Assessing European primary school performance through a conditional nonparametric model

José Manuel Cordero[1]*, Daniel Santín[2] and Rosa Simancas[1]

[1]University of Extremadura, Badajoz, Spain; and [2]Complutense University of Madrid, Madrid, Spain

This paper uses a fully nonparametric framework to assess the efficiency of primary schools using data about schools in 16 European countries participating in PIRLS 2011. This study represents an original enterprise since most of the empirical research in the field is restricted to evaluations at regional or national level and focused on secondary education. For our purpose, we adapt the metafrontier framework to compare and decompose the technical efficiency of primary schools operating in heterogeneous contexts, which in our case is represented by different educational systems or countries. Similarly, we use an extension of the conditional nonparametric robust approach to test the potential influence of a mixed set of environmental school factors and variables representing cultural values of each country. Our results indicate that the intergenerational transmission of non-cognitive skills such as responsibility or perseverance are significantly related to school efficiency, whereas most school factors do not seem to have a significant influence on school performance.

Introduction

All countries are interested in improving the level of education of their citizens because it is considered as one of the main sources of human development (Krueger and Lindahl, 2001) as well as an important source of economic growth (Hanushek and Kimko, 2000; Barro, 2001; Hanushek and Woessmann, 2008). This evidence can explain why so many studies have attempted to explore the potential determinants of education using an educational production function (Hanushek, 1979), in which the quality of education is usually measured by test scores. In this sense, the participation of the majority of nations on common international large-scale assessments like PISA (*Programme for International Student Assessment*), TIMSS (*Trends in International Mathematics and Science Study*) or PIRLS (*Progress in International Reading Literacy Study*) has provided researchers with rich and extensive cross-national databases that can be used to assess the performance of educational systems from a comparative perspective. As a result, we can find an extensive literature using the entire world as a laboratory to explore the underlying determinants of educational achievement (Bray and Thomas, 1995; Hanushek and Woessman, 2011) or to analyse specific aspects such as the differences between public and private schools (Vandenberghe and Robin, 2004; Dronkers and Roberts, 2008), the effects of tracking (Brunello and Checchi, 2007; Schuetz *et al*, 2008) or the influence of accountability

(Bishop, 1997; Fuchs and Woessmann, 2007) to cite only some examples.

Since their main focus is the identification of significant relationships between educational outcomes and students and school-related variables, most of those cross-national studies apply econometric techniques. More recently, some of them have started to apply more sophisticated methods in order to identify causal relationships in the international data on educational achievement (see Hanushek and Woessmann, 2014 or Strietholt *et al*, 2014 for a review), although they do not consider the potential existence of an unexpected level of inefficiency in the performance of schools (Levin, 1974). In this sense, the actual constraints of resources faced by most of countries and the great amount of national income devoted to education expenditures have led policy makers and researchers to become increasingly concerned with assessing the efficiency of schools. However, most of these efficiency evaluations have been restricted to schools operating in the same country or region (see Grosskopf *et al*, 2014 for a recent review of this literature).

To the best of our knowledge, only few studies have applied frontier methods to micro data from those international data sets to evaluate the performance of educational systems using a cross-country approach. These include De Jorge and Santín (2010) and Deutsch *et al* (2013), which use PISA data at student level to estimate the efficiency of EU and Latin America countries respectively, while Wilson (2005) use PISA data at school level to assess the performance of 40 countries around the world. Moreover, we can find some empirical works using

Correspondence: José Manuel Cordero, Departamento de Economía, Universidad de Extremadura, Av. Elvas s/n, Badajoz 06071, Spain.
E-mail: jmcordero@unex.es

data aggregated at country level from different samples of countries participating in PISA (Afonso and St Aubyn, 2006; Giambona *et al*, 2011; Thieme *et al*, 2012; Aristovnik and Obadić, 2014) or TIMSS (Giménez *et al*, 2007). Those studies use predominantly the nonparametric data envelopment analysis to obtain efficiency measures of performance. (Deutsch *et al*, 2013, use corrected least squares.) Only in some cases, a two-stage procedure is also applied to examine the potential influence of contextual variables on efficiency estimates (eg Afonso and St Aubyn, 2006; Verhoeven *et al*, 2007), but none of them incorporate this information into the estimation of efficiency scores.

In this paper we attempt to extend this scarce body of literature on cross-country efficiency analysis in the education sector by performing an assessment of primary schools operating in 16 European countries. For that purpose, we apply some recently developed nonparametric methods that allow us to overcome some of the main limitations of previous studies. Similarly, we incorporate into the analysis some additional data about contextual factors in each nation that can be extremely helpful to shed light on the divergences in performance across countries.

Our aim consists of applying some of the most recently developed methods in the literature of efficiency analysis to estimate measures of school performance in an international context. In particular, we use the robust order-*m* methodology described by Cazals *et al* (2002) to avoid some of the main drawbacks of nonparametric methods. This approach consists of constructing a partial frontier using only part of the sample (*m* observations) to determine efficiency score; thus, the impact of potential outliers in data is reduced. Subsequently, we adapt the metafrontier framework developed by Battese and Rao (2002), Battese *et al* (2004) and O'Donnell *et al* (2008) to decompose the estimated inefficiency between two different levels (school and country). Finally, we use the conditional nonparametric approach proposed by Dario and Simar (2005, 2007a, b) to incorporate the effect of environmental factors into the estimation of efficiency scores and explore their potential influence without assuming the typical restrictive separability condition of two-stage approaches.

Some of those methodologies have been previously applied in empirical studies with educational data for specific countries. For example, De Witte *et al* (2010) used the order-*m* approach to assess the performance of a sample of British secondary school pupils and Thieme *et al* (2013) combined this approach with a metafrontier approach to evaluate students in primary education in Chile. Similarly, the conditional approach has also been applied to evaluate public and private schools in Flanders (Cherchye *et al*, 2010), to evaluate the performance of teachers in Belgium (De Witte and Rogge, 2011), to analyse the impact of innovations on school performance in the Netherlands (Haelermans and De Witte, 2012) or to assess the performance of Dutch students (De Witte and Kortelainen, 2013). However, this paper represents the first combined application of those methods in a cross-country analysis using educational data from an international large-scale survey.

In particular, data used in our empirical analysis was retrieved from PIRLS 2011. This project, conducted by the International Association for the Evaluation of Educational Achievement (IEA), comprises data about students' reading achievement after four years of primary schooling. One of the main advantage of using this data set comes from the fact that most students have only attended one school, thus the effect of school is direct; while in studies based on PISA data students have usually studied in different centres (schools and high schools), thus part of the school effect is not captured by observed school variables since learning is a cumulative process. Moreover, the comparison among students in different educational systems is reasonably homogenous, since their results are not affected by tracking, which is not applied before the children are 10 years old in any country (see Brunello and Checchi, 2007, for details).

One of the main shortcomings of studies based on data from international large-scale assessments is their inability to understand the economic, cultural and social context of each country (Zhao *et al*, 2008; Täht and Must, 2013). In order to overcome this limitation we retrieved some additional information about economic indicators from the World Bank's Indicators database and collected data from some questions included in the World Values Survey (WVS) to approximate the cultural heritage of each country. Coco and Lagravinese (2014) also use this source of data to incorporate a proxy measure of hard work in their evaluation of education performance of OECD countries using PISA data. In our case, these variables are considered as potential factors that might affect the performance of schools operating in the same country when they are compared with schools in other countries.

The remainder of the paper is structured as follows. The next section describes the methodology. The subsequent section explains the main characteristics of the data and the variables selected for the empirical analysis. The penultimate section presents the main results and relates them to the existing literature. Finally, the paper ends with some concluding remarks in the final section.

Methodology

The robust FDH model

The definition of the production technology in the educational sector is a very difficult task. The only thing that we know is that pupils attending schools transform a set of heterogeneous inputs x, ($x \in \Re_+^p$), including their own abilities, school variables and parental background (Hanushek *et al*, 2013), into heterogeneous outputs y ($y \in \Re_+^q$), usually represented by test scores in a standardized assessment. The production technology is the set of all feasible input-output combinations:

$$\psi = \left\{ (x, y) \in \Re_+^{p+q} | x \text{ can produce } y \right\} \quad (1)$$

In order to estimate the relative efficiency of each school, we estimate a frontier that represents the best practice observations

following the main ideas developed in the seminal work of Farrell (1957). In particular, our model is based on the nonparametric Free Disposal Hull (FDH) methodology (Deprins *et al*, 1984). The output-oriented efficiency score $(\hat{\lambda}_{\text{FDH}})$ of an observation can be obtained by solving the linear programming problem in Equation (2):

$$\hat{\lambda}_{\text{FDH}} = \max\left\{ \lambda \,\middle|\, \lambda y \leqslant \sum_{i=1}^{N} \gamma_i y_i; \; x \geqslant \sum_{i=1}^{N} \gamma_i x_i; \; \sum_{i=1}^{N} \gamma_i = 1; \right.$$
$$\left. \gamma_i \in \{0, 1\}; \, i = 1, \, \ldots, \, n \right\} \tag{2}$$

However, since this technique is very sensitive to the presence of atypical observations in data, we adopt the robust order-*m* estimation method introduced by Cazals *et al* (2002). This approach is related to the FDH estimator, but instead of constructing a full frontier, it creates a partial frontier that envelops only *m* ($\geqslant 1$) observations randomly drawn with replacement from the empirical sample. This procedure is repeated *B* times resulting in multiple measures $(\hat{\lambda}_{mi}^1, \, \ldots, \, \hat{\lambda}_{mi}^B)$ from which the final order-*m* efficiency measure is computed as the simple mean $(\hat{\lambda}_{mi})$. Specifically, the order-*m* estimated efficiency score is derived from Equation (3) as follows:

$$\hat{\lambda}_m = E\left[\min_{i=1, \, \ldots, \, m} \left\{ \max_{j=1, \, \ldots, \, p} \left(\frac{x_i^j}{x^j} \right) \right\} \middle| y_i \geqslant y \right] \tag{3}$$

where the ρ-dimensional random variables $x_i, \ldots, x_m$ are drawn randomly and repeatedly from the conditional distribution of *X* given $y_i \geqslant y$. For acceptable *m* values, the efficiency scores will present values higher than unity, which indicates that schools are inefficient, as outputs can be increased without modifying the level of inputs. When $\hat{\lambda} < 1$, the unit can be labelled as super-efficient, since the order-*m* frontier exhibits lower levels of outputs than the unit under analysis (Daraio and Simar, 2007a).

The metafrontier approach

Given that our data has a hierarchical structure (schools operating in different countries), we adapt the concept of a metafrontier developed by Battese and Rao (2002), Battese *et al* (2004) and O'Donnell *et al* (2008). This approach measures the efficiency of units relative to separate best practice frontiers and allows us to decompose which part of the performance can be attributed to the schools and which part depends on country factors. This approach is basically an extension of the decomposition of school effect from students' inefficiency introduced by Ruggiero (2000), Silva-Portela and Thanassoulis (2001) and Thanassoulis and Silva-Portela (2002). Therefore, the case of schools operating within a country can be derived straightforward.

If we consider *K* different educational systems, each having its own distinctive features, a metafrontier is defined as the boundary of the unrestricted technology set. Hence, the metafrontier envelops each of the separate group frontiers (one frontier for each country). Separately, the local efficiency of the schools with regard to the special characteristics of the country where it is operating is measured relative to the n_k observations in the school sample. If all the countries have the same characteristics, all the observations can be pooled and schools can be evaluated relative to the same standards. As a result, two different frontiers are estimated: the local frontier specific for each educational system and the overall frontier. The distance to the local frontier depends only on the school efficiency (*SCE*) whereas the distance separating the local and the overall frontier can be interpreted as the country effect (*CNE*). This can be illustrated in Figure 1, where the efficiency level of each school *c* depends on the level of the output achieved (y_c) using their input endowment (x_c). This school is inefficient, since there are other schools operating in the same educational system obtaining better results (y') with the same amount of inputs (x_c). The
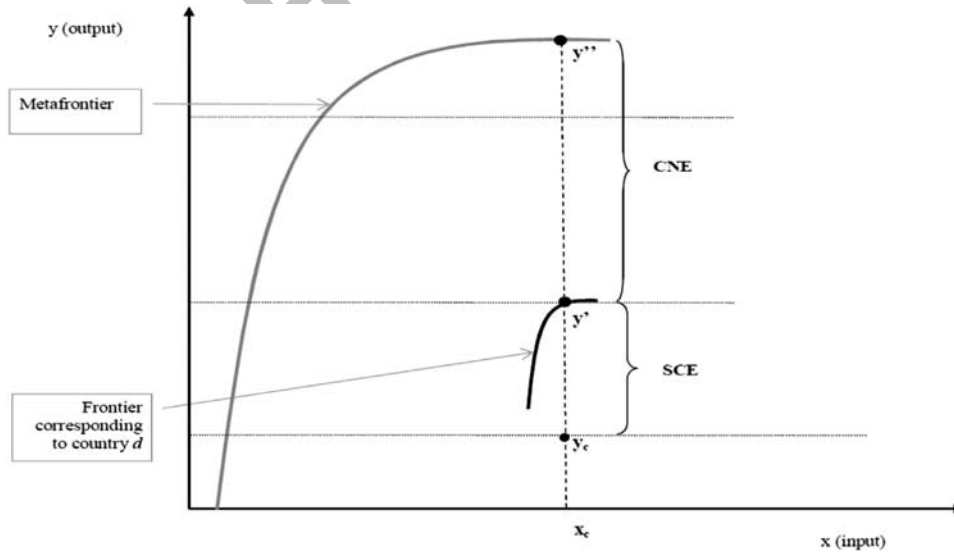
Figure 1 Metafrontier illustration (decomposition of school and country effect).

school inefficiency can be defined by the ratio between the local potential output divided by the actual output ($SCE = y'/y_c$). When this school is compared with the metafrontier, the overall efficiency (OE) can be defined as $OE = y''/y_c$. From those two measures of efficiency, the country effect can be automatically derived as $CNE = y''/y' = OE/SCE$. In summary, the global efficiency can be decomposed in two effects: $OE = SCE \times CNE$.

The robust and conditional FDH model

Once we have decomposed the efficiency of each school, the final step of our analysis consists of considering the effect of some exogenous variables $Z \in \Re_+^k$, affecting the performance of schools. If we do not consider the existing heterogeneity among schools, we would be implicitly assuming that all the schools are operating with the most favourable environment, which would not be real in many situations (Ruggiero, 1996). In our case, we are interested in testing the potential influence of some external variables at school level, but we also account for potential specific features at country level that can affect the performance of schools. For that purpose, we use the full nonparametric conditional approach developed by Cazals *et al* (2002) and Daraio and Simar (2005, 2007a, b), which assumes that both types of factors can have a direct influence on the shape of the best practice frontier (ie, this model does not assume a separability condition). Therefore, efficiency estimates are determined by inputs, outputs and exogenous variables. Using a probabilistic formulation, this conditional function can be defined as:

$$H_{XY|Z}(x, y|z)) = \Pr(Y \geqslant y | x \leqslant x, Z = z)$$

$$\Pr(X \leqslant x, Z = z)$$

$$= S_{Y|X,Z}(Y \geqslant y | X \leqslant x, Z = z)$$

$$F_X(X \leqslant x; Z = z) = S_Y(y|x, z)F_X(x|z) \qquad (4)$$

The function $H_{XY|Z}(x, y|z)$ represents the probability of a unit operating at level (x, y) being dominated by other units facing the same environmental conditions z. Using this formulation, the conditional order-m efficiency estimator can be defined as:

$$\hat{\lambda}_m(x, y|z) = \int_0^\infty \left[1 - \left(1 - \hat{S}_{Y|X}(uy|x, z) \right)^m \right] du \qquad (5)$$

The estimation of $S_Y(y|x, z)$ is more difficult than the unconditional case, because it requires using smoothing techniques for the exogenous variables in z (due to the equality constraint $Z = z$):

$$\hat{S}_{Y,n}(y|x, z) = \frac{\sum_{i=1}^n I(x_i \leqslant x, y_i \geqslant y)K_{\hat{h}}(z, z_i)}{\sum_{i=1}^n I(x_i \leqslant x)K_{\hat{h}}(z, z_i)} \qquad (6)$$

This approach relies therefore on the estimation of a nonparametric kernel function to select the appropriate reference partners and a bandwidth parameter h using some bandwidth choice method. In our case, we follow the data-driven selection approach developed by Badin *et al* (2010) and we use a generalized product kernel function, since we have mixed data (continuous and discrete variables) in our data set (see De Witte and Kortelainen, 2013, for details).

This conditional model allows us to evaluate the direction of the effect of external variables on the production process by comparing conditional with unconditional measures. In particular, Daraio and Simar (2005, 2007a) suggest using a scatter plot of the ratio between these measures ($Q^z = \hat{\lambda}_m(x, y|z)/\hat{\lambda}_m(x, y)$) against Z and its smoothed nonparametric regression line. In an output-oriented conditional model, an increasing regression line will indicate that Z is favourable to efficiency whereas a decreasing line will denote an unfavourable effect. Similarly, it is also possible to investigate the statistical significance of Z explaining the variations of Q. For that purpose, we use local linear least squares for regression estimation as recommended by Badin *et al* (2010) and Jeong *et al* (2010). We then apply the nonparametric regression significance test proposed by Li and Racine (2004) and Racine and Li (2004), which smooths both continuous and discrete variables. Specifically, we test the significance of each of the continuous and discrete variables using bootstrap tests proposed by Racine *et al* (2006) and Racine (1997), which can be interpreted as the nonparametric equivalent of standard *t*-tests in ordinary least squares regression (De Witte and Kortelainen, 2013).

Data and variables

In this study we use data from schools in European countries participating in PIRLS 2011. This data set provides international comparative data about students' reading achievement in the fourth year of primary schooling as well as a rich array of background information about students' socioeconomic status, the school environment and instructional practices (see Mullis *et al*, 2012, for details). This information comes from the responses given to different questionnaires completed by students, parents, teachers and school principals.

As we are interested in accounting for some specific characteristics of the countries where those schools are operating, we also retrieved data about different economic and social aspects from two additional sources. The economic information was collected from the World Bank Open Data section, while social indicators about cultural values come from pooled data about the five aggregate waves of the WVS.

Given that some of the European countries participating in PIRLS 2011 were not included in the WVS database, we had to restrict our analysis to only 16 countries for which we had data available from all the sources. Therefore, our final data set comprises a total number of 2398 schools distributed across countries as it is shown in Table 1.

The output variable is represented by the average of the results in reading of students attending the same school (PVREAD). These results are not expressed by only one value,

Table 1 Data set composition: number of schools in each country

Countries	Number of schools	Countries	Number of schools
Bulgaria	140	Lithuania	137
Czech Republic	153	Netherlands	128
Finland	121	Norway	101
France	138	Poland	116
Georgia	136	Romania	125
Germany	187	Slovenia	182
Hungary	124	Spain	278
Italy	190	Sweden	142
TOTAL	2398		

but by five denominated *plausible values* randomly obtained from the distribution function of test results derived from the answers in each test (Rasch, 1960, 1980), which can be interpreted as the representation of the ability range for each student (Mislevy *et al*, 1992; Wu and Adams, 2002). As baseline results, in our analysis we have only used one (the first) of the plausible values and, subsequently, we have used each of the other four values as a robustness test. Owing to space restrictions, we only present the estimations obtained with that first value, although the results of the empirical analysis are quite similar with the others.

The decision about which variables should be included as inputs is one of the main challenges of empirical studies using data from an international survey, since it usually includes an extensive list of potential indicators that can be considered. In this sense, most empirical papers attempting to measure efficiency of schools usually include some measure of human and capital resources (the interested reader can refer to Worthington, 2001 or Bradley *et al*, 2001 for early reviews or De Witte, 2014 for a recent update). In our case, we have decided to include the number of teachers in the model following the same criterion of many previous studies (eg Cordero *et al*, 2010; Brennan *et al*, 2014; Essid *et al*, 2014). Similarly, we include two variables that have been identified in the literature as main contributors to the attainment such as the instruction time (Lavy, 2010; Rivkin and Schiman, 2013) and the number of computers (Spiezia, 2010).

The inclusion of the characteristics of the students (usually represented by their socioeconomic status—SES—) among inputs is more discussed. Although the role of SES in explaining the students and schools' educational results is unquestionable (Haveman and Wolfe, 1995), thus it should be considered an input as in Agasisti (2014), Thieme *et al* (2013) or Crespo-Cebada *et al* (2014); however, some authors propose to include measures of SES as non-discretionary inputs (see Ruggiero and Vitaliano, 1999). In this paper, we have included this variable as a traditional input so that schools are evaluated according to their ability of making the most with their inputs (see Bradley *et al*, 2001 and Camanho *et al*, 2009, for details). Unfortunately, unlike PISA data set, PIRLS does not provide an

index representing SES; therefore, we have constructed it through principal component analysis (Jobson, 1992) as we describe below.

The following list summarizes our four input variables included in the empirical analysis based on the criterion of selecting variables representing different dimensions of school resources:

Number of teachers per a hundred students (TEACH100).

Instruction hours per week (INSTIME).

Number of computers per a hundred students (COMP100).

A composed index representing the socioeconomic status of pupils (SES). Following a similar criterion of the Economic Social and Cultural Status (ESCS) in PISA, we derived this index from data about the level of education of both parents and the number of books at home. As those variables were strongly correlated with each other, one factor was sufficient to explain most of the original variance (69%). Similarly, the original positive and negative values were rescaled to present positive values.

Regarding the variables representing the environment in which the school is operating, we take into account two different sources of heterogeneity: at school level and those specific country features. The first one includes a mixed set of six indicators with theoretical support in the literature. In particular, there are two continuous variables representing early literacy skills of students before entering the primary grades (EARLIT) (see Foy and Drucker, 2013, for details about its construction) and the average level of classroom disciplinary climate perceived by students attending the same school (DISCPL). There are also two ordered discrete variables that allow us to take into account the parental involvement at home (INVHOME) and in school (INVSCHL) and two (unordered) dummy variables regarding whether there is problem of absenteeism at the school (ABSENT) and whether the school is placed in a rural area (RURAL). These latter variables were rescaled to have a value equal to 1 for those conditions and equal to 2 otherwise.

The second group is composed of five continuous indicators about economic and cultural aspects collected at country level. The two economic variables are represented by the gross domestic product (GDP) per capita and the public expenditure per student in primary education as a percentage of GDP per capita for each country in the year 2011. On the other hand, as we mentioned previously, the source for information about cultural values in each country is the WVS. This data set provides information on individual socio-economic variables, attitudes and values regarding multiple aspects of life collected through a standardized survey. In particular we use information provided from a set of questions about which qualities are most valued when raising a child. Specifically, respondents are given a list of qualities (independence, hard work, feeling of responsibility, imagination, tolerance, thrift, perseverance, religious faith, unselfishness and obedience) that children can

Table 2 Descriptive statistics of variables included in the analysis

Variable	Type	Mean	Std. Dev.	Min	Max
PVREAD	Output	529.99	37.51	330.39	619.56
SES	Input	2.18	0.60	0.01	3.93
TEACH100	Input	4.80	1.22	1.08	18.18
COMP100	Input	5.53	4.35	0.07	73.05
INSTIME	Input	15.90	3.36	9.23	31.54
School factors					
EARLIT	Continuous	9.76	0.85	5.26	12.19
DISCPL	Continuous	10.12	1.65	3.65	13.16
INVHOME	Ordered discrete	3.32	0.52	1.00	4.00
INVSCHL	Ordered discrete	3.03	0.79	1.00	5.00
ABSENT	Unordered discrete	1.47	0.75	1.00	4.00
RURAL	Unordered discrete	1.40	0.49	1.00	2.00
Country features					
GDP pc	Continuous	31 486	21 301	3220	99 173
EXPEDUC	Continuous	21.84	4.46	13.21	31.1
HARDWORK	Continuous	0.48	0.28	0.07	0.90
RESP	Continuous	0.78	0.09	0.57	0.91
PERSEV	Continuous	0.41	0.10	0.26	0.56

be encouraged to learn at home and then asked to choose up to five that they think are most important.

According to Heckman (2011), there are 'Big Five' dimensions of personality skills (Conscientiousness, Openness to Experience, Extraversion, Agreeableness and Emotional Stability). Among them, factor 'Conscientiousness' can be defined as the tendency to be organized, responsible and hardworking. Heckman (2011) shows that this Conscientiousness factor is the most highly correlated with education outcomes (course grades and years of schooling). Borghans and Schils (2012) study the development of the performance of students during the test finding that Conscientiousness turns out to be associated with a smaller performance drop in correct responses between the beginning and the end of the test. For this reason, we have only selected answers for the three variables directly related to the conscientiousness factor: hard work (HARDWORK), responsibility (RESP) and perseverance (PERSEV). Table 2 reports the descriptive statistics for all these variables. In addition, Table 3 includes their mean values for each country in order to facilitate the interpretation of results shown in the next section.

Results

Table 4 reports the average estimated efficiency scores of schools operating in each country for the unconditional model as well as the decomposition between the school effect and the country effect using the metafrontier approach. Those scores

have been estimated using the robust order-*m* model with an output orientation, since we consider that schools are always attempting to maximize their attainment and cannot easily reduce their inputs, at least in the short term. The estimation of overall efficiency is obtained using the whole sample, whereas the decomposition between the school effect and country effect requires the estimation of 16 local frontiers (one for each country). To avoid the problem of bias that might arise because there is a dissimilarity in sample sizes across countries, we follow the same criterion established in De Witte and Marques (2009), which consists of setting *m* equal to the smallest local data set (ie Norway). Therefore, we set $m = 100$ to estimate all the local frontiers as well as the overall frontier; thus, every unit is compared with the same number of schools in all the estimations. For statistical inference, we use 200 bootstrap replications following the recommendation made by Daraio and Simar (2005, p 103).

In the last row of the table, we can observe that the average value of the overall efficiency for all the schools in the sample is 1.0759, which indicates that if all schools would perform as efficiently as the best school performers, the test scores could increase on average by 7% (or 10% if we only consider the inefficient units). Similarly, it is worth noting that some schools have a performance score below one. These super-efficient schools are performing better than the 100 schools they are benchmarked with. With regard to the average scores of units operating in different educational systems, the three top-listed countries are Finland, Bulgaria and the Netherlands, while the schools with the worst performance are operating in France, Slovenia, Norway and Georgia. Although there are some similarities with the ranking of countries according to their results shown in Table 3 (Finland and Georgia are placed at the top and the bottom, respectively, in both rankings), we can observe that the consideration of the inputs of educational process in the efficiency analysis makes some countries improve their position in the ranking. For instance, the incorporation of the socioeconomic status in the model benefits schools operating in Romania, since they have the lowest average levels. Something similar occurs with Bulgarian schools, which are placed at the second place in the ranking after considering their limited educational resources. In contrast, Germany occupies a lower position when educational inputs are taken into account since this country presents higher values than the average in most input variables.

If we focus on the decomposition of this overall efficiency, it is possible to find that, on average, the inefficiency levels detected can be attributed to a greater extent to the operating environment in the country (around 60%) than to the school context (around 40%). However, it is possible to find significant differences across countries. For instance, most part of the inefficiency in schools operating in Italy (86%) or Bulgaria (79%) depend on specific school factors, while in Norway, Georgia or Sweden the country effect is quite more relevant (85, 80 and 76%, respectively) to explain the levels of inefficiency of their schools.

Table 3 Mean values of all variables for each country (countries ranked by results in reading)

Countries	Output	Inputs			
	PVREAD	SES	TEACH100	COMP100	INSTIME
Finland	568.49	2.67	5.12	6.76	15.18
Czech Republic	548.57	2.23	4.75	7.07	15.16
Netherlands	544.34	2.14	4.15	5.41	20.71
Germany	541.95	2.29	4.74	6.29	16.54
Hungary	540.63	1.92	4.75	6.9	14.73
Italy	540.61	1.85	5.96	5.63	20.78
Sweden	537.61	2.51	4.72	4.37	16.26
Bulgaria	532.6	2.03	4.69	4.62	12.96
Lithuania	531.72	2.27	5.17	3.41	12.65
Poland	529.15	2.03	4.92	5.09	14.71
Slovenia	527.94	2.12	5.34	7.02	13.15
France	522.22	2.22	4.19	6.23	17.58
Spain	514.32	1.97	4.41	4.76	16.73
Norway	509.95	2.80	4.77	7.67	15.69
Romania	506.51	1.67	5.05	4.74	15.41
Georgia	488.07	2.62	4.72	4.11	14.06
Country avg.	529.99	2.18	4.84	5.59	15.90

Countries	Output	Environmental variables										
		School factors						Country factors				
	PVREAD	EARLIT	DISCPL	INVHOME	INVSCHL	Absent	Rural	GDP pc	EXPEDUC	Hardwork	RESP	PERSEV
Finland	568.49	10.21	10.15	3.01	3.16	1.37	1.46	48 678	21.1	0.15	0.89	0.55
Czech Republic	548.57	9.78	10.22	3.49	2.47	1.22	1.42	20 580	15.6	0.82	0.67	0.39
Netherlands	544.34	9.26	9.1	2.83	3.39	1.54	1.39	49 886	18.7	0.29	0.90	0.37
Germany	541.95	9.19	9.6	3.01	3.43	1.58	1.48	44 355	18.3	0.19	0.88	0.52
Hungary	540.63	8.82	9.86	3.32	3.04	1.61	1.44	13 964	22.51	0.30	0.58	0.31
Italy	540.61	9.34	9.69	3.24	3.22	1.71	1.65	36 180	24.1	0.39	0.87	0.44
Sweden	537.61	10.33	9.7	3.27	2.88	1.56	1.37	56 755	27.6	0.07	0.89	0.37
Bulgaria	532.6	9.81	10.62	3.19	2.76	1.29	1.39	7287	23.2	0.90	0.68	0.50
Lithuania	531.72	10.18	10.58	3.26	3.26	1.27	1.26	14 158	24.3	0.89	0.74	0.36
Poland	529.15	10.05	9.74	3.54	3.09	1.16	1.4	13 382	27.4	0.18	0.80	0.26
Slovenia	527.94	9.29	10.12	3.21	2.82	1.31	1.37	24 429	30.98	0.34	0.72	0.56
France	522.22	10.23	10.38	3.8	3.17	1.54	1.57	42 560	18.7	0.62	0.78	0.54
Spain	514.32	10.86	10.62	3.62	2.94	1.43	1.13	31 118	21.5	0.61	0.73	0.26
Norway	509.95	9.14	9.92	3.6	3.07	1.51	1.45	99 173	21.1	0.12	0.91	0.39
Romania	506.51	9.28	10.4	3.44	2.85	1.66	1.54	9064	19.6	0.75	0.74	0.44
Georgia	488.07	9.64	10.84	3.17	2.97	1.73	1.33	3220	13.2	0.85	0.73	0.32
Country avg.	529.99	9.76	10.12	3.32	3.03	1.47	1.4	31 486	21.84	0.48	0.78	0.41

Table 4   Decomposition of overall efficiency between school and country effect for each country

Countries	Overall efficiency				School effect				
	Mean	ST	Min	Max	Mean	%	ST	Min	Max
Finland	1.0432	0.0342	0.9764	1.1358	1.0208	48.11	0.0298	1.0000	1.1059
Bulgaria	1.0442	0.0772	0.9921	1.4440	1.0347	78.64	0.0735	1.0000	1.4439
Netherlands	1.0450	0.0478	0.9891	1.2111	1.0182	40.46	0.0289	0.9995	1.1551
Hungary	1.0456	0.0584	0.9782	1.2729	1.0160	35.18	0.0297	0.9995	1.1518
Lithuania	1.0476	0.0525	0.9996	1.2159	1.0228	47.86	0.0353	0.9998	1.1713
Czech Republic	1.0510	0.0480	0.9931	1.2500	1.0187	36.62	0.0285	0.9953	1.1164
Italy	1.0611	0.0685	0.9512	1.5774	1.0526	86.09	0.0594	0.9996	1.2119
Poland	1.0642	0.0614	0.9941	1.2935	1.0174	27.15	0.0330	0.9926	1.1689
Germany	1.0685	0.0577	0.9867	1.3744	1.0281	41.01	0.0412	0.9980	1.2178
Romania	1.0771	0.0925	0.9961	1.4585	1.0239	30.97	0.0481	1.0000	1.2447
Spain	1.0822	0.0725	0.9983	1.3822	1.0465	56.57	0.0563	0.9997	1.2937
Sweden	1.0824	0.0588	0.9968	1.3170	1.0190	23.01	0.0300	0.9969	1.1282
France	1.0871	0.0583	0.9909	1.2814	1.0335	38.48	0.0411	0.9993	1.1939
Slovenia	1.0970	0.0649	0.9944	1.3260	1.0335	34.48	0.0408	0.9842	1.2402
Norway	1.1606	0.0586	1.0000	1.3080	1.0207	12.91	0.0310	0.9994	1.1304
Georgia	1.1684	0.1106	1.0000	1.4840	1.0285	16.94	0.0448	0.9984	1.1946
Total	1.0759	0.0747	0.9512	1.5774	1.0293	38.57	0.0455	0.9842	1.4439

Countries	Country effect				
	Mean	%	ST	Min	Max
Finland	1.0222	51.46	0.0287	0.9764	1.1358
Bulgaria	1.0092	20.76	0.0244	0.9762	1.1473
Netherlands	1.0262	58.28	0.0343	0.9891	1.1917
Hungary	1.0293	64.16	0.0528	0.9782	1.2729
Lithuania	1.0243	51.00	0.0378	0.9986	1.2159
Czech Republic	1.0316	61.99	0.0338	0.9931	1.1335
Italy	1.0089	14.59	0.0544	0.9514	1.5774
Poland	1.0459	71.48	0.0484	0.9986	1.2319
Germany	1.0397	57.91	0.0479	0.9867	1.3744
Romania	1.0516	66.88	0.0689	0.9712	1.3054
Spain	1.0340	41.40	0.0395	0.9974	1.2385
Sweden	1.0625	75.76	0.0527	0.9968	1.3170
France	1.0519	59.63	0.0409	0.9894	1.2119
Slovenia	1.0613	63.18	0.0414	0.9945	1.2132
Norway	1.1371	85.33	0.0464	1.0000	1.2585
Georgia	1.1355	80.47	0.0903	1.0000	1.4546
Total	1.0455	59.87	0.0590	0.9514	1.5774

The main problem of this initial assessment is that they are based on the assumption that all the evaluated schools are operating in the same environment, so the estimated performance scores may not adequately represent their level of efficiency. Therefore, the next step consists of considering the existing heterogeneity among schools in our estimation of their efficiency scores. For that purpose, two alternative conditional efficiency models have been developed. These models include stepwise additional information about variables representing the environment as in Haelermans and De Witte (2012). In model 1 we only consider a mixed set of continuous and discrete variables related to the school environment (location,

students' early skills, parental involvement, absenteeism and disciplinary levels). Subsequently, we estimate a second conditional model (model 2) including an additional set of economic and cultural values associated with each country. The results obtained for both models are reported in Table 5, where we also distinguish the average efficiency of schools across countries.

Once we include information about environmental variables in the analysis the average efficiency decreases in both models (1.0536 for model 1 and 1.0315 for model 2). This is intuitive since the consideration of additional variables in the analysis implies that the reference group only includes schools with more similar characteristics. However, the most interesting

Table 5 Efficiency score distribution across countries in different models

Unconditional			*Conditional model 1*			*Conditional model 2*		
	Mean	*ST*		*Mean*	*ST*		*Mean*	*ST*
Finland	1.0432	0.0342	Hungary	1.0257	0.0417	Hungary	1.0166	0.0318
Bulgaria	1.0442	0.0772	Bulgaria	1.0268	0.0655	Finland	1.0194	0.0261
Netherlands	1.0450	0.0478	Netherlands	1.0289	0.0375	Czech Republic	1.0198	0.0276
Hungary	1.0456	0.0584	Czech Republic	1.0329	0.0364	Netherlands	1.0199	0.0306
Lithuania	1.0476	0.0525	Finland	1.0331	0.0319	Poland	1.0211	0.0362
Czech Republic	1.0510	0.0480	Lithuania	1.0347	0.0450	Sweden	1.0220	0.0311
Italy	1.0611	0.0685	Italy	1.0405	0.0627	Bulgaria	1.0240	0.0621
Poland	1.0642	0.0614	Poland	1.0442	0.0507	Germany	1.0268	0.0351
Germany	1.0685	0.0577	Romania	1.0451	0.0751	Lithuania	1.0293	0.0434
Romania	1.0771	0.0925	Spain	1.0513	0.0610	Norway	1.0302	0.0464
Spain	1.0822	0.0725	Germany	1.0523	0.0491	Slovenia	1.0345	0.0421
Sweden	1.0824	0.0588	Sweden	1.0594	0.0504	France	1.0352	0.0427
France	1.0871	0.0583	France	1.0626	0.0507	Romania	1.0371	0.0721
Slovenia	1.0970	0.0649	Slovenia	1.0665	0.0629	Spain	1.0424	0.0536
Norway	1.1606	0.0586	Georgia	1.1327	0.1057	Italy	1.0430	0.0543
Georgia	1.1684	0.1106	Norway	1.1399	0.0643	Georgia	1.0657	0.0800
Total	1.0759	0.0747	*Total*	1.0536	0.0656	*Total*	1.0315	0.0489

conclusions can be drawn by exploring the distribution of the efficiency scores across countries in both rankings. In this sense, the ranking of countries according to their average results with model 1 does not suffer significant alterations with respect to the unconditional model. Actually, the Spearman correlation coefficient between the two distributions of schools is quite high (0.849). The only remarkable change in the ranking of countries is represented by Hungary, which is considered the top performer within this model, while Finland descends from the top seed in the unconditional model to the fifth position in this model.

Nevertheless, the consideration of country features in model 2 modifies the picture to a large extent (Spearman correlation coefficient with the unconditional case is now 0.658). In particular, schools operating in Norway, Sweden and Georgia, where the country effect was very relevant according to the information showed in Table 4, experienced a substantial improvement in their average level of inefficiency, although the latter still maintains at the bottom of the ranking. In contrast, other countries in which the specific country features were not too relevant such as Italy or Bulgaria worsen their relative positions with respect to other countries. The importance of those changes leads us to presume that the heterogeneity among different countries is more relevant than heterogeneity among schools within the same country.

In order to examine the influence of those external factors on efficiency estimates, we regress the ratio between the conditional and the unconditional efficiency scores on the environmental variables using the local linear estimator described in the methodology section. As we have two different models (model 1 and model 2), we have also carried out two different estimations. Table 6 presents the median influence of these

variables and the *p*-values of the significance tests proposed by Li and Racine (2004) and Racine and Li (2004) after performing 1000 bootstrap samples. Moreover, we indicate whether a variable has a favourable or unfavourable correlation with efficiency according to the visualization of the partial regression scatter plots.

In model 1, almost all the external school variables seem to have a significant effect on explaining the level of inefficiency. In particular, we detect a negative influence of being placed in a rural area, which is usually related to the lack of competitive pressure rather than the location itself (Alchian, 1950). Similarly, the parental involvement at home and school has a positive and significant effect on performance as pointed out by multiple previous literature focused on the determinants of students' attainment (see Wilder, 2014, for an extensive review on this topic). Finally, variables representing school environment, such as the level of disciplinary climate in the classroom or the existence of a problem of absenteeism in the school, are also significantly related (positive and negative, respectively) to school efficiency. In contrast, the early literacy skills of students before entering the school is not considered as a relevant factor to explain school efficiency probably because its effect was captured by the SES input.

The results of model 2, in which we account for heterogeneity across countries, reveal that all variables representing specific country features of countries have a significant and positive influence on efficiency, whereas the influence of school variables becomes insignificant for all the considered indicators. Therefore, we can conclude that country factors seem to have a more relevant role in explaining differences in the performance of schools because there is more heterogeneity across countries than among schools. One potential explanation for this result is

Table 6 Influence of different factors on educational performance (estimation of nonparametric significance tests)

School variables	Model 1		Model 2	
	p-value	Influence (scatter plot)	p-value	Influence (scatter plot)
Disciplinary index	0.01*	Favourable	0.99	Favourable
Earlit	0.95	Favourable	0.99	Favourable
Parents' involvement at home	0.00***	Favourable	0.99	Favourable
Parents' involvement in school	0.01*	Favourable	0.99	Favourable
Rural area	0.00***	Unfavourable	0.99	Unfavourable
Absenteeism	0.00***	Unfavourable	0.99	Unfavourable
Country variables			p-value	Influence (scatter plot)
GDP pc			0.00***	Favourable
Expenditure in education			0.00***	Favourable
Hard work			0.00***	Favourable
Responsibility			0.00***	Favourable
Perseverance			0.00***	Favourable

***denotes statistical significance at 1%.
*denotes statistical significance at 10%.

that our sample only includes European countries, where the school context is relatively similar.

This finding about the scarce influence of specific school factors is in line with some previous results in the literature for a specific country (eg, Haelermans and De Witte, 2012, for the Netherlands) as well as the important role played by economic indicators to explain differences in efficiency performance across countries (eg, Afonso and St Aubyn, 2006). Nevertheless, the importance of cultural or personality factors as determinants of educational performance has only been considered as a relevant factor in some recent cross-country studies (eg, Borghans *et al*, 2008; Borghans and Schils 2012; Conti *et al*, 2011). Our contribution here is that coinciding with previous results we found that the conscientiousness factor also has a significant influence to explaining differences in efficiency estimates. From this result it is worth to remark that more attention should be devoted to non-cognitive skills in order to boost efficiency in education.

Concluding remarks

In this paper we have combined the use of two totally nonparametric methods to assess the performance of primary schools in 16 European countries using data from PIRLS 2011. Specifically, we adapt the metafrontier approach developed by Battese and Rao (2002), Battese *et al* (2004) and O'Donnell *et al* (2008) to our context with the aim of decomposing which

part of the estimated inefficiency can be attributed to the school and the country. Subsequently, we apply the robust conditional model developed by Daraio and Simar (2005, 2007a, b) and extended by De Witte and Kortelainen (2013) to account for two different sets of variables related to the school and the country environment. Similarly, nonparametric bootstraps based on significance tests have been applied to examine the statistical influence of those factors on efficiency scores.

Our results show that the ranking of countries based only on academic results modifies in a certain extent when we account for data about school inputs participating in the educational process in our efficiency analysis. As a result, schools in which the average socioeconomic status of the students is relatively low improve their relative position, while schools having better resource endowments descend in the ranking when this information is taken into account. Similarly, the decomposition of inefficiency between the school and country effect allows us to detect significant divergences across countries with regard to which are the main explanatory factors for their inefficient performance. In this sense, the results indicate that heterogeneity across countries is more relevant than among schools. In particular, we would like to highlight the significant and positive effect of cultural non-cognitive values, since those factors have been scarcely studied so far in the efficiency literature. However, the approach used in this paper does not allow for a causal interpretation of the results but it allows pointing out to a future line of research based on the search of the causes of inefficiency.

These findings provide some interesting insights into the analysis of determinants of educational attainment using a cross-country approach. However, more research will be needed in future to explore more in depth the results discussed here. First, it would be interesting to replicate this type of analysis using data about secondary schools, since the performance of students in those levels might also be affected by the existence of school tracking or external exams—accountability—in some countries (see Woessman *et al*, 2009 or Hanushek and Woessmann, 2014, for details). Those types of policies have not been considered in this study because students evaluated in PIRLS are enrolled in the fourth grade of primary school; thus, basically they are not affected by them.

Second, the analysis of divergences between public and private schools remains as an appealing field for future research using the model proposed in this paper. Unfortunately, in our empirical analysis we could not deal with this aspect, since this information is not available in PIRLS for the majority of countries. Similarly, the proposed analysis could be replicated using data at pupil level. Actually, these data have been used in some recent studies to assess the performance of students in a specific country (eg Perelman and Santín, 2011; De Witte and Kortelainen, 2013). However, it must be taken into account that the estimation of the kernel bandwidths and the efficiency scores with conditional models can take a great amount of time (even for months) given the huge sample sizes of international large-scale data sets.

References

Afonso A and 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 (2014). The efficiency of public spending on education: An empirical comparison of EU countries. *European Journal of Education* **49**(4): 543–557.

Alchian AA (1950). Uncertainty, evolution and economic theory. *Journal of Political Economy* **58**(3): 211–221.

Aristovnik A and Obadić A (2014). Measuring relative efficiency of secondary education in selected EU and OECD countries: The case of Slovenia and Croatia. *Technological and Economic Development of Economy* **20**(3): 419–433.

Badin L, Daraio C and Simar L (2010). Optimal bandwidth selection for conditional efficiency measures: A data-driven approach. *European Journal of Operational Research* **201**(2): 633–640.

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

Battese G and Rao D (2002). Technology gap, efficiency, and a stochastic metafrontier function. *International Journal of Business* **1**(2): 87–93.

Battese G, Rao D and O'Donnell C (2004). A metafrontier production function for estimation of technical efficiencies and technology gaps for firms operating under different technologies. *Journal of Productivity Analysis* **21**(1): 91–103.

Bishop JH (1997). The effect of national standards and curriculum-based exams on achievement. *The American Economic Review* **87**(2): 260–264.

Borghans L, Duckworth AL, Heckman JJ and Ter Weel B (2008). The economics and psychology of personality traits. *Journal of Human Resources* **43**(4): 972–1059.

Borghans L and Schils T (2012). The leaning tower of PISA: The effect of test motivation on scores in the international student assessment. Mimeo. http://www.sole-jole.org/13260.pdf, accessed 8 May 2015;

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

Bray M and Thomas RM (1995). Levels of comparison in educational studies: Different insights from different literatures and the value of multilevel analyses. *Harvard Educational Review* **65**(3): 472–490.

Brennan S, Haelermans C and Ruggiero J (2014). Nonparametric estimation of education productivity incorporating nondiscretionary inputs with an application to Dutch schools. *European Journal of Operational Research* **234**(3): 809–818.

Brunello G and Checchi D (2007). Does school tracking affect equality of opportunity? New international evidence. *Economic Policy* **22**(52): 781–861.

Camanho A, Portela MC and Vaz C (2009). Efficiency analysis accounting for non-discretionary factors: The assessment of retailing stores using data envelopment analysis. *Computers and Operations Research* **36**: 1591–1601.

Cazals C, Florens J and Simar L (2002). Nonparametric frontier estimation: A robust approach. *Journal of Econometrics* **106**: 1–25.

Cherchye L, De Witte K, Ooghe E and Nicaise I (2010). Efficiency and equity in private and public education: A nonparametric comparison. *European Journal of Operational Research* **202**(2): 563–573.

Coco G and Lagravinese R (2014). Cronyism and education performance. *Economic Modelling* **38**: 443–450.

Conti G, Heckman JJ and Urzúa S (2011). *Early endowments, education and health. Human capital and economic opportunity: A global working group*. Working Paper 2011-2001.

Cordero JM, Pedraja F and Santín D (2010). Enhancing the inclusion of non-discretionary inputs in DEA. *Journal of Operational Research Society* **61**(4): 574–584.

Crespo-Cebada E, Pedraja F and Santín D (2014). Does school ownership matter? An unbiased efficiency comparison for regions of Spain. *Journal of Productivity Analysis* **41**(1): 153–172.

Daraio C and Simar L (2005). Introducing environmental variables in nonparametric frontier models: A probabilistic approach. *Journal of Productivity Analysis* **24**(1): 93–121.

Daraio C and Simar L (2007a). *Advanced Robust and Nonparametric Methods in Efficiency Analysis. Methodology and Applications*. Springer: New York.

Daraio C and Simar L (2007b). Conditional nonparametric frontier models for convex and non convex technologies: A unifying approach. *Journal of Productivity Analysis* **28**(1): 13–32.

De Jorge J and Santín D (2010). Determinantes de la eficiencia educativa en la Unión Europea. *Hacienda Publica Española* **193**(2): 131–155.

Deprins D, Simar L and Tulkens H (1984). Measuring labor inefficiency in post offices. In: Marchand P, Pestieau P and Tulkens H (eds). *Concepts and Measurements*. Amsterdam: North Holland, pp 243–267.

Deutsch J, Dumas A and Silber J (2013). Estimating an educational production function for five countries of Latin America on the basis of the PISA data. *Economics of Education Review* **36**: 245–262.

De Witte K (2014). Efficiency in education. A review of literature and a way forward. paper presented at Workshop on Efficiency in Education, London.

De Witte K and Kortelainen M (2013). What explains performance of students in a heterogeneous environment? Conditional efficiency estimation with continuous and discrete environmental variables. *Applied Economics* **45**(17): 2401–2412.

De Witte K and Marques R (2009). Capturing the environment, a metafrontier approach to the drinking water sector. *International Transactions in Operational Research* **16**(2): 257–271.

De Witte K and Rogge N (2011). Accounting for exogenous influences in performance evaluations of teachers. *Economics of Education Review* **30**(4): 641–653.

De Witte K, Thanassoulis E, Simpson G, Battisti G and Charlesworth-May A (2010). Assessing pupil and school performance by nonparametric and parametric techniques. *Journal of the Operational Research Society* **61**(8): 1224–1237.

Dronkers J and 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.

Essid H, Ouellette P and Vigeant S (2014). Productivity, efficiency, and technical change of Tunisian schools: A bootstrapped Malmquist approach with quasi-fixed inputs. *Omega* **42**(1): 88–97.

Farrell MJ (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A* **120**(3): 253–290.

Foy P and Drucker KT (2013). *PIRLS 2011 User Guide for the International Database*. TIMSS and PIRLS International Study Center, Boston College: Chestnut Hill, MA.

Fuchs T and Woessmann L (2007). What accounts for international differences in student performance? *Empirical Economics* **32**(2–3): 433–464.

Giambona F, Vassallo E and Vassiliadis E (2011). Educational systems efficiency in European Union countries. *Studies in Educational Evaluation* **37**(2): 108–122.

Giménez V, Prior D and Thieme C (2007). Technical efficiency, managerial efficiency and objective-setting in the educational system: An international comparison. *Journal of the Operational Research Society* **58**(8): 996–1007.

Grosskopf S, Hayes KJ and Taylor LL (2014). Efficiency in education: Research and implications. *Applied Economic Perspectives and Policy* **36**(2): 175–210.

Haelermans C and De Witte K (2012). The role of innovations in secondary school performance: Evidence from a conditional efficiency model. *European Journal of Operational Research* **223**(2): 541–549.

Hanushek EA (1979). Conceptual and empirical issues in the estimation of educational production functions. *Journal of Human Resources* **14**(3): 351–388.

Hanushek EA and Kimko DD (2000). Schooling, labor-force quality, and the growth of nations. *The American Economic Review* **90**(5): 1184–1208.

Hanushek EA and Woessmann L (2008). The role of cognitive skills in economic development. *Journal of Economic Literature* **46**(3): 607–668.

Hanushek EA and Woessman L (2011). The economics of international differences in educational achievement. In: Hanushek EA, Machin S and Woessmann L (eds). *Handbook of the Economics of Education*. Vol. 3. North Holland: Amsterdam, pp 89–200.

Hanushek EA, Link S and Woessmann L (2013). Does school autonomy make sense everywhere? Panel estimates from PISA. *Journal of Development Economics* **104**: 212–232.

Hanushek EA and Woessmann L (2014). Institutional structures of the education system and student achievement: A review of cross-country economic research. In: Strietholt R, Bos W, Gustafsson JE and Rosen M (eds). *Educational Policy Evaluation through International Comparative Assessments*. Waxmann Verlag: Münster: pp 145–176.

Haveman R and Wolfe B (1995). The determinants of children's attainments: A review of methods and findings. *Journal of Economic Literature:* **33**(4): 1829–1878.

Heckman JJ (2011). *Integrating personality psychology into economics* (N° w17378). National Bureau of Economic Research.

Jeong SO, Park BU and Simar L (2010). Nonparametric conditional efficiency measures: Asymptotic properties. *Annals of Operations Research* **173**(1): 105–122.

Jobson JD (ed). (1992). Principal components factors and correspondence analysis. In: *Texts in Statistics. Applied Multivariate Data Analysis*. Vol. II. Categorical and Multivariate Methods. New York, Springer-Verlag, pp 345–482.

Krueger AB and Lindahl M (2001). Education for growth: Why and for whom. *Journal of Economic Literature* **39**(4): 1101–1136.

Lavy V (2010). *Do differences in school's instruction time explain international achievement gaps in Maths, Science and Language? Evidence from developed and developing countries*. CEE DP 118. Centre for the Economics of Education (NJ1).

Levin H (1974). Measuring efficiency in educational production. *Public Finance Quarterly* **2**(1): 3–24.

Li Q and Racine JS (2004). Cross-validated local linear nonparametric regression. *Statistica Sinica* **14**(2): 485–512.

Mislevy RJ, Beaton AE, Kaplan B and Sheehan KM (1992). Estimating population characteristics form sparse matrix samples of item responses. *Journal of Educational Measurement* **29**(2): 133–161.

Mullis IVS, Martin MO, Foy P and Drucker KT (2012). *PIRLS 2011 International Results in Reading*. TIMSS and PIRLS International Study Center, Boston College: Chestnut Hill, MA.

O'Donnell C, Rao D and Battese G (2008). Metafrontier frameworks for the study of firm-level efficiencies and technology ratios. *Empirical Economics* **37**(2): 231–255.

Perelman S and Santín D (2011). Measuring educational efficiency at student level with parametric stochastic distance functions: An application to Spanish PISA results. *Education Economics* **19**(1): 29–49.

Racine JS (1997). Consistent significance testing for nonparametric regression. *Journal of Business and Economic Statistics* **15**(3): 369–379.

Racine JS and Li Q (2004). Nonparametric estimation of regression functions with both categorical and continuous data. *Journal of Econometrics* **119**(1): 99–130.

Racine JS, Hart JD and Li Q (2006). Testing the significance of categorical predictor variables in nonparametric regression models. *Econometric Reviews* **25**(4): 523–544.

Rasch G (1960/1980). *Probabilistic Models for Some Intelligence and Attainment Tests*. Danish Institute for Educational Research: Copenhagen: Expanded edition (1980). The University of Chicago Press.

Rivkin SG and Schiman JC (2013). *Instruction time, classroom quality, and academic achievement* (No. w19464). National Bureau of Economic Research.

Ruggiero J (1996). On the measurement of technical efficiency in the public sector. *European Journal of Operational Research* **90**(3): 553–565.

Ruggiero J (2000). Nonparametric estimation of returns to scale in the public sector with an application to the provision of educational services. *Journal of the Operational Research Society* **51**(8): 906–912.

Ruggiero J and Vitaliano DF (1999). Assessing the efficiency of public schools using data envelopment analysis and frontier regression. *Contemporary Economic Policy* **17**(3): 321–331.

Schuetz G, Ursprung HW and Wößmann L (2008). Education policy and equality of opportunity. *Kyklos* **61**(2): 279–308.

Silva-Portela MC and Thanassoulis E (2001). Decomposing school and school-type efficiency. *European Journal of Operational Research* **132**(2): 357–373.

Spiezia V (2010). Does computer use increase educational achievements? Student-level evidence from PISA. *OECD Journal: Economic Studies* **1**: 1–22.

Strietholt R, Gustafsson JE, Rosen M and Bos W (2014). Outcomes and causal inference in international comparative assessments. In: Strietholt, R., Bos, W., Gustafsson, J. E., & Rosén, M. (Eds.). Educational Policy Evaluation through International Comparative Assessments. Waxmann Verlag. Münster, 9-19.

Täht K and Must O (2013). Comparability of educational achievement and learning attitudes across nations. *Educational Research and Evaluation* **19**(1): 19–38.

Thanassoulis E and Silva-Portela A (2002). School outcomes: Sharing the responsibility between pupil and school. *Education Economics* **10**(2): 183–207.

Thieme C, Giménez V and Prior D (2012). A comparative analysis of the efficiency of national education systems. *Asia Pacific Education Review* **13**(1): 1–15.

Thieme C, Prior D and Tortosa-Ausina E (2013). A multilevel decomposition of school performance using robust nonparametric frontier techniques. *Economics of Education Review* **32**: 104–121.

Vandenberghe V and Robin S (2004). Evaluating the effectiveness of private education across countries: A comparison of methods. *Labour Economics* **11**(4): 487–506.

Verhoeven M, Gunnarsson V and Carcillo S (2007). *Education and health in G7 countries: Achieving better outcomes with less spending* (No. 2007-2263). International Monetary Fund.

Wilder S (2014). Effects of parental involvement on academic achievement: A meta-synthesis. *Educational Review* **66**(3): 377–397.

Wilson PW (2005). *Efficiency in education production among PISA countries, with emphasis on transitioning economies*. World Bank Working Paper.

Woessmann L, Luedemann E, Schuetz G and West MR (2009). *School Accountability, Autonomy, and Choice Around the World*. Edward Elgar: Cheltenham, UK.

Worthington AC (2001). An empirical survey of frontier efficiency measurement techniques in education. *Education Economics* **9**(3): 245–268.

Wu M and Adams RJ (2002). *Plausible Values—Why They are Important*. International Objective Measurement Workshop: New Orleans.

Zhao Y, Zhang G, Yang W, Kirkland D, Han X and Zhang J (2008). A comparative study of educational research in China and the United States. *Asia Pacific Journal of Education* **28**(1): 1–17.