided (in this paper: half-normal); moreover, technical efficiency estimates can be obtained via $E\{\exp(-u_i)|\epsilon_i\}$, where u are the estimates of minus the natural log of the technical efficiency via $E(u_i|\epsilon_i)$ (for further details about the various existent SF models, see Kumbhakar & Lovell, 2003).

The estimated coefficients for the production function are in the Table 14; the descriptive statistics of the set of efficiency scores obtained through this SFA are reported in the Table 15; lastly, in the Table 16, I report the correlation indexes between DEA and SFA estimates (both those referred to mathematics and reading scores). In general terms, the results are very consistent with the DEA ones; the estimated scores are higher (i.e. around 0.90 versus 0.80), but the correlation indexes (both absolute and rank-based) are statistically significant at 1% conventional level, and high in magnitude for all the countries (in all of them except Spain, the indexes are >0.8). Also looking at the graphical distribution of scores (Figure 9), there is no evidence of schools that result as outliers, i.e. being individuated as very efficient by DEA but not by SFA and *vice versa*.

[12] For instance, Johnes (2006) compared DEA and multilevel models for measuring efficiency of English universities; with special reference to the setting of the present paper, Sutherland *et al.* (2009) used both DEA and Stochastic Frontier Analysis when estimating the efficiency of schools in the OECD area using PISA 2003 data.

Summarizing, the picture that emerges from the application of a Stochastic Frontier Analysis to the dataset is completely coherent with that obtained through DEA. Whatever the empirical method, then, the set of estimated efficiency scores is stable: it can be concluded that the "international frontier" used as benchmark for measuring the efficiency of the schools in all the five countries do not change substantially if a different techniques for estimating it is employed.

[Figure 9, Table 14, Table 15 and Table 16] around here

5.2. Data Envelopment Analysis: different combinations of inputs, and outputs

In this section, I report the results obtained by using a different combination of inputs and by using alternative Plausible Values as outputs. In all these cases, the empirical approach is based on the same bootstrap DEA procedure described in the section §3.1. More specifically, the robustness checks tested are the following:

- Three models (RC1, RC2 and RC3) where I did not use all the three inputs, but instead three different combinations of two inputs:
    - RC1: *StRatio* and *ESCS*;
    - RC2: *StRatio* and *Computer_n*;
    - RC3: *ESCS* and *Computer_n*.

  The resulting DEA models, then, have two inputs (the ones listed above in the three different cases) and two outputs (mathematics and reading school-average scores, as measured by the Plausible Values 1, *pv1math* and *pv1read*).
- In the remaining four models (RC4, RC5, RC6 and RC7), on the other side, the DEA model uses the three inputs together as in the baseline model, and instead the four different sets of outputs – i.e. plausible values (PV) from *pv2math* and *pv2read* to *pv5math* and *pv5read*.

From each of the seven models, I calculated the bootstrap, bias-corrected efficiency scores, and then the correlation indexes were computed. The aim is to check two major facts: the first being that the choice of inputs – and their combination – do not lead to data-specific results, and the second being that (as claimed in the technical notes about PISA, see OECD (2012)) the use of each of the five plausible value as reference output do not alter the main results of the empirical analysis. Pearson and Spearman correlation types are illustrated in the Table 17, and confirm a high validity (i.e. robustness) of the baseline DEA efficiency scores commented and used in the previous sections. All the relevant coefficients are >.95, with the notable exception of model RC2, where the school-average ESCS is not included among inputs – nevertheless, the correlation is still positive, high in magnitude, and statistically significant. Taken together, these indexes claim for the importance of considering the average socioeconomic status of students as the key measure when judging efficiency, while the others (human and structural capital invested in the educational process are not equally essential for this purpose).

[Table 17] around here

5.3. Double bootstrap procedure for the second-stage analysis

The methodological stream of the literature debates the best way to conduct second-stage analyses of DEA scores; while for a long time the use of a Tobit regression has been widely accepted, some authors argue that in certain circumstances the use of simple OLS should be preferable (see McDonald, 2009). However, a more radical concern has been moved by Simar & Wilson (2007); their claim is that second-stage regressions are inconsistent when they do not consider a coherent (and statistically robust) data-generating process. In practice, the authors argument is that it is not only necessary to derive DEA scores through a bootstrap approach (to introduce statistical properties to these scores, as done in this paper); but also, the second-stage regression should also be conducted through a double-bootstrap algorithm for allowing proper inference. Technically, it involves the specification of a truncated regression, that uses maximum likelihood methods for deriving coefficients, which are subsequently bootstrapped (and that allows building confidence intervals for them). Hence, as a robustness check, I present here the results obtained by following the algorithm #2 suggested by Simar & Wilson (2007; p. 42-43). As shown in the Table 18, the estimated coefficients are very close (almost identical in magnitude) to those discussed in the previous section, with the expected sign in all cases. Therefore, the statistical significance of some of these coefficients is changed. Specifically, the results that maintain their statistical robustness after the check provided by the S&W estimations are: the negative efficiency differential for *isced2* schools, and the positive differential of efficiency for the schools in the most academic track in Germany (*prong_germany1*); the negative association with the proportion of students below the proficiency level 2 (*pv1_belowprof2*); the negative differential of efficiency for private schools (*private*); the positive relationships between efficiency and the proportion of females and immigrants in the school; the index for the quality of educational resources (*sc_matedu*); the indexes about climate (*st_truancy* and *t_morale*). Also, the check confirms that Spanish schools are on average more efficient, when comparing them net of the other factors influencing efficiency.

[Table 18] around here

6. Concluding remarks

This paper presented an efficiency analysis conducted on over 2,500 schools in five big European countries (Italy, Germany, France, Spain and UK). The interest of such international perspective resides in more than just rankings of most efficient schools or countries; allowing for the existence of an international frontier of efficient schools gives more spaces for policy making in European and national contexts. For instance, observing the combination of inputs and outputs in schools operating in different contexts can be extremely beneficial for promoting reforms, policies and/or managerial practices that are not already considered in each country's landscape. Moreover, understanding the factors associated with efficiency levels (something done here by means of second-stage regressions) could pave the way to more profound analyses about the drivers for a better use of (scarcer) resources. In the light of the acknowledged contribution of education/skills for economic growth (Hanushek & Wößmann, 2008; Sauer & Zagler, 2012), and considering the estimates about how much better educational results can foster European Union's growth (Hanushek & Wößmann,

2012), this paper contributes to the growing literature about the most effective (and efficient, in this case) educational policies for a better economic future.

One of the available information from efficiency analyses is the simulation of outputs' improvement of each school in the case of their productive (technical) efficiency. I calculated for each school the expansion of output that can be realised, given the available inputs and the estimated efficiency frontier. In the Figure 10, I reported the histograms that indicate how much output is missing for considering the schools as efficient, holding the inputs constant (figures consider the expansion all in the direction of mathematics test score, pv1math, and without considering improvements in reading)[13]. The results indicate the, on average, schools should show a much higher score in mathematics (if keeping the score in reading as constant), between 79 (France) and 100 (UK) points; the real figures are likely to be half than this number, if considering a radial expansion towards the efficiency frontier that consider an improvement in both mathematics and reading scores – say then, between 40 and 50 points. In PISA terms, this improvement is substantial, as it ranges from almost one to more than one years of schooling (conventionally, in PISA it is set around 41 points, again with heterogeneity between 23 in UK and 64 in Spain – see OECD, 2014; table A1.2). The Figure 10 indicates how this "efficient target" of outputs is very different across schools, however, depending on their relative inefficiency score: for some schools, the expected expansion of outputs is very modest, while for others is likely to be impossible to be gained, at least in the short term. These results from simulation must be kept with caution, however. First, there is no practical indication of how to get this improvement in outputs (unless leveraging on factors that are statistically associated with efficiency scores). Second, there is no judgment about the real possibility of reaching a simulated level of output – this is, indeed, simulation. In other words, the expansion of output here is useful as an indicative target, more than a prescriptive one; and it serves as a reference target for suggesting an order of magnitude for potential (efficiency) improvements. And third, no hypotheses about the eventual reduction of inputs, or their more productive use is done, so the resulting estimates are somehow upward biased.

[Figure 10] around here

The present paper innovates the previous literature, in that it departs from analysing schools' efficiency within a single country, and instead claims for the importance of cross-country comparisons. To my knowledge, only Sutherland *et al.* (2009) approached the topic in this way; but their contribution used PISA 2003 data (whilst this paper focuses on 2012 edition); it put together many countries, without exploring specific country-specific factors; it focused more on country-level variables that explain variation in efficiency, not on school-related indicators, which are instead at the *core* of this work.

I am aware that the empirical analysis presented here still suffers a number of limitations, partly due to the specific of PISA design and implementation. Based on this work, some

[13] The calculations are made through the package "Benchmarking" in R; a simple DEA model is used, instead of the bootstrap version (pairwise correlation >0.98).

suggestions can be formulated for making the dataset more suitable for efficiency studies in the future.

First, as effectively pointed out by Portela *et al.* (2013), the idea two inputs for computing a value-added measure (VAM) of schools' efficiency are (i) students' background and (ii) prior achievement. In PISA, the indicator ESCS does a good job as a proxy for the former; unfortunately, the latter is completely missing. This is a major weakness; while a relationship between the two dimensions can be assumes (so that, ESCS is a proxy also for prior achievement), having a direct measure of students' skills at the beginning of each educational step can be extremely useful for improving estimates of schools' efficiency. In future PISA editions, a specific attention should be posed to the possibility of deriving proxy indicators for the different starting points in achievement. This additional information would also make the investigation more coherent with the description of education as a "cumulative" process, as modelled by Cuhna *et al.* (2010).

Second, the present work is focused only on PISA 2012 edition, so it provides a picture of the relative efficiency of the schools in these countries only at a precise point of time. A first line of research for extending this paper would be to include a comparison with PISA 2003 edition (when mathematics was the main subject tested). While a real longitudinal comparison would be prevented altogether – students and schools in PISA 2003 and 2012 are obviously different – a static comparison between relative efficiency of schools in the selected countries could nevertheless provide interesting insights. However, this extension would be not without problems; for instance, the selection of variables to be included in the models would be affected by indicators availability and comparability across editions.

Third, the PISA dataset does not contain any direct measure of cost or expenditure at school level. This lack of information impedes any measure of efficiency beyond the technical one; in other words, it is not possible to extend efficiency measurements for including allocative and/or cost efficiency. When the information about costs is available, the kinds of analyses that can be realised allow understanding in much a wider way the factors that drive efficiency of schools themselves (Levin, 2976); some applications of the allocative efficiency concept to study of schools are in Grosskopf *et al.* (2001), Banker *et al.*, (2004), and Haelermans & Ruggiero (2013). I think that promoting the inclusion of cost/expenditure indicators (at school level) into PISA could be beneficial for all those who intend study the efficiency of schools in a comparative setting.

Fourth, the results presented here do not confirm the key findings obtained by Sutherland *et al.* (2009) about the positive impact on efficiency of teachers' quality. The availability of few data about teachers' characteristics, attitudes and activities; as well as the focus on five well-developed countries, can be at the basis of this discordance – which, however, should be investigated further in future research.

Lastly, as always in these empirical works, the production model assumes to capture the output through test scores on cognitive skills. The literature already shows that non-cognitive skills are as important as cognitive ones (Levin, 2012); and Levin (2013) calls for including indicators for measuring them on a large-scale educational assessment. When such additional information will be available, the quality and usefulness of efficiency studies will be substantially improved.

References

Afonso, A., & St Aubyn, M. (2005). Non-parametric approaches to education and health efficiency in OECD countries. Journal of Applied Economics, 8(2), 227-246.

Afonso, A., & St Aubyn, M. (2006). Cross-country efficiency of secondary education provision: A semi-parametric analysis with non-discretionary inputs. Economic Modelling, 23(3), 476-491.

Agasisti, T. (2013). The efficiency of Italian secondary schools and the potential role of competition: a data envelopment analysis using OECD-PISA2006 data. Education Economics, 21(5), 520-544.

Agasisti, T. (2014). The Efficiency of Public Spending on Education: an empirical comparison of EU countries. European Journal of Education, forthcoming.

Alexander, W. R. J., Haug, A. A., & Jaforullah, M. (2010). A two-stage double-bootstrap data envelopment analysis of efficiency differences of New Zealand secondary schools. Journal of Productivity Analysis, 34(2), 99-110.

Banker, R. D., Janakiraman, S., & Natarajan, R. (2004). Analysis of trends in technical and allocative efficiency: An application to Texas public school districts. European Journal of Operational Research, 154(2), 477-491.

Bogetoft, P., & Otto, L. (2010). Benchmarking with DEA, SFA, and R (Vol. 157). Springer.

Borge, L. E., & Naper, L. R. (2006). Efficiency potential and efficiency variation in Norwegian lower secondary schools. FinanzArchiv: Public Finance Analysis, 62(2), 221-249.

Bradley, S., Johnes, G., & Millington, J. (2001). The effect of competition on the efficiency of secondary schools in England. European Journal of Operational Research, 135(3), 545-568.

Brunello, G., & Checchi, D. (2007). Does school tracking affect equality of opportunity? New international evidence. Economic Policy, 22(52), 781-861.

Burney, N. A., Johnes, J., Al-Enezi, M., & Al-Musallam, M. (2013). The efficiency of public schools: the case of Kuwait. Education Economics, 21(4), 360-379.

Chakraborty, K., & Blackburn, V. C. (2013). Efficiency and Equity in Funding for Government Schools in Australia. Australian Economic Papers, 52(3-4), 127-142.

Clements, B. (2002). How efficient is education spending in Europe? European Review of Economics and Finance, 1(1), 3-26.

Cooper, W. W., Seiford, L. M., & Tone, K. (2007). Data envelopment analysis: A comprehensive text with models, applications, references and DEA-Solver Software. Second editions. Springer, ISBN, 387452818, 490.

Cordero-Ferrera, J. M., Pedraja-Chaparro, F., & Salinas-Jiménez, J. (2008). Measuring efficiency in education: an analysis of different approaches for incorporating non-discretionary inputs. Applied Economics, 40(10), 1323-1339.

Costrell, R. M., Hanushek, E., & Loeb, S. (2008). What do cost functions tell us about the cost of an adequate education? Peabody Journal of Education, 83, 198-223.

Cunha, F., Heckman, J. J., & Schennach, S. M. (2010). Estimating the technology of cognitive and noncognitive skill formation. Econometrica, 78(3), 883-931.

Dronkers, J., & Robert, P. (2008). Differences in Scholastic Achievement of Public, Private Government-Dependent, and Private Independent Schools A Cross-National Analysis. Educational Policy, 22(4), 541-577.

EU Commission. (2006). Communication from the Commission to the Council AND to the European Parliament: Efficiency and equity in European education and training systems COM/2006/0481 final.

Farrell, M. J. (1957). The measurement of productive efficiency. Journal of the Royal Statistical Society. Series A (General), 253-290.

Fried, H. O., Lovell, C. K., & Schmidt, S. S. (Eds.). (2008). The measurement of productive efficiency and productivity growth. Oxford University Press.

Grosskopf, S., Hayes, K. J., Taylor, L. L., & Weber, W. L. (2001). On the determinants of school district efficiency: Competition and monitoring. Journal of Urban Economics, 49(3), 453-478.

Haelermans, C., & Ruggiero, J. (2013). Estimating technical and allocative efficiency in the public sector: A nonparametric analysis of Dutch schools. European Journal of Operational Research, 227(1), 174-181.

Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. Journal of Economic Literature, 1141-1177.

Hanushek, E. A. (1989). The impact of differential expenditures on school performance. Educational Researcher, 18(4), 45-62.

Hanushek, E. A. (1997). Assessing the effects of school resources on student performance: An update. Educational Evaluation and Policy Analysis, 19(2), 141-164.

Hanushek, E. A. (2003). The Failure of Input‐based Schooling Policies. The Economic Journal, 113(485), F64-F98.

Hanushek, E. A., & Wößmann, L. (2008). The role of cognitive skills in economic development. Journal of Economic Literature, 607-668.

Hanushek, E. A., & Wößmann, L. (2010). The economics of international differences in educational achievement. NBER Working Paper No. w15949, National Bureau of Economic Research.

Hanushek, E. A., & Wößmann, L. (2012). The economic benefit of educational reform in the European Union. CESifo Economic Studies, 58(1), 73-109.

Kim, Y. D., Zabel, J. E., Stiefel, L., & Schwartz, A. E. (2006). School Efficiency and Student Subgroups: Is a Good School Good for Everyone?. Peabody Journal of Education, 81(4), 95-117.

Kirjavainen, T., & Loikkanent, H. A. (1998). Efficiency differences of Finnish senior secondary schools: an application of DEA and Tobit analysis. Economics of Education Review, 17(4), 377-394.

Koenker, R., & Hallock, K. (2001). Quantile regression: An introduction. Journal of Economic Perspectives, 15(4), 43-56.

Kumbhakar, S. C., & Lovell, C. K. (2003). Stochastic frontier analysis. Cambridge University Press.

Johnes, J. (2004). 16 Efficiency measurement. In Johnes G. & Johnes J. (eds.), International handbook on the economics of education, 613-742.

Johnes, J. (2006). Measuring efficiency: A comparison of multilevel modelling and data envelopment analysis in the context of higher education. Bulletin of Economic Research, 58(2), 75-104.

Levin, H.M. (1976). Concepts of economic efficiency and educational production, in Froomkin J.T., Jamison D.T. & Radner R. (eds.). Education as an industry, (149-198) National Bureau of Economic Research, Cambridge Massachusetts.

Levin, H. M. (2012). More than just test scores. Prospects, 42(3), 269-284.

Levin, H. M. (2013). The utility and need for incorporating non-cognitive skills into large-scale educational assessments. In The role of international large-scale assessments: Perspectives from technology, economy, and educational research (67-86). Springer, Netherlands.

Mancebón, M. J., Calero, J., Choi, Á., & Ximénez-de-Embún, D. P. (2012). The efficiency of public and publicly subsidized high schools in Spain: Evidence from PISA-2006. Journal of the Operational Research Society, 63(11), 1516-1533.

McDonald, J. (2009). Using least squares and tobit in second stage DEA efficiency analyses. European Journal of Operational Research, 197(2), 792-798.

Millimet, D. L., & Collier, T. (2008). Efficiency in public schools: Does competition matter?. Journal of Econometrics, 145(1), 134-157.

Oakes, J. (2005), Keeping track: how schools structure inequality, 2nd edition, New Haven, CT: Yale University Press.

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

OECD (2014). PISA2012 Results: what students know and can do – student performance in mathematics, reading and science (volume I, revised version, February 2014), OECD Publishing, Paris.

Portela, M.C.S., & Thanassoulis, E. (2001). Decomposing school and school-type efficiency. European Journal of Operational Research, 132(2), 357-373.

Portela, M. C., Camanho, A. S., & Keshvari, A. (2013). Assessing the evolution of school performance and value-added: trends over four years. Journal of Productivity Analysis, 39(1), 1-14.

Ruggiero, J., & Vitaliano, D. F. (1999). Assessing the efficiency of public schools using data envelopment analysis and frontier regression. Contemporary Economic Policy, 17(3), 321-331.

Sauer, P., & Zagler, M. (2012). Economic growth and the quantity and distribution of education: a survey. Journal of Economic Surveys, 26(5), 933-951.

Schleicher, A. (2014). Charting a way towards equity and excellence. In Equity, excellence and inclusiveness in education – policy lessons from around the world. OECD Publisher, Paris.

Simar, L., & Wilson, P. W. (2000). A general methodology for bootstrapping in non-parametric frontier models. Journal of Applied Statistics, 27(6), 779-802.

Simar, L., & Wilson, P. W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. Journal of Econometrics, 136(1), 31-64.

Sutherland, D., Price, R., & Gonand, F. (2009). Improving public spending efficiency in primary and secondary education. OECD Journal: Economic Studies, 2009(1), 1-30.

Teddlie, C., & Reynolds, D. (Eds.). (2000). The international handbook of school effectiveness research. Psychology Press.

Thanassoulis, E., Portela, M.C.S., & Descpic, O. (2008). Data envelopment analysis: the mathematical programming approach to efficiency analysis. In (Fried, H.O., Knox Lovell, C.A. & Schmidt, S.S. (eds.). The measurement of productive efficiency and productivity growth, Oxford University Press, New York.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. Econometrica: Journal of the Econometric Society, 26(1), 24-36.

Wößmann, L. (2003). Schooling resources, educational institutions and student performance: the international evidence. Oxford Bulletin of Economics and Statistics, 65(2), 117-170.

Wooldridge, J.M. (2013). Introductory econometrics – a modern approach, 5th Edition, Sage Publications.

Worthington, A. C. (2001). An empirical survey of frontier efficiency measurement techniques in education. Education Economics, 9(3), 245-268.

Table 1 Descriptive statistics, inputs and outputs, by country

		Inputs		Outputs	
Germany	ESCS	StRatio	Computer_n	pv1math	pv1read
N	194	194	194	194	194
mean	0.125	0.080	0.698	506.127	499.447
cv	4.354	0.869	2.146	0.155	0.153
sd	0.545	0.070	1.497	78.299	76.234
min	-1.382	0.020	0.000	303.658	273.928
max	1.511	0.851	18.000	633.988	627.027
Spain	ESCS	StRatio	Computer_n	pv1math	pv1read
N	841	841	841	841	841
mean	-0.166	0.103	0.723	488.871	488.364
cv	-3.258	0.635	0.869	0.096	0.097
sd	0.540	0.065	0.628	46.888	47.608
min	-2.360	0.007	0.000	98.232	262.940
max	1.423	0.900	8.000	607.231	623.460
France	ESCS	StRatio	Computer_n	pv1math	pv1read
N	193	193	193	193	193
mean	-0.071	0.090	0.565	491.619	502.184
cv	-6.313	0.287	0.969	0.153	0.169
sd	0.447	0.026	0.548	75.189	84.793
min	-1.402	0.047	0.000	283.658	258.048
max	0.969	0.223	5.000	683.382	717.762
UK	ESCS	StRatio	Computer_n	pv1math	pv1read
N	447	447	447	447	447
mean	0.233	0.070	0.978	487.539	496.491
cv	1.628	0.256	0.653	0.102	0.102
sd	0.379	0.018	0.639	49.968	50.786
min	-0.831	0.046	0.000	365.621	345.711
max	1.364	0.233	4.423	648.769	640.594
Italy	ESCS	StRatio	Computer_n	pv1math	pv1read
N	1044	1044	1044	1044	1044
mean	-0.125	0.128	0.540	477.247	477.940
cv	-4.411	0.716	0.891	0.152	0.168
sd	0.552	0.091	0.481	72.724	80.180
min	-2.580	0.014	0.000	253.630	221.319
max	1.578	1.582	4.500	652.770	636.546
Total	ESCS	StRatio	Computer_n	pv1math	pv1read
N	2719	2719	2719	2719	2719
mean	-0.057	0.104	0.682	485.615	487.469
cv	-9.382	0.706	1.019	0.131	0.139
sd	0.536	0.074	0.694	63.378	67.715
min	-2.580	0.007	0.000	98.232	221.319
max	1.578	1.582	18.000	683.382	717.762

Table 2 Variables for the second-stage regression(s) – list, description, mean values

Variable name	Description	Mean (or %)	N
School characteristics			
isced2	=1 if the school is ISCED 2	44.2%	2718
progn_Germany1	=1 if the school is of the 1st track in Germany (academic)	3.0%	2718
orgen	Orientation of the school: general	80.6%	2718
pv1mathsd	Standard devation of scores, within school (pv1math)	70.303	2710
ESCSsd	Standard deviation of ESCS index, within school	0.798	2709
pv1_belowprof2	Proportion of students below the proficiency level 2 (pv1math)	24.6%	2718
private	=1 if the school is private (no distintion if gvmt dependent)	19.7%	2718
size	Number of students	710.543	2718
clsize_small	=1 if the average class has <15 students	3.4%	2718
Students' characteristics			
immig_1	Proportion of immigrant students, first generation	7.1%	2718
female	Proportion of female students	48.9%	2718
out_hours	Hours devoted to studying with additional activities, per week	6.703	2700
School processes and practices			
sc_matedu	PISA index for the quality of educational material resources	0.103	2703
mminutes	Minutes of curricular actitivities, in mathematics	220.660	2692
prop_cert	Proportion of teachers with a certification to teach	91.4%	2631
competition	=1 if the school competes with 2 or more schools for students	52.6%	2718
st_truancy	Proportion of students who reported having skipped days of schools	29.4%	2718
t_morale	Principal reports that teaching morale is very posiitve	16.9%	2718
monitor_p	Teachers are monitored by school principal or senior staff	29.9%	2718
select_1	The school uses always at least one factor to select students	37.8%	2718
grouping	Ability grouping for some or all classes	40.8%	2718

Notes: more descriptive statistics are available, on request, from the author. Also a wider list of potential second-stage covariates is available.

Table 3 Efficiency scores, by country – international frontier

	mean	median	sd	min	max	N
Germany	0.801	0.811	0.090	0.541	0.952	194
Spain	0.804	0.807	0.054	0.554	0.949	841
France	0.804	0.826	0.093	0.446	0.957	193
UK	0.772	0.772	0.059	0.566	0.974	447
Italy	0.781	0.790	0.089	0.413	0.990	1,043
Total	0.790	0.795	0.076	0.413	0.990	2,718

Notes: the scores are obtained using the boostrap DEA procedure included in the package *Benchmarking* in R; 2,000 replicates have been used. Efficiency scores are calculated in a output-oriented setting, and then transformed as 1/eff_score for presenting scores in the range [0;1]. The model was unable to estimate the scores for two schools that were consequently dropped by these statistics.

Table 4 The number of "more efficient than the average" and "less efficient than the average" schools, by country – international frontier

	N	n_efficient	% efficient	n_inefficient	% inefficient
Germany	194	106	54.6%	66	34.0%
Spain	841	495	58.9%	215	25.6%
France	193	117	60.6%	63	32.6%
UK	447	151	33.8%	215	48.1%
Italy	1,043	498	47.7%	433	41.5%
Total	2,718	1,367	50.3%	992	36.5%

Notes: a school is considered "more efficient than the average" is the upper bound limit of the confidence interval (95%) is lower than the overall mean efficiency score; conversely, it is considered as "less efficient than the average" is the lower bound is higher than the overall mean efficiency score. Calculations are made using the output-oriented measures, which range is $[1; \infty]$ and 1 is the best efficiency score.

Table 5 The characteristics of very efficient and very inefficient schools: descriptive statistics – international frontier

Very efficient school (>95% percentile)	ESCS	StRatio	Computer_n	pv1math	pv1read
Germany (n=21; 11%)	0.47	0.08	0.44	600.17	582.10
Spain (n=24, 3%)	-0.54	0.11	0.69	511.10	530.83
France (n=22, 11%)	0.18	0.09	0.43	569.15	592.81
UK (n=7, 2%)	0.66	0.07	0.73	609.59	602.06
Italy (n=63, 6%)	0.12	0.12	0.44	574.10	569.83
Total (n=137)	0.09	0.10	0.50	568.08	570.22

Very efficient school (<5% percentile)	ESCS	StRatio	Computer_n	pv1math	pv1read
Germany (n=13, 7%)	-0.48	0.13	2.24	357.95	349.50
Spain (n=11, 1%)	-0.68	0.14	0.51	352.62	341.44
France (n=12, 6%)	-0.41	0.08	0.62	354.31	343.88
UK (n=9, 2%)	-0.10	0.07	1.11	388.45	378.86
Italy (=90, 9%)	-0.52	0.13	0.43	354.99	334.57
Total (n=135)	-0.49	0.12	0.68	357.25	340.35

Notes: the distribution over which it the percentiles are defined is the unique, international one; hence, the definition of "very efficient" or "very inefficient" school is an absolute one, and is not country-specific.

Table 6 Scale efficiency scores (descriptive statistics), by country – international frontier

	mean	median	sd	min	max	N
Germany	0.309	0.316	0.095	0.033	0.726	194
Spain	0.237	0.224	0.092	0.028	1.000	841
France	0.261	0.266	0.067	0.098	0.530	193
UK	0.287	0.286	0.066	0.110	0.466	447
Italy	0.216	0.204	0.089	0.017	1.000	1,044
Total	0.244	0.239	0.091	0.017	1.000	2,719

Notes. As described in the section §4.1, scale efficiency is defined as the ratio $Eff_{SCALE} = Eff_{CRS}/Eff_{VRS}$.

Table 7 Efficiency scores, by country – country-specific frontier

	mean	median	sd	min	max	N
Germany	0.877	0.899	0.080	0.594	0.985	192
Spain	0.876	0.882	0.058	0.571	0.989	840
France	0.860	0.874	0.080	0.503	0.987	192
UK	0.855	0.859	0.058	0.611	0.975	446
Italy	0.813	0.829	0.094	0.415	0.985	1,043
Total	0.847	0.860	0.082	0.415	0.989	2,713

Notes: the scores are obtained using the boostrap DEA procedure included in the package *Benchmarking* in R; 2,000 replicates have been used. Efficiency scores are calculated in a output-oriented setting, and then transformed as (1/efficiency score) for presenting scores in the range [0;1]. The model was unable to estimate the scores for six schools that were consequently dropped by these statistics.

Table 8 The number of "more efficient than the average" and "less efficient than the average" schools, by country – country-specific frontier

cnt	N	n_efficient	% efficient	n_inefficient	% inefficient
Germany	192	61	31.8%	101	52.6%
Spain	840	151	18.0%	561	66.8%
France	192	72	37.5%	95	49.5%
UK	446	41	9.2%	372	83.4%
Italy	1,043	232	22.2%	686	65.8%
Total	2,713	557	20.5%	1815	66.9%

Notes: a school is considered "more efficient than the average" is the upper bound limit of the confidence interval (95%) is lower than the overall mean efficiency score; conversely, it is considered as "less efficient than the average" is the lower bound is higher than the overall mean efficiency score.

Table 9 The difference between the estimated efficiency scores: country-specific vs international frontier

	mean	median	sd	min	max	N
Germany	0.076	0.074	0.037	-0.003	0.343	192
Spain	0.072	0.071	0.028	-0.010	0.213	840
France	0.057	0.048	0.046	-0.013	0.234	192
UK	0.083	0.083	0.033	-0.032	0.221	446
Italy	0.032	0.031	0.028	-0.033	0.199	1,042
Total	0.058	0.057	0.038	-0.033	0.343	2,712

Notes: the indicator is calculated as the difference between the efficiency score estimated with the country-specific frontier and that estimated under the international-frontier assumption. The model was unable to estimate the scores for ten schools that were consequently dropped by these statistics.

Table 10 The factors associated with efficiency scores: results from second-stage, Tobit regression

model	Model 1: only school characteristics	Model 2: students and schools' characteristics	Model 3: complete
isced2	**-0.03220*** **	**-0.03123*** **	**-0.03382*** **
	0.005	0.005	0.005
progn_germany1	**0.05864*** **	**0.05538*** **	**0.05828*** **
	0.005	0.005	0.005
orgen	**0.02476*** **	**0.02035*** **	**0.00871***
	0.004	0.004	0.004
pv1mathsd	-0.0001	0.00004	-0.00001
	0	0	0
pv1_belowprof2	**-0.21013*** **	**-0.20003*** **	**-0.19908*** **
	0.005	0.005	0.005
private	**-0.01245*** **	**-0.01494*** **	**-0.01719*** **
	0.003	0.003	0.003
size	**-0.00001** **	**-0.00001*** **	**-0.00001*** **
	0	0	0
clsize_small	**-0.01985*** **	**-0.02248*** **	**-0.01953*** **
	0.004	0.005	0.005
ESCSsd		**-0.02131*** **	**-0.02069*** **
		0.005	0.006
immig_1		0.01502	**0.03685*** **
		0.008	0.008
female		**0.05009*** **	**0.05453*** **
		0.005	0.005
out_hours		**-0.00418*** **	**-0.00617*** **
		0.001	0.001
sc_matedu			**0.00455*** **
			0.001
mminutes			**0.00028*** **
			0
prop_cert			0.00671
			0.005
competition			**0.00338[w]**
			0.002
st_truancy			**-0.03521*** **
			0.008
t_morale			**0.00988*** **
			0.003
monitor_p			**0.00662***
			0.003
select_1			-0.00437
			0.002
grouping			-0.00136
			0.002

model	Model 1: only school characteristics	Model 2: students and schools' characteristics	Model 3: complete
Spain	**0.04332*** **	**0.04504*** **	**0.05314*** **
	0.004	0.004	0.005
France	**0.01219*** **	**0.00951** **	**0.01318*** **
	0.003	0.003	0.004
UK	**-0.02282*** **	**-0.01998*** **	**-0.02936*** **
	0.006	0.006	0.006
Italy	-0.00244	0.00395	**0.01231***
	0.004	0.005	0.006
Constant	**0.84039*** **	**0.84898*** **	**0.80613*** **
	0.007	0.008	0.01
sigma	0.19908***	0.18903***	0.17961***
N	2710	2694	2589
loh-likelihood	2825.218	2926.669	2907.71

Notes: *,** and *** mean statistically significant of coefficients at standard 5%, 1%, and .1% respectively. The coefficients reported are the marginal effects as computed at the mean of the dependent variable.

Table 11 The factors associated with efficiency scores: heterogeneity – results from the second-stage, quantile regression

	Model C (complete)		
	25th	median	75th
isced2	**-0.01407****	**-0.01757*****	**-0.01227***
	0.005	0.004	0.005
progn_germany1	**0.05199*****	**0.05817*****	**0.05889*****
	0.006	0.005	0.007
orgen	0.00557	**0.01298****	**0.01552****
	0.005	0.005	0.006
pv1mathsd	**0.00058*****	**0.00032*****	0.00011
	0	0	0
pv1_belowprof2	**-0.23059*****	**-0.22297*****	**-0.20806*****
	0.006	0.006	0.007
private	**-0.01583*****	**-0.01874*****	**-0.01476*****
	0.003	0.003	0.004
size	-0.00001	**-0.00001*****	0
	0	0	0
clsize_small	**-0.02484*****	**-0.01566****	0.00445
	0.006	0.006	0.007
ESCSsd	-0.00003	-0.00655	**-0.03078*****
	0.007	0.006	0.008
immig_1	**0.03402****	**0.03048****	0.01921
	0.01	0.009	0.012
female	**0.05576*****	**0.05008*****	**0.05100*****
	0.007	0.006	0.007
out_hours	**-0.00651*****	**-0.00611*****	**-0.00560*****
	0.001	0.001	0.001
sc_matedu	**0.00292***	**0.00326****	**0.00605*****
	0.001	0.001	0.002
mminutes	**0.00024*****	**0.00016*****	**0.00023*****
	0	0	0
prop_cert	-0.00753	**0.01086***	**-0.01381***
	0.006	0.005	0.007
competition	**-0.00674***	**0.00929*****	**0.01072*****
	0.003	0.002	0.003
st_truancy	**-0.02688****	**-0.03824*****	**-0.03690*****
	0.009	0.008	0.01
t_morale	0.00298	0.00273	**0.02067*****
	0.003	0.003	0.004
monitor_p	**0.00733***	0.00019	**0.01533*****
	0.003	0.003	0.004
select_1	**-0.00852****	**-0.00664****	**-0.00996****
	0.003	0.003	0.003
grouping	-0.00198	0.00104	-0.00561
	0.003	0.002	0.003

	Model C (complete)		
	25th	median	75th
Germany	**0.01719***	**0.01293***	**0.02355****
	0.007	0.006	0.008
France	**0.03512*****	**0.03537*****	**0.04636*****
	0.007	0.006	0.008
Italy	**0.04288*****	**0.05059*****	**0.07403*****
	0.007	0.006	0.007
Constant	0.72838***	0.77414***	0.79924***
	0.014	0.013	0.016

Notes: *,** and *** mean statistically significant of coefficients at standard 5%, 1%, and .1% respectively. The function qreg in Stata© has been used for the estimations.

Table 12 Pairwise statistical correlations between indicators of efficiency and equity – international frontier

Germany	deab_effscore	pv1mathsd	pv1_belowprof2
deab_effscore	1		
pv1mathsd	-0.0234	1	
pv1_belowprof2	**-0.8533***	-0.0336	1
Spain	deab_aga	pv1mat~d	pv1_be~2
deab_aga	1		
pv1mathsd	-0.0787	1	
pv1_belowp~2	**-0.5120***	**0.1763***	1
France	deab_aga	pv1mat~d	pv1_be~2
deab_aga	1		
pv1mathsd	**0.2333***	1	
pv1_belowp~2	**-0.8383***	**-0.2534***	1
UK	deab_aga	pv1mat~d	pv1_be~2
deab_aga	1		
pv1mathsd	**-0.2410***	1	
pv1_belowp~2	**-0.7182***	**0.2478***	1
Italy	deab_aga	pv1mat~d	pv1_be~2
deab_aga	1		
pv1mathsd	0.0592	1	
pv1_belowp~2	**-0.8185***	-0.0635	1
Total	deab_aga	pv1mat~d	pv1_be~2
deab_aga	1		
pv1mathsd	0.0235	1	
pv1_belowp~2	**-0.7573***	-0.0317	1

Notes: * indicates that the correlation is statistically significant at the 1% level.

Table 13 Indicators of equity at various points of the efficiency distribution

>75th	deab_effscore	pv1_belowprof2	pv1mathsd
Germany	0.889	0.024	64.046
Spain	0.871	0.142	76.758
France	0.884	0.041	65.566
UK	0.875	0.053	69.040
Italy	0.883	0.042	63.524
Total	**0.880**	**0.070**	**68.023**

75th-median	deab_effscore	pv1_belowprof2	pv1mathsd
Germany	0.819	0.095	62.983
Spain	0.819	0.179	78.942
France	0.820	0.151	63.451
UK	0.816	0.115	73.620
Italy	0.819	0.141	63.331
Total	**0.818**	**0.152**	**71.495**

median-25th	deab_effscore	pv1_belowprof2	pv1mathsd
Germany	0.771	0.234	70.644
Spain	0.772	0.258	80.976
France	0.771	0.277	60.924
UK	0.771	0.215	79.539
Italy	0.771	0.250	61.865
Total	**0.772**	**0.245**	**73.457**

<25th	deab_effscore	pv1_belowprof2	pv1mathsd
Germany	0.679	0.528	64.230
Spain	0.710	0.429	79.418
France	0.677	0.588	59.918
UK	0.707	0.389	78.430
Italy	0.678	0.590	62.550
Total	**0.689**	**0.519**	**68.292**

Notes: the efficiency scores are the bootstrap, bias-corrected ones derived from the baseline model. The indicators for equality and equity (*pv1_belowprof2* and *pv1mathsd*) are defined as described in the section §3.3.

Table 14 Robustness check: Stochastic Frontier Analysis – production function's coefficients

	ln(pv1math)	ln(pv1read)
ln(ESCS)	-0.23	-0.023
	0.12	0.126
ln(StRatio)	0.005	-0.036
	0.055	0.063
ln(Computer_n)	0.022	0.053
	0.04	0.038
ln(ESCS)^2	0.318***	0.286***
	0.037	0.043
ln(StRatio)^2	-0.001	-0.002
	0.008	0.01
ln(Computer_n)^2	-0.009**	-0.010*
	0.003	0.005
ln(ESCS)_ln(StRatio)	-0.075	0.039
	0.091	0.113
ln(ESCS)_ln(Computer)	-0.141**	-0.126*
	0.046	0.052
ln (StRatio)_ln(Computer)	-0.033	0.005
	0.027	0.035
Spain	0.064***	0.077***
	0.012	0.013
France	0.002	0.030*
	0.013	0.014
UK	-0.021	0
	0.012	0.013
Italy	0.02	0.018
	0.013	0.013
Constant	6.046***	5.927***
	0.118	0.127
Sigma(v)	-5.706***	-5.799***
	0.155	0.218
Sigma(u)	-3.887***	-3.503***
	0.17	0.15
N	2601	2601

Notes: All the variables (including the dependent variable) are transformed in natural logarithms. Sigma(v) and Sigma(u) are the variance for random error and inefficiency components, respectively. The results are obtained through the command Frontier in STATA©.

Table 15 Robustness check: Stochastic Frontier Analysis – efficiency scores

Germany	Mathematics	Reading
Mean	0.913	0.898
Median	0.930	0.915
SD	0.057	0.065
Min	0.668	0.688
Max	0.981	0.982
N	184	184
Spain	Mathematics	Reading
Mean	0.902	0.881
Median	0.913	0.888
SD	0.047	0.057
Min	0.729	0.609
Max	0.981	0.984
N	813	813
France	Mathematics	Reading
Mean	0.921	0.902
Median	0.941	0.933
SD	0.060	0.081
Min	0.611	0.522
Max	0.980	0.982
N	179	179
UK	Mathematics	Reading
Mean	0.902	0.893
Median	0.908	0.900
SD	0.043	0.051
Min	0.719	0.665
Max	0.977	0.979
N	429	429
Italy	Mathematics	Reading
Mean	0.905	0.892
Median	0.921	0.917
SD	0.062	0.079
Min	0.585	0.440
Max	0.987	0.987
N	996	996
Total	Mathematics	Reading
Mean	0.905	0.890
Median	0.917	0.904
SD	0.055	0.068
Min	0.585	0.440
Max	0.987	0.987
N	2,601	2,601

Notes: estimated through the command Frontier, STATA©. The model was unable to estimate the scores for 118 schools that were consequently dropped by these statistics.

Table 16 Robustness check: Stochastic Frontier Analysis – correlation of efficiency scores with DEA ones

	Pearson's correlations			Spearman's (rank) correlations		
Germany	DEA	SFA (math)	SFA (read)	DEA	SFA (math)	SFA (read)
DEA	1			1		
SFA (math)	0.8089*	1		0.7601*	1	
SFA (read)	0.8116*	0.8761*	1	0.7435*	0.8675*	1
Spain	DEA	SFA (math)	SFA (read)	DEA	SFA (math)	SFA (read)
DEA	1			1		
SFA (math)	0.6949*	1		0.7054*	1	
SFA (read)	0.7607*	0.8335*	1	0.7681*	0.8156*	1
France	DEA	SFA (math)	SFA (read)	DEA	SFA (math)	SFA (read)
DEA	1			1		
SFA (math)	0.8873*	1		0.8810*	1	
SFA (read)	0.9333*	0.8952*	1	0.9144*	0.8576*	1
UK	DEA	SFA (math)	SFA (read)	DEA	SFA (math)	SFA (read)
DEA	1			1		
SFA (math)	0.7764*	1		0.7622*	1	
SFA (read)	0.8142*	0.8168*	1	0.8061*	0.8159*	1
Italy	DEA	SFA (math)	SFA (read)	DEA	SFA (math)	SFA (read)
DEA	1			1		
SFA (math)	0.8166*	1		0.8097*	1	
SFA (read)	0.8496*	0.7835*	1	0.8452*	0.7740*	1

Notes: The symbol (*) means that the correlation index is statistically significant at conventional 1% level. In grey, the relevant correlations (i.e. those between (bootstrap) DEA efficiency score and the two sets of scores obtained through SFA.

Table 17 Robustness check: alternative DEA models – correlation of efficiency scores with baseline ones

Pearson's correlations	DEAb (baseline)	RC1	RC2	RC3	RC4	RC5	RC6	RC7
DEAb (baseline)	1							
RC1	0.9819*	1						
RC2	0.8328*	0.8167*	1					
RC3	0.9848*	0.9662*	0.8091*	1				
RC4	0.9751*	0.9657*	0.8280*	0.9634*	1			
RC5	0.9795*	0.9691*	0.8240*	0.9683*	0.9783*	1		
RC6	0.9723*	0.9612*	0.8089*	0.9603*	0.9752*	0.9707*	1	
RC7	0.9732*	0.9666*	0.8145*	0.9618*	0.9761*	0.9761*	0.9719*	
Spearman's correlations	DEAb (baseline)	RC1	RC2	RC3	RC4	RC5	RC6	RC7
DEAb (baseline)	1							
RC1	0.9806*	1						
RC2	0.7957*	0.7742*	1					
RC3	0.9798*	0.9600*	0.7633*	1				
RC4	0.9761*	0.9644*	0.7972*	0.9606*	1			
RC5	0.9810*	0.9677*	0.7896*	0.9661*	0.9790*	1		
RC6	0.9703*	0.9576*	0.7730*	0.9552*	0.9748*	0.9704*	1	
RC7	0.9741*	0.9650*	0.7800*	0.9587*	0.9785*	0.9771*	0.9744*	1

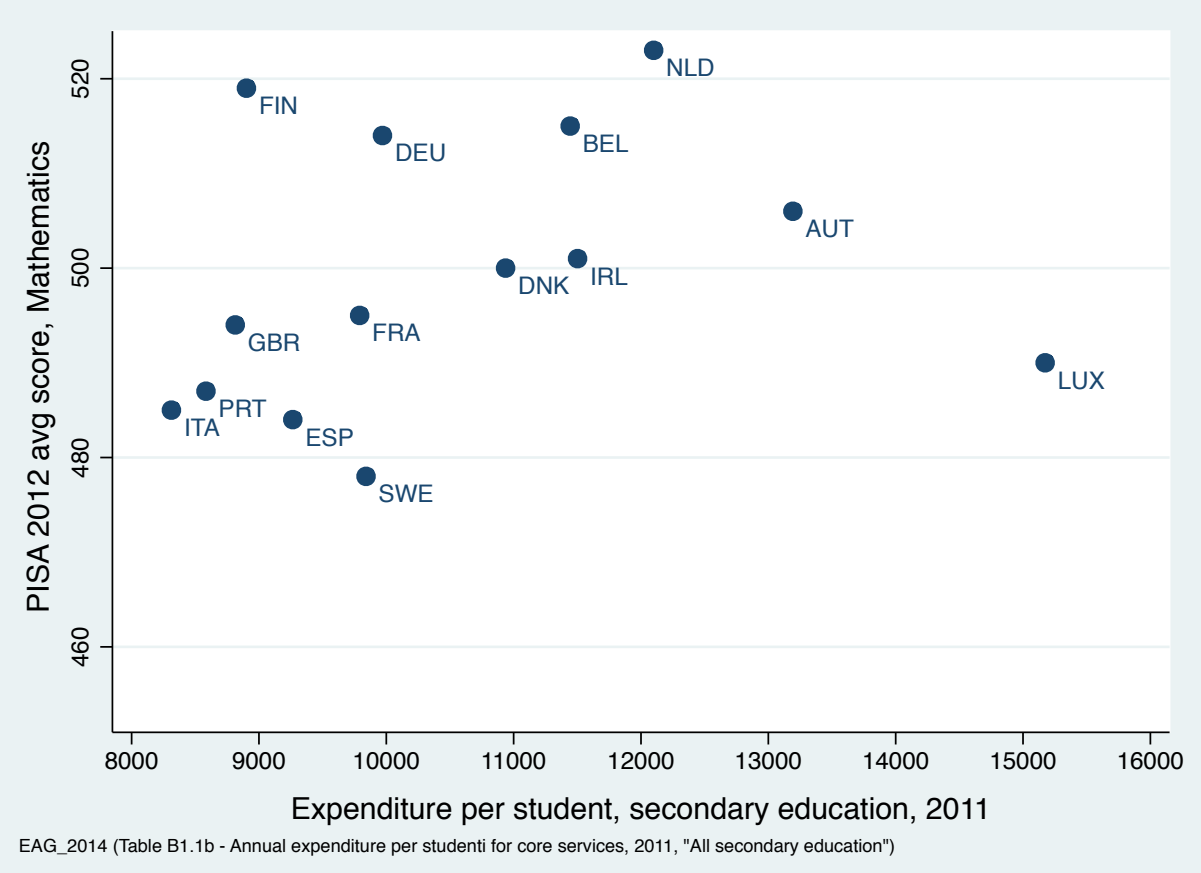
Notes: models from RC1 to RC7 are defined as illustrated in the section § 5.2. The symbol (*) means that the correlation index is statistically significant at conventional 1% level.

Table 18 Robustness check: Simar & Wilson (2007) double-bootstrap second-stage regression – comparison of results with Tobit regression

	Simar & Wilson (2007)			Tobit		
	Coeff.	Lower bound	Upper bound	Coeff.	Lower bound	Upper bound
isced2	**0.053**	0.009	0.075	**0.048**	0.031	0.065
progn_germany1	**-0.085**	-0.099	-0.034	**-0.075**	-0.093	-0.058
orgen	0.001	-0.025	0.035	0.004	-0.011	0.019
pv1mathsd	0.000	-0.001	0.000	0.000	-0.001	0.000
ESCSsd	0.047	-0.001	0.090	**0.046**	0.026	0.066
pv1_belowprof2	**0.394**	0.344	0.426	**0.389**	0.371	0.408
private	**0.034**	0.011	0.051	**0.032**	0.022	0.042
size	0.000	0.000	0.000	0.000	0.000	0.000
clsize_small	**0.062**	0.014	0.102	**0.061**	0.045	0.077
immig1	**-0.077**	-0.129	-0.006	**-0.073**	-0.104	-0.043
female	**-0.126**	-0.161	-0.081	**-0.124**	-0.142	-0.105
out_hours	**0.014**	0.008	0.019	**0.013**	0.011	0.016
scmatedu	**-0.009**	-0.015	-0.001	**-0.008**	-0.012	-0.005
mminutes	-0.001	-0.001	0.000	-0.001	-0.001	0.000
prop_cert	-0.028	-0.068	0.012	**-0.027**	-0.044	-0.011
competition	0.001	-0.013	0.018	0.002	-0.006	0.010
st_truancy	**0.067**	0.010	0.110	**0.064**	0.036	0.092
t_morale	**-0.021**	-0.038	-0.001	**-0.020**	-0.029	-0.011
monitor_p	-0.001	-0.019	0.020	0.000	-0.010	0.009
select_1	0.014	-0.004	0.031	**0.013**	0.005	0.022
grouping	-0.002	-0.017	0.015	-0.002	-0.010	0.006
Spain	**-0.085**	-0.109	-0.043	**-0.081**	-0.097	-0.064
France	-0.013	-0.048	0.026	-0.012	-0.026	0.003
UK	0.042	-0.011	0.078	**0.038**	0.015	0.061
Italy	-0.019	-0.066	0.030	-0.018	-0.041	0.004

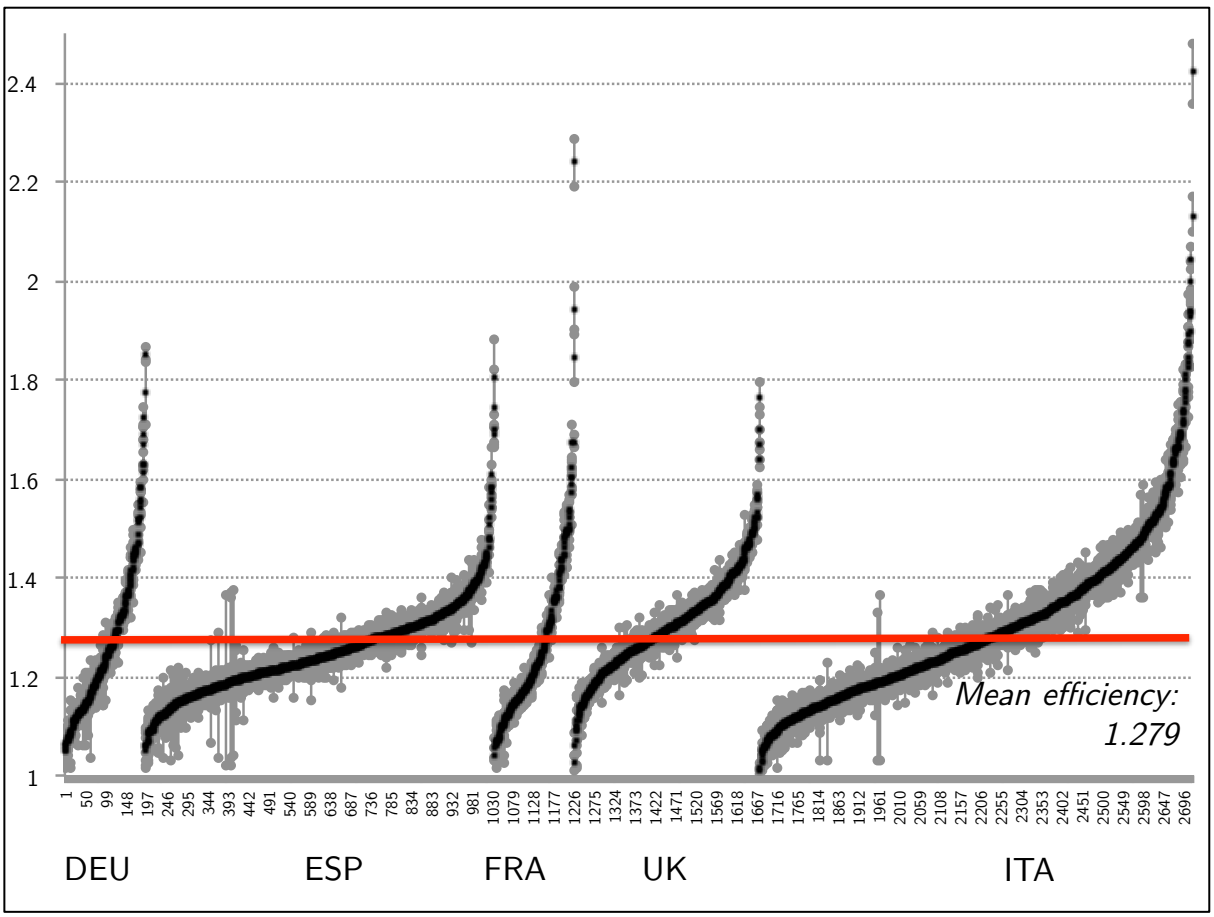
Notes. In bold, I signal the statistically significant coefficients. The efficiency scores used as dependent variable are the original, output-oriented ones (which range is between $[1, \infty]$ where 1 is the maximum level of efficiency). Thus, they should be interpreted as the factors associated with in-efficiency – the interpretation is the contrary of the one presented in previous tables.

Figure 1 The relationship between expenditure per student and results in PISA 2012 (mathematics), EU-15 countries



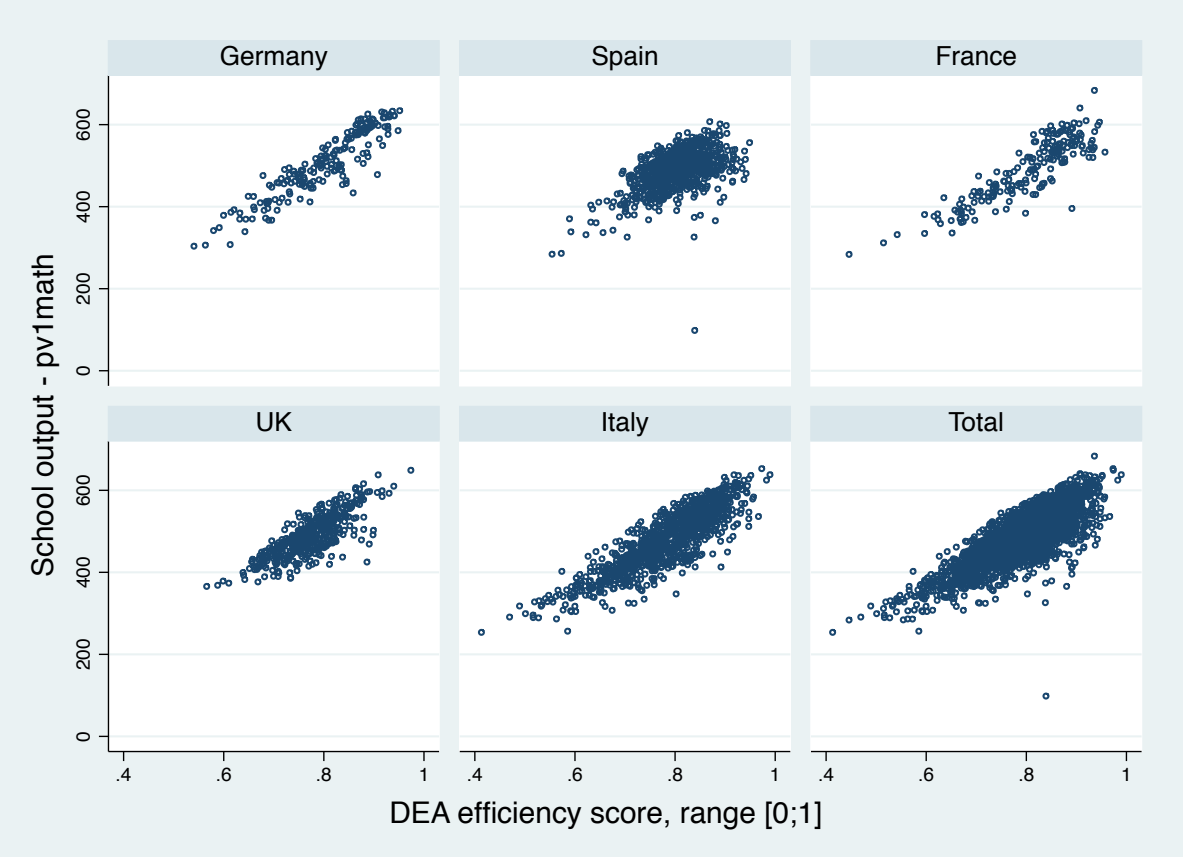
Notes: author's elaborations on OECD (2013a) and OECD (2014). Average scores in Mathematics are those reported in OECD (2013a), Table I.2.3a; expenditure per students refer to 2011, and is reported in the (web-only) table B1.1b; the correct complete definition is: Annual expenditure per student for core services, 2011, "All secondary education"

Figure 2 A graphical representation of bias-corrected (bootstrap) efficiency scores, international frontier, by country



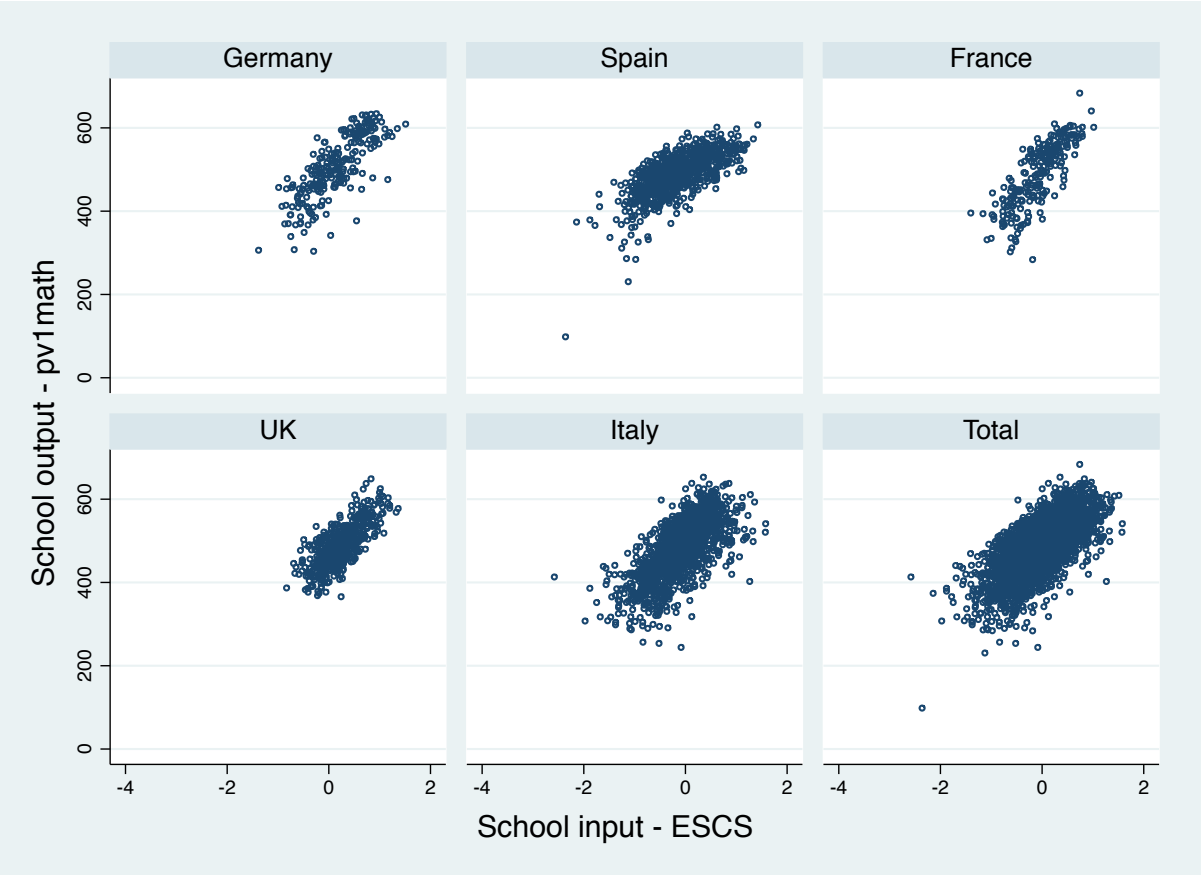
Notes: output-oriented scores are reported (max efficiency=1; inefficient schools if eff_score>1). The red line indicates the overall mean efficiency score (=1.249165). Schools (horizontal axis) are ordered (by country) on the basis of growing efficiency score. Grey lines represent 95% confidence interval.

Figure 3 The relationship between average performance in mathematics (vertical axis) and efficiency scores (horizontal axis), by country and total (international frontier)



Notes: output-oriented scores are reported on the horizontal axis (max efficiency=1; inefficient schools if eff_score>1); the horizontal axis is ordered on the basis of growing efficiency score. The score in mathematics is pv1math as defined in the section about Data.

Figure 4 The relationship between performance (pv1math) and one input (ESCS), by country



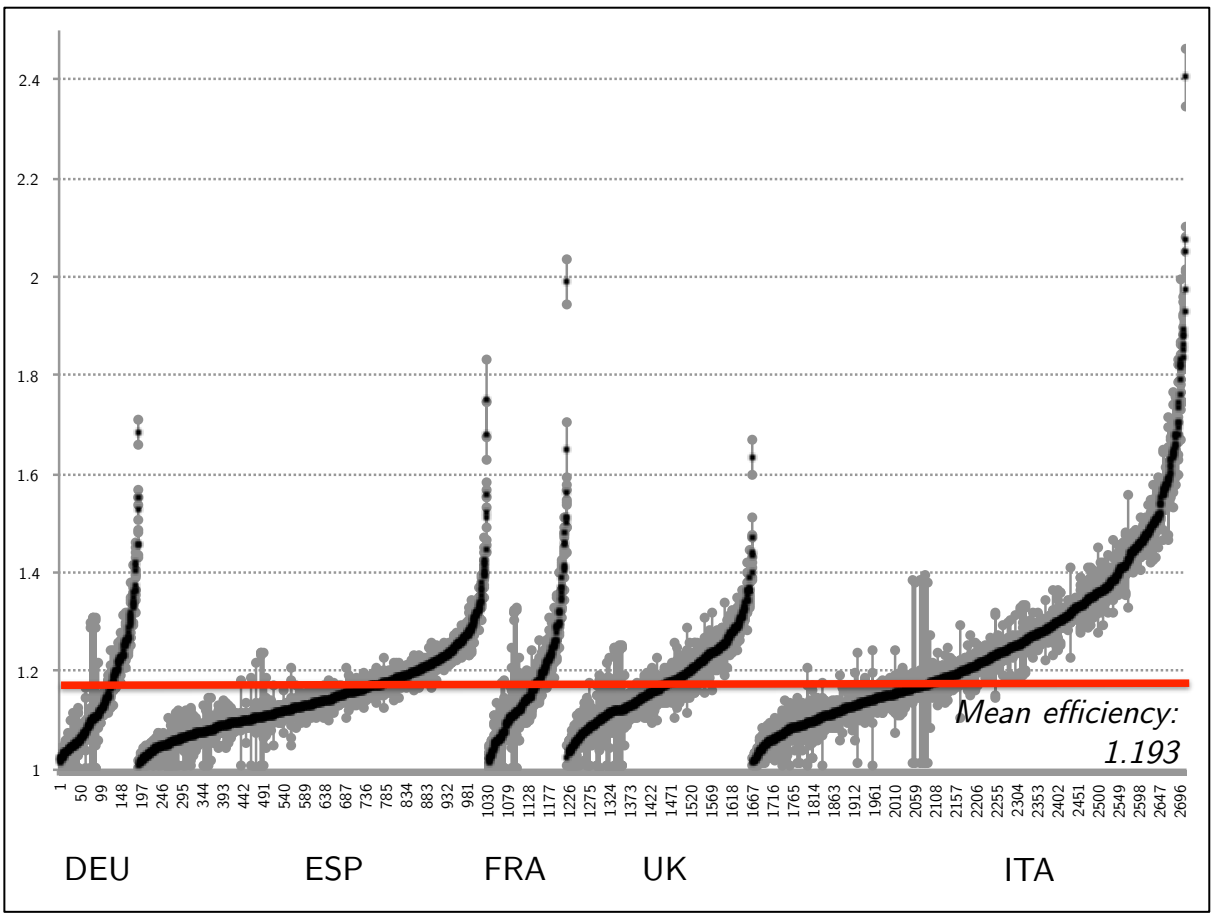
Notes: the figure illustrates the extent to which scale economies can be graphically detected. On the horizontal axis, one input (school-average ESCS) is reported; on the vertical axis, one of the two outputs is reported (the analogous figure for the second output is available, in request, from the author).

Figure 5 The correlation between size (horizontal axis), performance (left panel) and efficiency (right panel)



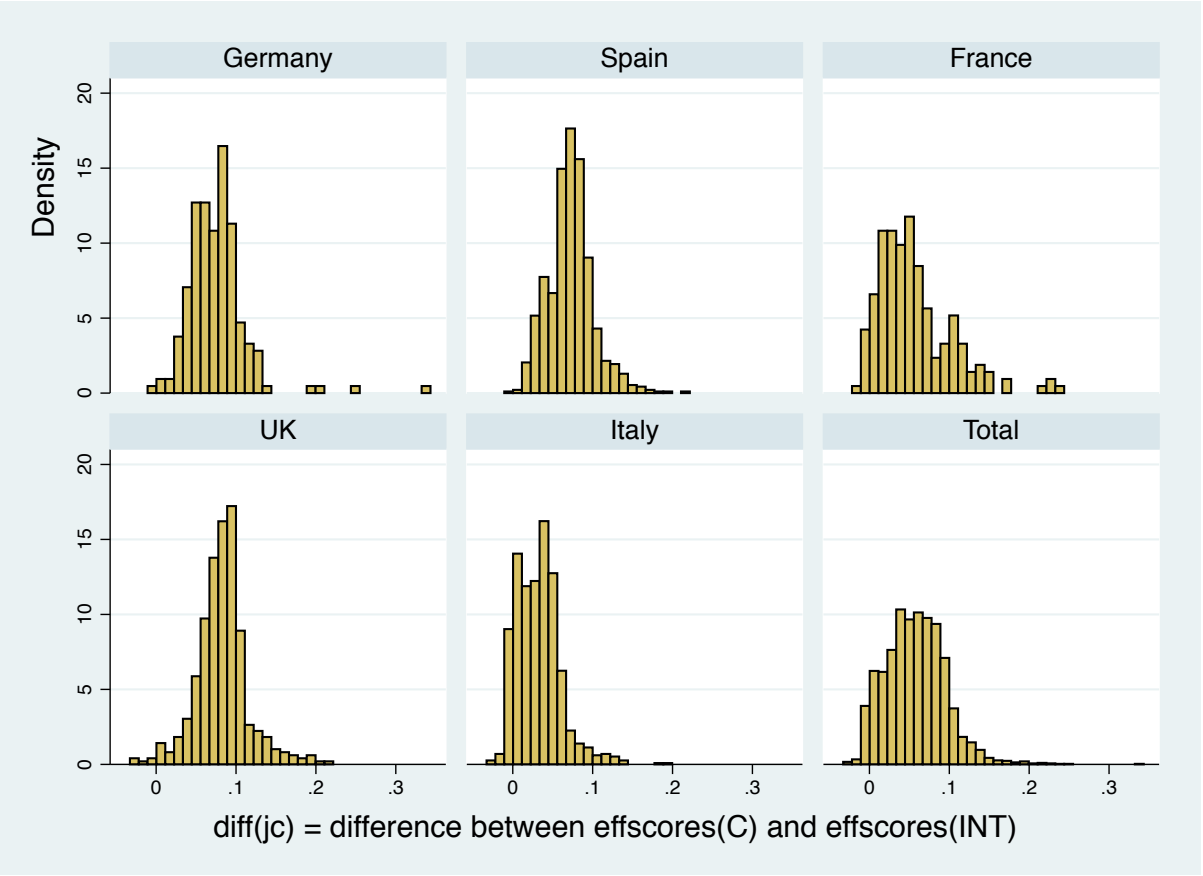
Notes: the figure illustrates the extent to which size economies (or efficiencies) can be graphically detected. On the horizontal axis, school size is reported; on the vertical axis: (i) in the left panel one of the two outputs is reported (the analogous figure for the second output is available, on request, from the author); (ii) in the right panel, it is the efficiency score (bootstrap, bias-corrected, from the baseline model).

Figure 6 A graphical representation of bias-corrected (bootstrap) efficiency scores, country-specific frontier, by country



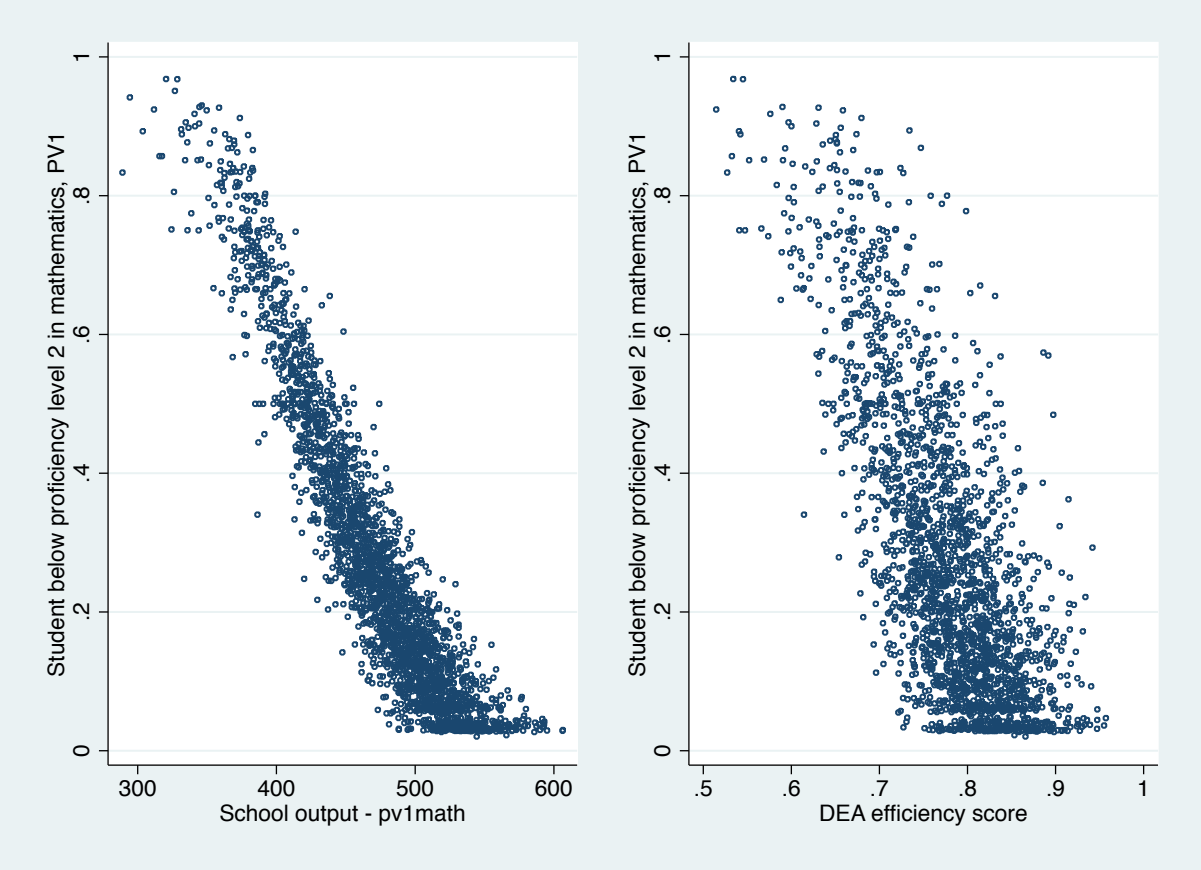
Notes: output-oriented scores are reported (max efficiency=1; inefficient schools if eff_score>1). The red line indicates the overall mean efficiency score (=1.171). Schools (horizontal axis) are ordered (by country) on the basis of growing efficiency score. Grey lines represent 95% confidence interval.

Figure 7 The difference between the estimated efficiency scores: country-specific vs international frontier



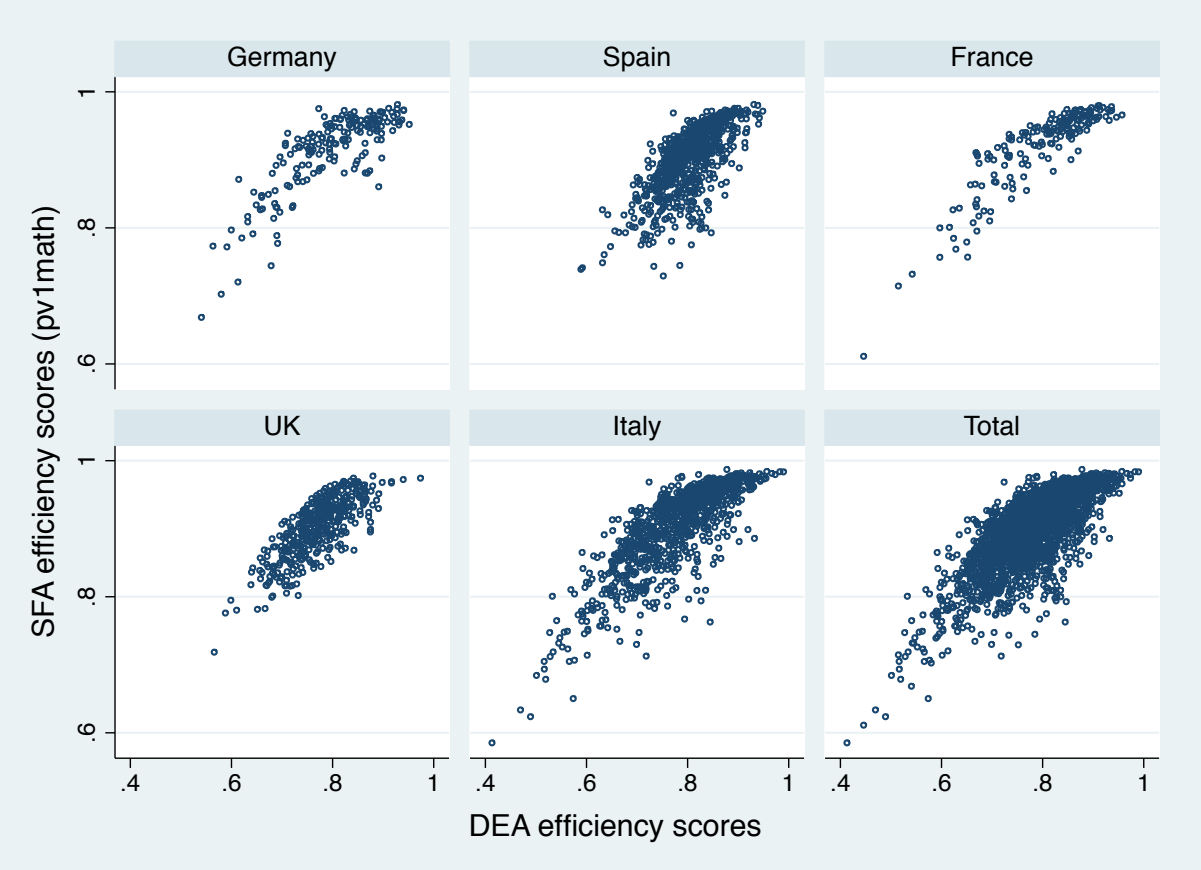
Notes: density is reported on the vertical axis. In the horizontal axis, the variable is calculated as the difference between the efficiency score estimated with a country-specific frontier, and the score estimated by considering an international frontier. Mathematically: $\boldsymbol{diff_{jc} = eff_{ic(C)} - eff_{jc(INT)}}$ (see section §4.2).

Figure 8 Efficiency and equity: relationships between performance (test scores), efficiency and the proportion of students below proficiency level 2



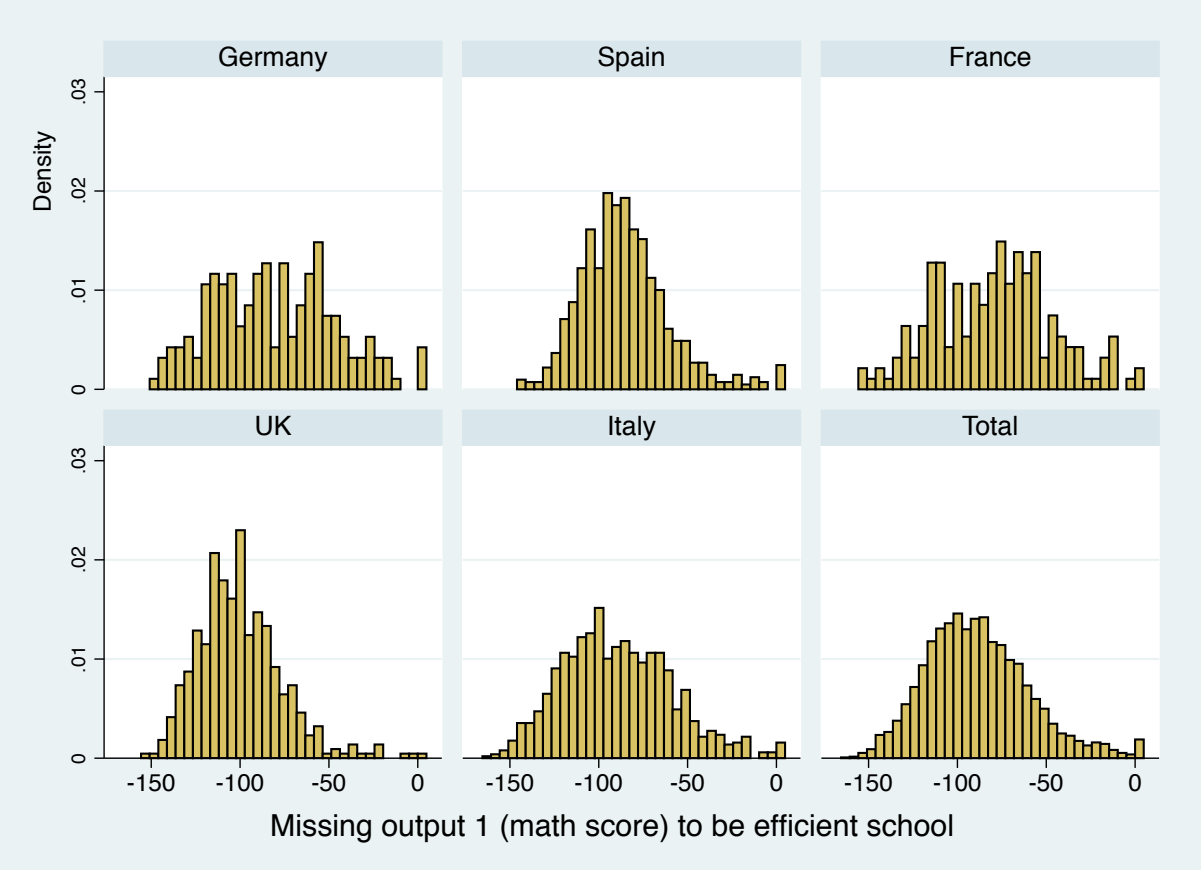
Notes: the schools with pv1_belowprof2 =0 or =1 have been excluded by the graphs.

Figure 9 Robustness check: Stochastic Frontier Analysis – graphical correlation between DEA and SFA efficiency scores (output: mathematics)



Notes: the analogous graph for SFA efficiency scores where the output=reading is available, on request, from the author.

Figure 10 Potential output expansion based on efficiency scores' calculations, by country (reference output: mathematics score [pv1math], holding inputs and pv1read constant)



Notes: The calculations are made with the package "Benchmarking" in R; the model is the baseline (traditional, not bootstrap) DEA as specified in the section §3.1.

Annex 1

Table A1.1 The distribution of students across grades, by country – original PISA2012 dataset

Grade	Spain	UK	Italy	Germany	France
7	16		82	32	2
8	2,090		432	497	84
9	5,451	2	4,663	**2,606**	1,254
10	**17,744**	166	**25,313**	**1,829**	**3,117**
11	12	**10,330**	582	37	150
12		2,160	1		6
13		1			
Total	25,313	12,659	31,073	5,001	4,613

Notes: data from the student-level dataset, PISA2012

Table A1.2 Level of education, years in the school of the students in the modal grade, and expected duration of that school level, by country

	Level	Years in the school	Normal legal duration
Germany	ISCED-2	6	6
Spain	ISCED-2	4	4
France	ISCED-3	1	3
UK	ISCED-3	4	7
Italy	ISCED-3	2	5
Mean	-	3.4	5

Notes: author's elaborations on data extracted from OECD's Education GPS (http://gpseducation.oecd.org).

Table A1.3 Number of students enrolled in the schools included in the final dataset

	In the original sample	In the final sample	%
Germany	5,001	4,258	85.1%
Spain	25,313	23,645	93.4%
France	4,613	3,982	86.3%
UK	12,659	11,188	88.4%
Italy	30,173	27,464	91.0%
Total	77,759	70,537	90.7%

Notes: these are the numbers of students included in the dataset before collapsing it at school level.

Annex 2

Table A2.1 Pairwise statistical correlations between inputs and outputs, by country

Germany	ESCS	StRatio	Computer_n	pv1math	pv1read
ESCS	1				
StRatio	-0.1087	1			
Computer_n	-0.1387	-0.0174	1		
pv1math	0.7918*	-0.1028	-0.1539	1	
pv1read	0.7683*	-0.0864	-0.2022*	0.9428*	1
Spain	ESCS	StRatio	Computer_n	pv1math	pv1read
ESCS	1				
StRatio	-0.2113*	1			
Computer_n	0.0532	0.0949*	1		
pv1math	0.6970*	-0.1715*	0.1262*	1	
pv1read	0.6610*	-0.1821*	0.0988*	0.8662*	1
France	ESCS	StRatio	Computer_n	pv1math	pv1read
ESCS	1				
StRatio	0.0129	1			
Computer_n	-0.1818*	0.1420	1		
pv1math	0.8068*	0.1034	-0.1068	1	
pv1read	0.7790*	0.072	-0.0757	0.9485*	1
UK	ESCS	StRatio	Computer_n	pv1math	pv1read
ESCS	1				
StRatio	0.1308*	1			
Computer_n	-0.0458	0.2122*	1		
pv1math	0.7430*	0.1692*	0.0343	1	
pv1read	0.7039*	0.1249*	0.0605	0.9155*	1
Italy	ESCS	StRatio	Computer_n	pv1math	pv1read
ESCS	1				
StRatio	-0.1592*	1			
Computer_n	-0.0324	0.2184*	1		
pv1math	0.6876*	-0.1603*	0.0458	1	
pv1read	0.7125*	-0.1612*	0.0013	0.9043*	1
Total	ESCS	StRatio	Computer_n	pv1math	pv1read
ESCS	1				
StRatio	-0.2031*	1			
Computer_n	0.0060	0.0471	1		
pv1math	0.6853*	-0.1502*	0.0221	1	
pv1read	0.6841*	-0.1612*	0.0063	0.9067*	1

Notes: * indicates that the correlation is statistically significant at 1% level. Recall that the indicators are measured at school (not individual) level.

Annex 3

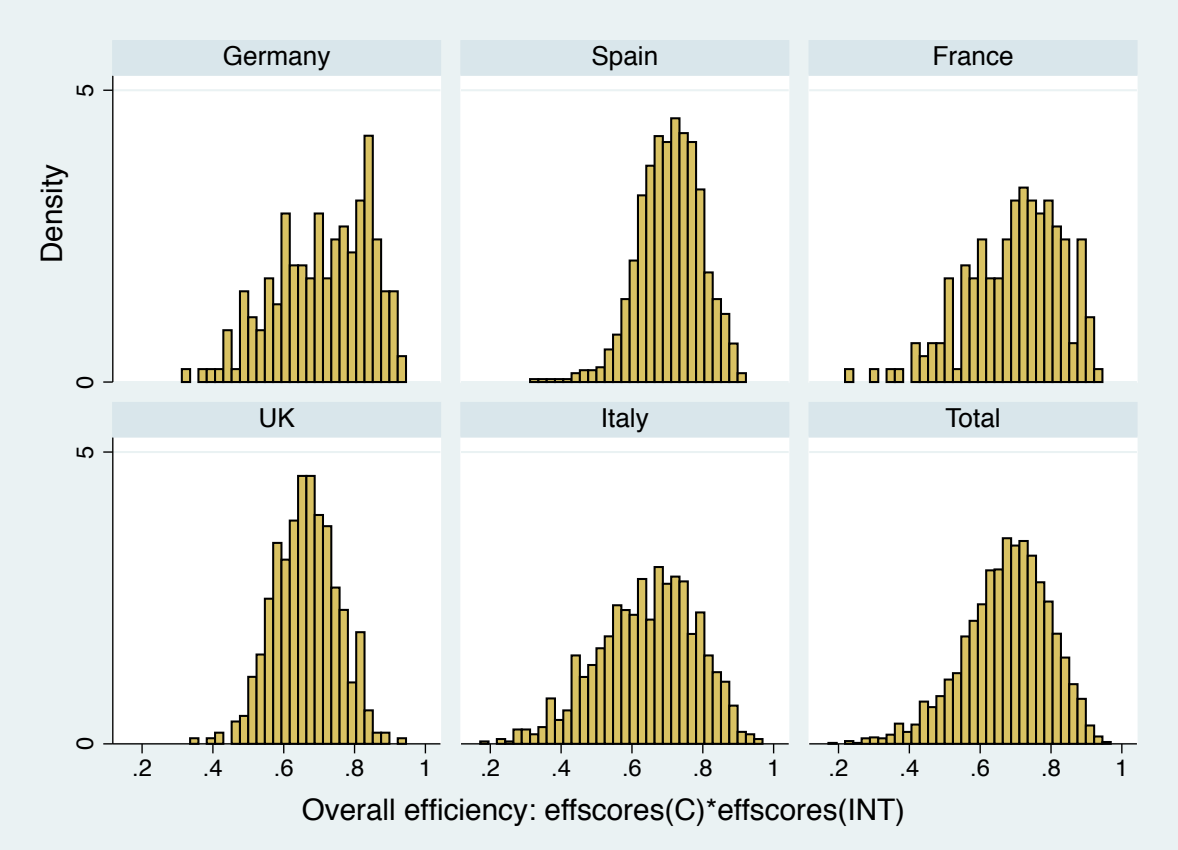
In this annex, I report the main results (descriptive statistics and distribution) about the "overall" schools' efficiency, defined as the quotient between the efficiency scores obtained using the country-specific and the international frontier as a benchmark.

Table A3.1 "Overall" efficiency scores, by country

	mean	median	sd	min	max	N
Germany	0.709	0.727	0.134	0.321	0.931	192
Spain	0.707	0.712	0.089	0.316	0.915	840
France	0.698	0.717	0.133	0.224	0.926	192
UK	0.662	0.664	0.090	0.346	0.922	446
Italy	0.643	0.654	0.139	0.172	0.969	1,042
Total	0.675	0.685	0.121	0.172	0.969	2,712

Notes: Overall efficiency scores are calculated as described in the section §4.2:
$overall(eff_{jc}) = eff_{jc(C)} \times eff_{jc(INT)}$

Figure A3.1 The distribution of "overall" efficiency scores, by country



Notes: density is reported on the vertical axis