Assessing European primary school performance through a conditional nonparametric model

José Manuel Cordero Ferrera[a*]

Carmen Mª García Gómez[a]

Daniel Santín González[b]

Rosa Simancas Rodríguez[a]

[a]Department of Economics, University of Extremadura, Spain
[b]Department of Applied Economics VI, Complutense University of Madrid, Spain

Abstract

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. Likewise, 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 like responsibility or perseverance are significantly related to school efficiency, whereas most school factors do not seem to have a significant influence on school performance.

Key words: Education, Cross-country analysis, Nonparametric, Conditional Efficiency

[*] *Corresponding author at: Departamento de Economía, Universidad de Extremadura, Av. Elvas s/n, 06071 Badajoz, Spain*
E-mail address: *jmcordero@unex.es; Tlf. +34 924289300, Fax: +34 924272509*

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 Woessman, 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 analyze 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 Cecchi, 2007; Schuetz et al., 2008) or the influence of accountability (Bishop, 1997; Fuchs and Woessman, 2007) to cite only some examples.

Since their main focus is the identification of significant relationship 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 Woessman, 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 lead 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 datasets 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 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 Obadic, 2014) or TIMSS (Gimenez 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 (e.g. 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 which allow us to overcome some of the main limitations of previous studies. Likewise, 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.

From a methodological perspective, the contributions of the paper are four-fold. First, we use the robust order-*m* methodology described by Cazals et al, (2002) to avoid some of the main drawbacks of the nonparametric methods. This approach consists of constructing a partial frontier using only part of the sample (*m* observations) to determine efficiency scores. Thus, it mitigates the impact of outliers and potential errors in data and the bias that can arise when the evaluated units (schools) are grouped into units (countries) of different size (Zhang and Bartels, 1998). Second, we adapt the metafrontier framework developed by Battese and Rao (2002), Battese et al (2004) and O'Donnell et al (2008) to the context of our study. This approach allows us to decompose the estimated inefficiency between two different levels (school and country). Third, we use the conditional nonparametric approach proposed by Dario and Simar (2005, 2007a, 2007b) to incorporate the effect of environmental factors at school and country level into the estimation of efficiency scores. Thus, the efficiency score assigned to each school truly reflect the portion of the production process for which that unit is responsible. Four, this method allows us to explore the potential influence of multiple factors at different levels (school and country) without assuming the restrictive separability condition of two-stage approaches in order to provide meaningful results.

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. Likewise, 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 analyze 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 dataset 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 centers (schools and high schools), thus part of the school effect is not caught 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 year old in any country (see Brunello and Cecchi, 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; Thät 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. Section 2 describes the methodology. Section 3 explains the main characteristics of the data and the variables selected for the empirical analysis. Section 4 presents the main results and relates them to the existing literature. Finally, the paper ends with some concluding remarks in Section 5.

METHODOLOGY

The 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. This can be represented by Equation (1):

$$\psi = \{(x, y) \in \Re_+^{p+q} \mid x \text{ can produce } y\} \qquad (1)$$

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 nonpatrametric Free Disposal Hull (FDH) methodology (Deprins et al, 1984), which is well-suited to this setting because it does not require any *a priori* assumption on the functional form of the production process. Although DEA is more popular among practitioners using nonparametric techniques, in our study we opt for using FDH because it has comparatively superior asymptotic properties (Park et al., 2000; Simar and Wilson, 2000), it does not require assuming convexity and for this reason ensure that all reference units are real. The output oriented efficiency score ($\hat{\lambda}_{FDH}$) of an observation can be obtained by solving the mixed integer linear programming problem in Equation (2):

$$\hat{\lambda}_{FDH} = \max\left\{ \lambda \,\middle|\, \lambda y \leq \sum_{i=1}^{N} \gamma_i y_i ; x \geq \sum_{i=1}^{N} \gamma_i x_i ; \sum_{i=1}^{N} \gamma_i = 1; \gamma_i \in \{0,1\}; i = 1,....,n \right\} \qquad (2)$$

where $\hat{\lambda}_{FDH} = 1$ denotes an efficient school, while $\hat{\lambda}_{FDH} > 1$ implies that the school is inefficient. However, this nonparametric approach presents some significant shortcomings that should be born in mind if it is used to estimate efficiency measures of school performance. Firstly, statistical inference is not possible due to its deterministic nature. Secondly, it is very sensitive to the presence of outliers and measurement errors in data. Finally, it experiences dimensionality problems due to their slow convergence rates. In the next sections, we explain some approaches that can be used in order to overcome these limitations.

The robust FDH model

The first attempts to improve the robustness of nonparametric methods were the works of Kneipp et al (1998) and Simar and Wilson (2000). Subsequently, Cazals et al (2002) introduced the robust order-*m* estimation. This approach is related to the FDH estimator, but instead of constructing a full frontier, it creates a partial frontier that envelops only *m* (≥1) observations randomly drawn from the empirical sample. This procedure is repeated *B* times resulting in multiple measures ($\hat{\lambda}_{mi}^{1},...,\hat{\lambda}_{mi}^{B}$) 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:

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

where the ρ-dimensional random variables $x_i,\ldots,x_m$ are drawn randomly and repeatedly from the conditional distribution of X given $y_i \geq y$. This estimator allows us to compare the efficiency of an observation with that of m potential units that have a production larger or equal to y. As it does not include all the observations, it is less sensitive to outliers, extreme values or noise in the data. Moreover, Cazals et al (2002) show that the convergence rate of this order-m estimator is comparable to parametric estimators, thus this estimator avoids the curse of dimensionality problem. As m increases, the expected order-m estimator tends to the FDH efficiency score ($\hat{\lambda}_{FDH}$). For acceptable m values, normally the efficiency scores will present values higher than unity, which indicates that students 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 average m observations in its reference sample (Daraio and Simar, 2007a). This is not possible in the traditional nonparametric framework where by construction $\hat{\lambda} \geq 1$.

Moreover, this approach allows us to avoid the problem of bias that can arise when we compare groups of units on a different size since the mean level of efficiency generally depends on the existing number of schools in each country (Zhang and Bartels, 1998). This problem can be reduced by using the same m parameter for every country, which implies assuming that the performance of every unit is compared to the same number of units independently of the number of schools included in the sample for each country. In our case, we determine the value of m that equals the size of the smallest number of schools in the dataset, since it fits better in the metafrontier framework (see below).

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 ideas developed by Silva-Portela and Thanassoulis (2001) and Thanassoulis and Silva-Portela (2002) to decompose the effect of school from students' inefficiency. Therefore, the extension to 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:

$$\lambda^k(x_k, y_k) = \inf\left\{\theta^k \middle| (\theta^k x_k, y_k) \in \psi^k\right\} \quad (4)$$

where the technology set for group k is defined as

$$\psi^k = \left\{(x_k, y_k) \in \Re_+^{p+q} \;\middle|\; x_k \text{ can produce } y_k\right\} \quad (5)$$

If all the countries have the same characteristics, all the observations can be pooled and schools can be evaluated relative to the same standards. Thus, the metafrontier can be represented by the technology set defined by:

$$\psi = \left\{(x, y) \in \Re_+^{p+q} \;\middle|\; x \text{ can produce } y\right\} \quad (6)$$

Therefore we have two different frontiers: 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 school inefficiency can be defined by the ratio between the local potential output divided by the actual output ($SCE = \alpha' = y'/y_c$). When this school is compared with the metafrontier, the overall efficiency (*OE*) can be defined as $OE = \alpha'' = y''/y_c$. 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: $OE = SCE \times CNE$.

(Figure 1 around here)

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 cases. 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, 2007b), which assumes that both types of factors can have a direct influence on the shape of the best practice frontier (i.e., this model does not assume a separability condition). Therefore, efficiency estimates are determined by both the inputs, outputs and exogenous variables. Using a probabilistic formulation, this conditional function can be defined as:

$$H_{XY|Z}(x, y|z) = \Pr(X \leq x, Y \geq y|Z = z) \tag{7}$$

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*. This can also be decomposed into:

$$\begin{aligned} H_{XY|Z}(x,y|z) &= \Pr(Y \geq y | x \leq x, Z = z) \Pr(X \leq x, Z = z) \\ &= S_{Y|X,Z}(Y \geq y | X \leq x, Z = z) F_X(X \leq x; Z = z) \\ &= S_Y(y|x,z) F_X(x|z) \end{aligned} \tag{8}$$

Therefore, the output efficiency measure can be analogously defined as:

$$\lambda(x,y|z) = \sup\{\lambda > 0 | S_{Y|XZ}(\lambda y | X \leq x, Z = z) > 0\} \tag{9}$$

The conditional order-*m* efficiency measure can be analogously defined using the expression:

$$\lambda_m(x,y|z) = \int_0^\infty \left[1 - (1 - \hat{S}_{Y|X}(uy|x,z))^m\right] du \tag{10}$$

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*):

$$\hat{S}_{Y,n}(y|x,z) = \frac{\sum_{i=1}^{n} I(x_i \leq x, y_i \geq y) K_{\hat{h}}(z, z_i)}{\sum_{i=1}^{n} I(x_i \leq x) K_{\hat{h}}(z, z_i)} \tag{11}$$

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. This would be straightforward if all the Z variables are continuous, but it becomes more complex if we have mixed data (continuous and discrete variables) as it is the case in our empirical study. De Witte and Kortelainen

(2013) proposed a standard multivariate product kernel for continuous, ordered discrete and unordered discrete variables, in order to smooth these mixed variables and obtain a generalized product kernel function ($K_{\hat{\hat{h}}}$) that can substitute $K_{\hat{h}}$ in equation 11. Regarding the estimation of the bandwidth parameters, we follow the data-driven selection approach developed by Badin et al (2010), which can be easily adapted to the case of mixed external variables. Subsequently, the conditional estimator $\hat{\lambda}(x, y|z)$ can be obtained by plugging in the new $\hat{S}_{Y,n}(y|x, z)$ in Equation 8.

This conditional approach allows us to evaluate the direction of the effect of external variables on the production process by comparing conditional with unconditional measures. In particular, when Z is continuous and univariate, Daraio and Simar (2005, 2007a) suggest using a scatter plot of the ratio between these measures ($Q^z = \hat{\lambda}(x, y|z) / \hat{\lambda}(x, y)$) against Z and its smoothed nonparametric regression line. In an output-oriented conditional model, an increasing regression line will indicate that Z is favorable to efficiency whereas a decreasing line will denote an unfavorable effect.

In addition, 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 dataset 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 dataset comprises a total number of 2,398 schools distributed across countries as it is shown in Table 1.

(*Table 1 around here*)

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, 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). Although PIRLS analysts recommend to use all of them to obtain more consistent estimations (see Martin and Mullis, 2012), in our analysis we calculate the mean value of those five plausible values, since the robustness of results is guaranteed by the use of the order-m approach, which reduces the impact of measurement error by drawing repeatedly (B times) observations from the sample.

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 related to the output. In this sense, the studies that use the school as the level of analysis usually include one indicator representing each of the following three groups of variables (e.g. Wilson, 2005): (i) the characteristics of pupils (abilities or socioeconomic background), (ii) indicators representing human resources related to teachers, (i.e., total number, experience or level of education) and (iii) variables related to school resources (expenditure per pupil, quality of educational resources, books in the library or computers for instruction). In addition, we follow the criteria of selecting variables that are positively correlated with the output as well as other basic rules like not mixing indices and volume measures or selecting continuous variables (see Dyson et al, 2001 for details). The following list shows our final input variables selected from the dataset:

- A composed index representing early literacy skills of students before entering the primary grades (EARLIT) as a proxy for students' abilities (see Foy and Drucker, 2013 for details about its construction), since PIRLS does not provide a continuous index representing the socioeconomic status of pupils like PISA.
- Number of teachers per a hundred students (TEACH100).
- Number of computers per a hundred students (COMP100).

Regarding the variables representing the environment in which the school is operating, we distinguish two different groups: the school environment and the country features. The first one includes a mixed set of seven indicators with theoretical support in the literature. In particular, there are two continuous variables representing the total instructional hours per year (INSTIME) and the average level of classroom disciplinary climate perceived by the 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 at school (INVSCHL) and three (unordered) dummy variables regarding whether there is problem of absenteeism at the school (ABSENT), whether the proportion of students from a disadvantaged background exceeds or not the 50% (PDESADV) and if the school is placed in a rural area or not (RURAL). The three variables were rescaled to have a value equal to 1 for those conditions and equal to 2 otherwise.

The second group is composed by five continuous indicators about economic and cultural aspects collected at country level. The two economic variables are represented by the gross domestic product 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 dataset 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 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, Openess 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 ans Schils (2011) study the development of the performance of students during the test finding that Conscientiousness turns out to be associated with a smaller performance drop. For this reason, we have only selected responses 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.

(Table 2 around here)
(Table 3 around here)

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 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). Regarding the value of the parameter *m*, which determines the sample size for comparisons, we use *m=100* because this is the size of the sample for the country with least observations (Norway). Likewise, we checked that from this value the decrease in super-efficient observations stabilizes (Daraio and Simar, 2005). This parameter is used 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 and avoid potential bias due to different sample sizes. For statistical inference, we use 200 bootstrap replications following the recommendation made by Daraio and Simar (2005, p. 103).

(*Table 4 around here*)

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 1.1418, which indicates that if all students would perform as efficiently as the best practice students, the test scores could increase on average by 14% (or 15% if we only consider the inefficient students). Likewise, it is worth noting that some schools have a performance score below one. These super-efficient schools are performing better than the average 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, Netherlands and the Czech Republic, while the schools with the worst performance are operating in Georgia, Romania, Norway and Spain. These ranking basically coincides with the ranking of countries according to their results shown in Table 3, thus the consideration of inputs in the analysis of efficiency does not have a great impact on efficiency scores, probably because the existing variation across countries is not too high (see Table 3).

If we focus on the decomposition of this overall efficiency, it is possible to detect that, on average, the inefficiency is almost exactly shared between the school and the country operating environment (1.070 vs. 1.068). However, it is possible to find significant differences across countries. For instance, most part of the inefficiency in schools operating in Hungary (96%) or Italy (77%) depend on specific school factors, while in Romania and Lithuania the country effect is quite more relevant (70% and 67%, respectively) to explain the levels of inefficiency of their schools.

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' background and parental involvement, absenteeism and disciplinary levels and instructional time). 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.

(*Table 5 around here*)

Once we include information about environmental variables in the analysis the average efficiency decreases in both models (1.057 for model 1 and 1.040 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 main interesting conclusions can be drawn by exploring the distribution of the efficiency scores across countries. In this sense, if we observe the average efficiency scores estimated with model 1, it is possible to detect that Hungary and Italy now are included among the top three countries in the ranking. This is noteworthy, since in those countries the inefficiency attributed to the school environment was really high, as we could check in our previous analysis, so when we take into account variables representing this context, the schools operating in those countries are placed at the top of the ranking. In contrast, the average level of efficiency of schools operating in Finland, which were considered as the top performers in the unconditional model, now is not so high, thus they are placed at the fifth position of the ranking of countries.

Something similar occurs to Sweden, Slovenia and Spain, which also have an inferior position when school environmental factors are taken into account, whereas Romania and Lithuania ascend in the ranking despite the school factors did not seem to explain a great part of their unconditional inefficiency. However, the consideration of country features in model 2 modifies the picture in a great extent. According to this classification, schools operating in Norway are, on average, the best performers, although they were included in the group of the worst performers according to the classification in the unconditional and the conditional model 1. In contrast, Hungary drops from the second in the previous rank to the ninth position now and Italy from the third to the twelfth. Likewise, two countries where the country environment seems to matter like Romania and Lithuania now climb up to the third and forth position. 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 section 2.4. 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 favorable or unfavorable correlation with efficiency according to the visualization of the partial regression scatter plots.

In model 1, two of the school environmental factors have a clear significant influence on efficiency, the location (negative for rural areas) and the parental involvement at home (positive), although the parental involvement in school and the proportion of students from a disadvantaged background also has a significant correlation with efficiency. This evidence confirms the well-known importance of students' background on explaining the performance of schools (Hanushek, 2003) and particularly the important role of parental involvement (see Wilder, 2014 for an extensive review on this topic). In contrast, our results about the significant influence of the location and the insignificance of some specific school factors such as the instructional time or the discipline in the class contrast with the majority of recent evidence on cross-country analyses (e.g. Fuchs and Woessman, 2007; Lavy, 2010; Rivkin and Schiman, 2013), although those studies are based on test scores data about secondary schools participating in PISA for which the variation among schools is usually higher.

The results of model 2 reveal that all five variables representing specific features of countries have a significant and positive influence on efficiency, while the influence of school variables becomes insignificant for all the considered indicators. The country factors seem to have a more relevant role in explaining differences in the performance of schools because there is more heterogeneity across countries in those variables than in the school frameworks. One potential explanation for this result is that our sample only includes European countries, where the school environment 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 (e.g., 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 (e.g., 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 (e.g., 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.

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, 2007b) and extended by De Witte and Kortelainen (2013) to account for two different sets of variables related to the school and the country environment. Likewise, nonparametric bootstraps based on significance tests have been applied to examine the statistical influence of those factors on efficiency scores.

Our results show that an efficiency analysis that only take into account some basic inputs apart from the academic performance as output leads to analogous results that a simple ranking of countries based on their students´ results. However, when we include in the analysis some factors representing the environment where those schools are operating the ranking of countries changes dramatically. Hence, some countries considered as the worst performers according to the results of their students, like Norway or Romania, are placed among the best when those environmental factors are accounted for. Likewise, 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 or their inefficient performance. In this sense, the results indicate that country variables have more influence on school performance than the characteristics of the students attending or the traditional school factors such as the instructional time or the maintenance of certain levels of discipline in the classes. 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 Woessman, 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 or another alternative approach (for instance, see Crespo-Cebada et al, 2014). 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. Likewise, the proposed analysis could be replicated using pupil level data. Actually, these data have been used in some recent studies to assess the performance of students in a specific country (e.g. Perelman and Santín, 2011; De Witte and Kortelainen, 2013). However, it must been 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 datasets.

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.

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*, in press.

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: 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* 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

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.

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

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 2011-001.

Crespo 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: 13–32.

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

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 Rogge N (2011). Accounting for exogenous influences in performance evaluations of teachers. *Economics of Education Review* 30: 641-653.

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

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 Siber 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.

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.

Dyson RG, Allen R, Camanho AS, Podinovski VV, Sarrico CS and Shale EA (2001). Pitfalls and protocols in DEA. *European Journal of Operational Research* 132: 245-259.

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

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

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: 541-549.

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

Hanushek E A (2003). The Failure of Input-based Schooling Policies. *The Economic Journal* 113(485): 64-98.

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* 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. Amsterdam: North Holland, pp 89-200.

Hanushek, E. A., Link, S., & 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, pp 145-176.

Heckman JJ (2011). *Integrating personality psychology into economics* (Nº 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: 105-122.

Kneip A, Park BU and Simar L (1998). A note on the convergence of nonparametric DEA estimators for production efficiency scores. *Econometric Theory* 14(6): 783-793.

Krueger AB and Lindahl M (2001). Education for Growth: Why and For Whom. *Journal of Economic Literature* 39: 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: 3-24.

Li Q and Racine JS (2004). Cross-Validated Local Linear Nonparametric Regression. *Statistica Sinica* 14(2): 485-512.

Martin MO and Mullis IVS (2012). *Methods and procedures in TIMSS and PIRLS 2011*. Chestnut Hill, MA: TIMSS and PIRLS International Study Center, Boston College.

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: 133-161.

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

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.

Park B, Simar L and Weiner C (2000). The FDH estimator for productivity efficiency scores. *Economic Theory* 16(6): 855–877.

Perelman S and Santin 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, 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: 523–544.

Rasch G (1960/1980) *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Danish Institute for Educational Research, 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.

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.

Simar L and Wilson PW (2000). Statistical inference in nonparametric frontier models: The state of the art. *Journal of Productivity Analysis* 13(1): 49-78.

Strietholt R, Gustafsson JE, Rosen M and Bos W (2014). Outcomes and Causal Inference in International Comparative Assessments. *Educational Policy Evaluation through International Comparative Assessments.*

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.* Cheltenham, UK: Edward Elgar.

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.

Zhang Y and Bartels R (1998). The effect of sample size on the mean efficiency in DEA with an application to electricity distribution in Australia, Sweden and New Zealand. *Journal of Productivity Analysis* 9(3): 187-204.

TABLES

Table 1.Dataset 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	**2,398**		

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
EARLIT	Input	9.76	0.85	5.26	12.19
TEACH100	Input	4.80	1.22	1.08	18.18
COMP100	Input	5.53	4.35	0.07	73.05
School factors					
INSTIME	Continuous	826.77	174.55	480	1640
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
PDESADV	Unordered discrete	1.15	0.35	1.00	2.00
RURAL	Unordered discrete	1.40	0.49	1.00	2.00
Country features					
GDP pc	Continuous	31,485.7	21,301.1	3,220.0	99,173.0
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

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

COUNTRIES	OUTPUT	INPUTS			ENVIRONMENTAL VARIABLES											
					SCHOOL FACTORS							COUNTRY FACTORS				
	PVREAD	**EARLIT**	**TEACH100**	**COMP100**	**INSTIME**	**DISCPL**	**INVHOME**	**INVSCHL**	**ABSENT**	**PDESADV**	**RURAL**	**GDP pc**	**EXPEDUC**	**HARDWORK**	**RESP**	**PERSEV**
Finland	568.49	10.21	5.12	6.76	789.25	10.15	3.01	3.16	1.37	1.03	1.46	48,678	21.10	0.15	0.89	0.55
Czech Republic	548.57	9.78	4.75	7.07	788.56	10.22	3.49	2.47	1.22	1.05	1.42	20,580	15.60	0.82	0.67	0.39
Netherlands	544.34	9.26	4.15	5.41	1076.85	9.10	2.83	3.39	1.54	1.11	1.39	49,886	18.70	0.29	0.90	0.37
Germany	541.95	9.19	4.74	6.29	860.04	9.60	3.01	3.43	1.58	1.10	1.48	44,355	18.30	0.19	0.88	0.52
Hungary	540.63	8.82	4.75	6.90	765.90	9.86	3.32	3.04	1.61	1.26	1.44	13,964	22.51	0.30	0.58	0.31
Italy	540.61	9.34	5.96	5.63	1080.78	9.69	3.24	3.22	1.71	1.11	1.65	36,180	24.10	0.39	0.87	0.44
Sweden	537.61	10.33	4.72	4.37	845.35	9.70	3.27	2.88	1.56	1.11	1.37	56,755	27.60	0.07	0.89	0.37
Bulgaria	532.60	9.81	4.69	4.62	673.72	10.62	3.19	2.76	1.29	1.28	1.39	7,287	23.20	0.90	0.68	0.50
Lithuania	531.72	10.18	5.17	3.41	657.69	10.58	3.26	3.26	1.27	1.20	1.26	14,158	24.30	0.89	0.74	0.36
Poland	529.15	10.05	4.92	5.09	765.09	9.74	3.54	3.09	1.16	1.09	1.40	13,382	27.40	0.18	0.80	0.26
Slovenia	527.94	9.29	5.34	7.02	684.00	10.12	3.21	2.82	1.31	1.12	1.37	24,429	30.98	0.34	0.72	0.56
France	522.22	10.23	4.19	6.23	914.11	10.38	3.80	3.17	1.54	1.17	1.57	42,560	18.70	0.62	0.78	0.54
Spain	514.32	10.86	4.41	4.76	870.07	10.62	3.62	2.94	1.43	1.12	1.13	31,118	21.50	0.61	0.73	0.26
Norway	509.95	9.14	4.77	7.67	815.82	9.92	3.60	3.07	1.51	1.01	1.45	99,173	21.10	0.12	0.91	0.39
Romania	506.51	9.28	5.05	4.74	801.16	10.40	3.44	2.85	1.66	1.42	1.54	9,064	19.60	0.75	0.74	0.44
Georgia	488.07	9.64	4.72	4.11	731.30	10.84	3.17	2.97	1.73	1.20	1.33	3,220	13.20	0.85	0.73	0.32
Country Avg.	**530.29**	**9.71**	**4.84**	**5.63**	**819.98**	**10.10**	**3.31**	**3.03**	**1.47**	**1.15**	**1.42**	**32.174**	**21.74**	**0.47**	**0.78**	**0.41**

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

Countries	Overall Efficiency				School Effect					Country Effect				
	Mean	**ST**	**Min**	**Max**	**Mean**	**%**	**ST**	**Min**	**Max**	**Mean**	**%**	**ST**	**Min**	**Max**
Finland	1.0673	0.0388	0.9924	1.1909	1.0327	48.53%	0.0375	0.9994	1.1361	1.0341	50.65%	0.0343	0.9882	1.1429
Netherlands	1.0995	0.0610	1.0000	1.2864	1.0456	45.80%	0.0481	0.9976	1.2250	1.0515	51.77%	0.0312	1.0000	1.1341
Czech Republic	1.1088	0.0525	1.0000	1.2709	1.0494	45.40%	0.0506	0.9963	1.2454	1.0574	52.71%	0.0403	1.0000	1.2152
Hungary	1.1129	0.0979	0.9970	1.5183	1.1088	96.41%	0.1002	1.0000	1.4958	1.0042	3.74%	0.0281	0.9808	1.2108
Germany	1.1210	0.0718	0.9987	1.4284	1.0805	66.54%	0.0721	0.9998	1.3193	1.0385	31.79%	0.0429	0.9876	1.3371
Sweden	1.1255	0.0706	1.0000	1.3099	1.0538	42.90%	0.0594	0.9965	1.2548	1.0685	54.57%	0.0459	1.0000	1.2747
Italy	1.1258	0.0829	0.9875	1.6141	1.0968	76.98%	0.0894	0.9991	1.6173	1.0276	21.96%	0.0382	0.9884	1.1920
Bulgaria	1.1276	0.1164	0.9861	1.7704	1.0837	65.60%	0.1099	0.9992	1.7732	1.0415	32.54%	0.0528	0.9868	1.2343
Lithuania	1.1393	0.0716	1.0000	1.3847	1.0423	30.39%	0.0470	0.9991	1.1721	1.0939	67.38%	0.0658	1.0000	1.3531
Poland	1.1459	0.0659	0.9899	1.2841	1.0773	48.44%	0.0607	0.9875	1.2460	1.0643	47.08%	0.0391	0.9997	1.2370
Slovenia	1.1546	0.0550	1.0167	1.3286	1.0581	50.00%	0.0493	0.9832	1.2444	1.0918	41.57%	0.0355	1.0322	1.2480
France	1.1646	0.0739	1.0000	1.4053	1.0810	35.28%	0.0739	0.9999	1.3595	1.0786	55.75%	0.0463	1.0000	1.2785
Spain	1.1802	0.0842	1.0000	1.4643	1.0902	44.95%	0.0737	1.0000	1.3922	1.0832	43.61%	0.0472	1.0000	1.3008
Norway	1.1807	0.0714	1.0000	1.3923	1.0436	49.90%	0.0482	0.9979	1.2321	1.1316	46.07%	0.0506	1.0000	1.2576
Romania	1.1881	0.1305	0.9861	1.6482	1.0784	23.16%	0.1025	0.9958	1.447	1.1033	69.97%	0.0879	0.9902	1.6482
Georgia	1.2277	0.1160	1.0000	1.5888	1.0742	34.41%	0.0799	0.9973	1.3698	1.1433	45.38%	0.0767	1.0000	1.4124
TOTAL	**1.1418**	**0.0897**	**0.9860**	**1.7704**	**1.0707**	**48.15%**	**0.0754**	**0.9832**	**1.7732**	**1.0687**	**45.77%**	**0.0603**	**0.9808**	**1.6482**

Table 5. Efficiency score distribution across countries in different models

UNCONDITIONAL			CONDITIONAL MODEL 1			CONDITIONAL MODEL 2		
	Mean	ST		Mean	ST		Mean	ST
Finland	1.0673	0.0388	Netherlands	1.0247	0.0426	Norway	1.0211	0.0388
Netherlands	1.0995	0.0610	Hungary	1.0358	0.0630	Netherlands	1.0247	0.0388
Czech Republic	1.1088	0.0525	Italy	1.0385	0.0630	Romania	1.0307	0.0729
Hungary	1.1129	0.0979	Romania	1.0425	0.0914	Lithuania	1.0312	0.0533
Germany	1.1210	0.0718	Finland	1.0435	0.0414	Georgia	1.0339	0.0678
Sweden	1.1255	0.0706	Germany	1.0445	0.0652	Finland	1.0345	0.0391
Italy	1.1258	0.0829	Lithuania	1.0461	0.0652	Sweden	1.0359	0.0535
Bulgaria	1.1276	0.1164	Czech Republic	1.0474	0.0533	Czech Republic	1.0363	0.0472
Lithuania	1.1393	0.0716	Bulgaria	1.0563	0.1055	Hungary	1.0364	0.0618
Poland	1.1459	0.0659	Sweden	1.0600	0.0694	Slovenia	1.0381	0.0479
Slovenia	1.1546	0.0550	Georgia	1.0636	0.1056	Germany	1.0398	0.0627
France	1.1646	0.0739	Poland	1.0674	0.0742	Italy	1.0402	0.0636
Spain	1.1802	0.0842	Norway	1.0679	0.0771	Poland	1.0445	0.0655
Norway	1.1807	0.0714	France	1.0733	0.0805	Bulgaria	1.0500	0.1016
Romania	1.1881	0.1305	Slovenia	1.0848	0.0717	France	1.0580	0.0719
Georgia	1.2277	0.1160	Spain	1.0914	0.0957	Spain	1.0634	0.0777
TOTAL	**1.1432**	**0.0897**	**TOTAL**	**1.0576**	**0.0784**	**TOTAL**	**1.0405**	**0.0644**

Tabla6. Influence of different factors on educational performance (Estimation of nonparametric significance tests)

	Model 1		Model 2	
School variables	**p-value**	**Influence (scatter plot)**	**p-value**	**Influence (scatter plot)**
Disadvantage background	0.01*	Unfavorable	1.00	Unfavorable
Rural area	(<2e-16)***	Unfavorable	1.00	Unfavorable
Parents' involvement at home	(<2e-16)***	Favorable	1.00	Favorable
Parents' involvement in school	0.04*	Favorable	1.00	Favorable
Absenteeism	0.14	Unfavorable	1.00	Unfavorable
Disciplinary index	0.98	Favorable	1.00	Favorable
Instructional time	0.86	Favorable	1.00	Favorable
Country variables			**p-value**	**Influence (scatter plot)**
GDP pc			(<2e-16)***	Favorable
Expenditure in education			(<2e-16)***	Favorable
Hard work			(<2e-16)***	Favorable
Responsibility			(<2e-16)***	Favorable
Perseverance			(<2e-16)***	Favorable

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

FIGURES

Figure 1. Metafrontier illustration (decomposition of school and country effect)