Omega 79 (2018) 9–20

Contents lists available at ScienceDirect

Omega

journal homepage: www.elsevier.com/locate/omega

Using non-radial DEA to assess school efficiency in a cross-country perspective: An empirical analysis of OECD countries☆

Juan Aparicio[a], Jose M. Cordero[b,*], Martin Gonzalez[a], Jose J. Lopez-Espin[a]

[a] Center of Operations Research (CIO), University Miguel Hernandez of Elche (UMH), 03202 Elche (Alicante), Spain
[b] Department of Economics, University of Extremadura (UEX), 06006 Badajoz, Spain

ARTICLE INFO

Article history:
Received 23 January 2017
Accepted 21 July 2017
Available online 26 July 2017

Keywords:
Data envelopment analysis
Non-radial measures
Least distance
Education

ABSTRACT

In this paper we use data from OECD countries participating in PISA 2012 to assess the efficiency of schools in a cross-country framework. In the analysis, and in contrast to previous applications, we consider that schools might concentrate their efforts on improving the results in one dimension of the educational output to a greater extent than in the other. To do this, we rely on non-radial efficiency measures of performance and the estimation of an educational production function based upon Data Envelopment Analysis (DEA) techniques. Specifically, DEA non-radial measures allow for identifying different levels of inefficiency for each output considered (reading and maths). In particular, we apply a non-radial measure based on Ando et al. [5] and Aparicio et al. [12]. Our results show that the majority of schools in OECD countries tend to be less efficient in reading than in mathematics.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

The participation of the majority of nations in international large-scale comparative studies in education has provided researchers with rich and extensive cross-national databases that can be used to assess the performance and effectiveness of educational systems. As a result, comparative education studies have become increasingly popular in education sciences today [49], since researchers can look at the entire world as a natural laboratory to view the multiple ways in which societal factors, educational policies and practices may vary across countries [19].

Most studies adopting a cross-country perspective are situated within the field of educational effectiveness research, which explores the main determinants of educational achievement using an econometric approach to estimate an equation in the form of an educational production function (see [32,51,78]). This strand of literature investigates how inputs are statistically related with outputs. However, the potential existence of an unexpected level of inefficiency in the performance of students, schools or educational systems also needs to be considered [57]. In this sense, the existing constraints of resources faced by most countries and the great amount of national income devoted to educational costs, policy makers and researchers have become increasingly concerned with assessing the efficiency of schools, although until now most of existing literature has been devoted to assessing schools operating in the same country or region.[1]

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. This line of research includes several works using data that has been aggregated at a country level from different samples of countries participating in international tests such as PISA [1,14,43,76,81]; Agasisti [3,45] or TIMSS [44]. Likewise, we can also find studies that compare the performance of educational systems in different countries using data at school level. For example, Sutherland et al. [75] study the performance of schools from 30 OECD countries participating in PISA 2003; Agasisti and Zoido [4] derives efficiency measures for more than 8600 schools in 30 countries using PISA 2012 data comparing efficiency scores and measures of equity; Cordero et al. [31] evaluates performance using the metafrontier framework to compare and decompose the technical efficiency of primary schools from 16 European countries participating in PIRLS 2011. Finally, De Jorge and Santín [34] and Deutsch et al. [37] use PISA data at a student level to estimate the efficiency of EU and Latin American countries, respectively.

Those studies predominantly use nonparametric techniques like DEA or FDH [24,36]. These methods are generally based on Farrell–Debreu radial efficiency measures, i.e. they reflect the ability of the unit to increase different outputs (e.g. test scores in maths

☆ This manuscript was processed by Associate Editor Chien-Ming Chen.

* Corresponding author.
E-mail address: jmcordero@unex.es (J.M. Cordero).

[1] Recent literature reviews on efficiency in education include De Witte and López-Torres [35], Johnes [55] and [48].

http://dx.doi.org/10.1016/j.omega.2017.07.004
0305-0483/© 2017 Elsevier Ltd. All rights reserved.

Table 1
Ranking of OECD countries according to results in maths and reading.

Country	Maths	Rank	Reading	Rank
Korea	554	1	536	2
Japan	536	2	538	1
Switzerland	531	3	509	12
Netherlands	523	4	511	10
Estonia	521	5	516	7
Finland	519	6	524	3
Canada	518	7	523	5
Poland	518	8	518	6
Belgium	515	9	509	11
Germany	514	10	508	13
Austria	506	11	490	21
Australia	504	12	512	9
Ireland	501	13	523	4
Slovenia	501	14	481	29
Denmark	500	15	496	18
New Zealand	500	16	512	8
Czech Rep.	499	17	493	19
France	495	18	505	14
UK	494	19	499	16
Iceland	493	20	483	28
Luxembourg	490	21	488	24
Norway	489	22	504	15
Portugal	487	23	488	25
Italy	485	24	490	20
Spain	484	25	488	23
Slovak Rep.	482	26	463	32
USA	481	27	498	17
Sweden	478	28	483	27
Hungary	477	29	488	22
Israel	466	30	486	26
Greece	453	31	477	30
Turkey	448	32	475	31
Chile	423	33	441	33
Mexico	413	34	424	34

and reading) equiproportionately. Nevertheless, schools sometimes might concentrate their efforts on improving the results in one dimension of the educational outcome more than in other one, thus there might be trade-offs between outputs that cannot be identified through radial efficiency measures. This intuition arises from the fact that there is relevant cross-country divergence in test results between maths and reading. Table 1 illustrates this evidence by showing the ranking of OECD countries in these subjects. Although, they are essentially similar, it is possible to observe some countries with relatively better results in maths (Switzerland, Netherlands, Austria or Slovenia) or in reading (Ireland, New Zealand or United States).

In this paper we would like to explore the potential existence of these trade-offs between reading and maths in an assessment of the performance of schools from all OECD countries participating in PISA 2012 adopting a cross-national framework. For this purpose, we rely on non-radial efficiency measures, which do not require an equiproportional increase in all the considered outputs, thus we can calculate different projections on the frontier for each output included in the production function. This possibility allows us to detect whether some schools may be more efficient in promoting their students´proficiency in reading, while other schools could be more prone to enhance the results in mathematics. Likewise, the proposed approach can also be useful to explore other potential trade-offs between educational outcomes such as the relationship between cognitive and non-cognitive skills [30][2] or educational inequality and average achievement [45,76].[3] Moreover, since we adopt a cross-national framework, we can derive some interesting insights about the average performance of schools from the same country, making it possible to construct different ranking of countries according to the levels of efficiency demonstrated in promoting different educational outcomes. This is an unusual approach in the literature on efficiency measurement in the educational sector, since most studies tend to analyze the performance of schools treating test scores as the unifying outcome that needs to be improved.

Some research in the literature has sought to construct different non-radial efficiency measures such as the Russell measure [38,66], the additive model [25], the slacks-based measure [79] or the Multi-directional Efficiency Analysis (MEA) approach of Bogetoft and Hougaard [18] and Asmild et al. [15]. In our case, we apply a recent methodology, fundamentally based on the application of the 'output-oriented' version of the Russell measure that determines the closest targets and the least distance to the strongly efficient frontier in DEA, based on Bilevel Linear Programming [12]. In addition, with the aim of satisfying monotonicity, a correction of this model is proposed based on Ando et al. [5]. Given that Ando et al. [5] did not show how to implement their approach without previously determining the explicit characterization of the set of points belonging to the strongly efficient frontier, we show for the first time in this paper how this methodology can be implemented in practice.

Moreover, we examine the potential determinants of existing divergence in the setting of production targets across schools and countries using a two-stage approach (DEA and regression). Among potential drivers of schools´performance we distinguish between school factors in the surrounding context and variables representing students´attitudes toward mathematics and reading with the aim of exploring whether schools with students devoting more time to one of those subjects tend to concentrate their effort in promoting that subject or the other one.

The contribution of this paper is threefold. First, as we are aware, there is no paper in the literature devoted to the estimation of technical efficiency in education for OECD countries that applies non-radial measures. So, in this sense, the empirical application that we present is original. Second, our analysis allows us to determine, for the first time, which dimension (in particular, between reading and maths) presents more technical inefficiency for schools in OECD countries. Third, as Aparicio [13] argues, current methodologies associated with the determination of the least distance in DEA lack real applications. Therefore, this paper represents an example of the use of this type of techniques, which permits the implementation of the Principle of Least Action in DEA [9].

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. Finally, the paper ends with some concluding remarks in Section 5.

2. Methodology

Data Envelopment Analysis was introduced by Charnes et al. [24] under constant returns to scale for multiple inputs and outputs and later extended by Banker et al. [16] to variable returns to

[2] The relationship between those dimensions of educational outcomes is usually difficult to explore due to the difficulties of establishing a standard definition for non-cognitive skills. However, recently, the PISA survey has added some component tests designed to capture aspects of non-cognitive skills including openness, locus of control, and motivation [66].

[3] In their empirical analysis of the performance of different countries participating in PISA, Giménez et al. [45] claim that "*for a given country, the results of the educational process should not be constrained to the knowledge students acquire at school, but should also include other outcomes such as the standard deviation of test scores (an undesirable outcome of the educational process, in terms of educational inequality)*".

scale. Nevertheless, Farrell [39] was the first in showing, for a single output and multiple inputs, how to estimate an isoquant enveloping all the observations, implementing the seminal ideas of Shephard [71] regarding the input distance function. In all these cases, the technical efficiency assessment of a Decision Making Unit (DMU) is based upon an 'oriented' measure of distance, which identifies a point on the isoquant of the technology with the same mix of inputs (input orientation) or outputs (output orientation) of that of the evaluated unit. The conservation of this mix in movements toward the boundary of the technology is the characteristic that gives a radial measure.

However, many real-life situations require non-radial measures of technical efficiency to be used. Any measure in DEA that does not adopt equi-proportional reductions of inputs or outputs is non-radial. Indeed, a well-known drawback of radial measures is the arbitrariness in imposing targets on the isoquant preserving the mix within inputs or within outputs, depending on the selected orientation of the model, when the firm's very reason to change its input/output levels might often be the desire to actually change that mix due to their differing opportunity costs (see [23], and [68]). In particular, in the context of education, some units (schools, for example) could be tempted, directly or indirectly, to upgrade some specific dimension, like science or mathematics, due to, for example, cultural characteristics and traditions inherent to the country where they are geographically located. Additionally, from a DMU point of view, it may mainly be interested in the easiest way of being classified as technically efficient (especially for public "production" like schools, and the regulatory pressure that could arise from being classified as inefficient rather than efficient). This type of benchmarking strategy will be the focus of our contribution in this paper.

Regarding the existence of non-radial oriented models, DEA endows practitioners with a toolbox full of possibilities. The first approach in this respect was due to Färe et al. [38], who introduced the Russell input and output measures of technical efficiency. After that, other oriented and non-radial measures were defined seeking more flexibility than that provided by the radial measures as, for example, the directional distance function [22] or the weighted additive models (see [47] or [28]). In contrast to these two last types of technical efficiency measures, the Russell input and output measures present some interesting properties. First, the directional distance function does not correspond to the Pareto–Koopmans definition of technical efficiency [56]. This implies that it ignores the possible existence of slacks associated with the projected points on the boundary of the technology. In other words, the directional distance function neglects some additional sources of technical inefficiency. Conversely, the Russell measures always generate non-dominated projection points in the corresponding input/output space. Second, in contrast to the weighted additive models, which aggregate slacks by a weighted scheme, the interpretation of the Russell measures is easier. In particular, the value of the Russel output measure can be interpreted as the average of proportional rates of output expansion needed to be technically efficient.

Regarding the Russell measures, we want to highlight that there are two clearly different paradigms for determining them nowadays. On the one hand, we have the traditional approach, which is associated with the identification of demanding targets. The targets are specifically the coordinates of the projection point on the boundary of the technology and thus represent levels of operation that would make the evaluated DMU perform efficiently. This first philosophy is followed by the original definition of the Russell input and output measures [38], where the total technical effort required by a DMU to become technically efficient is maximized instead of minimized, thereby generating the furthest projection points on the frontier. On the other hand, a recent proposal

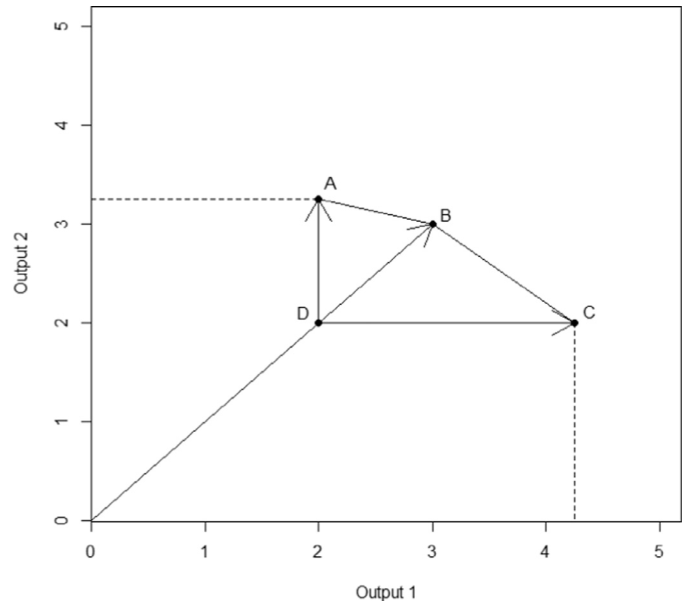

Fig. 1. Differences between radial and non-radial measures.

[12] has suggested determining the closest efficient targets through the oriented Russell measures, minimizing, instead of maximizing, the corresponding technical effort in order to reach the frontier. This second approach follows a well-known line of research in the DEA literature related to the determination of the least distance to the efficient frontier, the identification of closest targets and the application of the Principle of Least Action (see, for example, [7–11,21,40,67,68], Fukuyama et al. [42] and Ruiz and Sirvent [69]). Next, we graphically illustrate the main differences between the Russell measures under the two philosophies and their comparison with the radial measure (see Fig. 1). To do that, we focus our simple example on the Russell output measure of technical efficiency.

In Fig. 1, an output production set in two dimensions is represented. In this simple example, the only technically inefficient DMU is unit D = (2, 2). If it is assessed by the output-oriented radial measure, then unit B = (3, 3) is its corresponding projection point with a radial score of 1.5. This same value can be interpreted in terms of the Russell measures as the average of output changes associated with the radial projection: $1.5 = 0.5 \cdot (\frac{3}{2} + \frac{3}{2})$. In the case of the traditional 'non-radial' Russell output measure, the projection point on the isoquant would be unit C (4.25, 2), which produces an average of proportional rates of output expansion equals $0.5 \cdot (\frac{4.25}{2} + \frac{2}{2}) = 1.5625$. However, there is also an alternative projection point on the frontier that yields the least value for the average of output expansions. We are referring to unit A (3.25, 2): $0.5 \cdot (\frac{2}{2} + \frac{3.25}{2}) = 1.3125 < 1.5 < 1.5625$. This last projection corresponds to the approach by Aparicio et al. [12], which, in general, determines the Pareto-efficient point more easily achievable for the evaluated DMU. Note that, for DMU D, it is possible to be technically efficient improving output 2 only by 62.5%, instead of enhancing output 1 and output 2 by 50% (radial projection) or increasing output 1 only by 112.5% (traditional Russell projection).

Other key issue in the measurement of technical inefficiency is the satisfaction of some interesting properties. One of them is monotonicity. Monotonicity relates the notion of efficiency to Pareto optimality. Specifically, if unit A dominates unit B, in the Pareto sense, then the measure of technical inefficiency associated to A should be less than the measure of technical inefficiency of B. Briec [21] proved that Hölder distance functions meet weak monotonicity over the weakly efficient frontier. Ando et al. [5] were the first in showing that Hölder distance functions do not meet strong monotonicity on the strongly efficient frontier and suggested a solution for satisfying weak monotonicity on the strongly efficient frontier. Later, Aparicio and Pastor [8] proved that the output-oriented version of the Russell measure is a well-defined efficiency measure, satisfying strong monotonicity on the strongly efficient frontier, if efficiency is evaluated with respect to an extended facet production possibility set based on Full Dimensional Efficient Facets (FDEF) instead of the standard DEA technology.

Indeed, Aparicio and Pastor [10] showed that this drawback of the Hölder norms is associated with the dimensionality of the strongly efficient frontier. On the other hand, Fukuyama et al. [41,42] showed an alternative to the extension of FDEFs for endowing least distances with the property of strong monotonicity. Their approach is based on transforming the traditional definition of the measure introducing in the corresponding optimization model an auxiliary point that is dominated by the evaluated unit, and that it is used to calculate the least distance to the strongly efficient frontier. Finally, Ando et al. [6] proved the strong monotonicity property for the input and output oriented DEA models based on the Hölder norms in the context of two inputs and two outputs, respectively, as well as the strong monotonicity for non-oriented DEA models in two dimensions (one input – one output). Additionally, they reviewed the minimum distance inefficiency measures based on the extended facet approach, and discussed the problems of unmeasurability and overestimation. All this means that the least distance measure proposed in Aparicio et al. [12], as defined originally, satisfies neither weak nor strong monotonicity. Nevertheless, according to the above discussion, the technology or the measure could be modified in order to try to meet the property. In particular, in this paper, we opt for the transformation of the inefficiency measure *a la* Ando et al. [5] in order to correct the problem since the modification of the traditional DEA technology requires to detect the set of all FDEFs, if they exist. Additionally, given that Ando et al. [5] did not provide the way of implementing their approach without an explicit description of the strongly efficiency frontier, we show for the first time how it can be carried out using an implicit characterization of this set of the frontier. Thanks to the correction of the original model of Aparicio et al. [12], we can be sure that the proposed measure in this paper satisfies at least weak monotonicity in the output-oriented context.

Definition 1. Let $I : R^m_+ \times R^s_+ \to R_+$ be an output-oriented inefficiency index. I satisfies weak monotonicity if $I(x, y) \le I(x, \tilde{y})$ for all feasible vectors (x, y) and $(x, \tilde{y})$ with $y \ge \tilde{y}$.

Moreover, implementing in practice the approach based on the determination of closest targets is not easy from a computational point of view [20]. This difficulty is consequence of the complexity of determining the least distance to the frontier of a DEA technology (a polyhedral set) from an interior point (inefficient DMU). For this reason, Aparicio et al. [12] introduce a new methodology to implement this approach in the context of the DEA-oriented measures. Their method is based upon Linear Bi-level Programming (LBP).

A Bilevel Programming model refers to a mathematical programming problem where one of the constraints is an optimization problem. This theory has been successfully applied to model different real situations with a common feature: the existence of a hierarchical structure (see [85]). A Bilevel Programming problem where both the objective functions and the constraints are linear is called a Bilevel Linear Programming problem. Denote by $z \in Z \subset R^p$ and $t \in T \subset R^q$ the decision variables corresponding to the first and second level, respectively. The general formulation of a Bilevel Linear Programming (BLP) problem is as follows:

$$
\begin{aligned}
\underset{z,t}{Min}\quad & c_1 z + d_1 t \\
s.t.\quad & \\
& A_1 z + B_1 t \le b_1, \\
& \underset{t}{Min}\ c_2 z + d_2 t \\
& s.t. \\
& \quad A_2 z + B_2 t \le b_2, \\
& z \ge 0,\ t \ge 0
\end{aligned}
\tag{1}
$$

Program (1) consists of two subproblems. On the one hand, the higher level decision problem and, the other hand, the lower level decision problem, which appears as a constraint in (1). Both problems are connected in a way that the higher problem sets parameters influencing the lower level problem and the higher problem, in turn, is affected by the outcome of the lower level problem.

It is known that even for the Bilevel Programming problem where all the functions are linear, like in (1), the model to be solved is non-convex and NP-hard. This complexity is the reason why many different techniques have been proposed in the literature to study the computational aspects of Bilevel Programming problems. The formulation of optimality conditions for this type of problems usually starts with a suitable reformulation of the problem as a one-level model. One possibility is to transform the original problem into a single optimization problem by applying the well-known Karush–Kuhn–Tucker (KKT) optimality conditions of the lower level problem [72].

Regarding the solutions of a BLP problem, $(z^*, t^*) \ge 0$ is a feasible solution of (1) if t^* is an optimal solution of the lower level program with $z = z^*$ and, at the same time, $A_1 z^* + B_1 t^* \le b_1$. In this way, (z^*, t^*) is an optimal solution if additionally $c_1 z^* + d_1 t^* \le c_1 z + d_1 t$ for all feasible solution (z, t) of (1), being $c_1 z^* + d_1 t^*$ the corresponding optimal value of the BLP problem.

Aparicio et al. [12] proposed the following model for calculating the Russell output measure based on the least distance philosophy for the DMU_0:

$$
\begin{aligned}
\underset{\phi,\lambda,s^+,\gamma}{Min}\quad & \frac{1}{s}\sum_{r=1}^{s}\phi_r && && (2.1)\\
s.t.\quad & && && \\
& \sum_{j\in E_{VRS}} \lambda_j x_{ij} \le x_{i0}, && i = 1,\ldots,m && (2.2)\\
& \sum_{j\in E_{VRS}} \lambda_j y_{rj} \ge \phi_r y_{r0}, && r = 1,\ldots,s && (2.3)\\
& \sum_{j\in E_{VRS}} \lambda_j = 1, && && (2.4)\\
& \sum_{r=1}^{s} s_r^+ = 0, && && (2.5)\\
& \underset{s^+,\gamma}{Max}\quad \sum_{r=1}^{s} s_r^+ && && (2.6)\\
& s.t. && && \\
& \quad \sum_{j\in E_{VRS}} \gamma_j x_{ij} \le x_{i0}, && i = 1,\ldots,m && (2.7)\\
& \quad \sum_{j\in E_{VRS}} \gamma_j y_{rj} = \phi_r y_{r0} + s_r^+, && r = 1,\ldots,s && (2.8)\\
& \quad \sum_{j\in E_{VRS}} \gamma_j = 1, && && (2.9)\\
& \phi_r \ge 1,\ \lambda_j,\ s_r^+,\ \gamma_j \ge 0, && \forall r, j && (2.10)
\end{aligned}
\tag{2}
$$

where $x_j = (x_{1j}, \ldots, x_{mj}) \in R^m_{++}$ denotes the inputs and $y_j = (y_{1j}, \ldots, y_{sj}) \in R^s_{++}$ denotes the outputs for a sample of $j = 1, \ldots, n$ observations (DMUs), and E_{VRS} denotes the set of extreme efficient units in the case of assuming variable returns to scale (VRS).

In (2), (2.1)–(2.4) coincide with the constraints of the traditional Russell output measure of technical efficiency ([38], p. 149) except for the fact that the objective function is minimized instead of maximized as happens with the original definition of the Russell measure, while the lower level problem, (2.6)–(2.9) is an output-oriented version of the additive model in DEA [25] for evaluating the Pareto-efficiency of $(\phi_1 y_{10}, ..., \phi_s y_{s0})$. Constraint (2.5) ensures that the optimal $(\phi_1^* y_{10}, ..., \phi_s^* y_{s0})$ is not dominated.

Note that model (3) is mathematically equivalent to a model where (2.1) is changed by $1 + Min\{\frac{1}{s}\sum_{r=1}^{s}\frac{t_r}{y_{r0}}\}$, (2.3) substituted by $\sum_{j \in E_{VRS}} \lambda_j y_{rj} \geq y_{r0} + t_r$ and (2.8) by $\sum_{j \in E_{VRS}} \gamma_j y_{rj} = y_{r0} + t_r + s_r^+$, with $t_r \geq 0$, $r = 1, ..., s$.

Now, with the aim of guaranteeing weak monotonicity when the above measure is calculated on the strongly efficient frontier, we need to modify (2) *a la* Ando et al. [5]. To do that, we need to include in the constraints a generic point $(y_1, ..., y_s) \in R_+^s$ such that it is dominated by the assessed unit in the sense of Pareto, i.e., $y_{r0} \geq y_r$, for all $r = 1, ..., s$. The corresponding Bilevel Programming problem for evaluating DMU_0 would be then as follows.

$$
\begin{array}{llll}
\underset{\phi,\lambda,y,s^+,\gamma}{Min} & 1 + \dfrac{1}{s}\displaystyle\sum_{r=1}^{s}\dfrac{t_r}{y_{r0}} & & (3.1)\\
s.t. & \displaystyle\sum_{j \in E_{VRS}} \lambda_j x_{ij} \leq x_{i0}, & i = 1, \ldots, m & (3.2)\\
 & \displaystyle\sum_{j \in E_{VRS}} \lambda_j y_{rj} \geq y_r + t_r, & r = 1, \ldots, s & (3.3)\\
 & y_r \leq y_{r0}, & r = 1, \ldots, s & (3.4)\\
 & \displaystyle\sum_{j \in E_{VRS}} \lambda_j = 1, & & (3.5)\\
 & \displaystyle\sum_{r=1}^{s} s_r^+ = 0, & & (3.6)\\
 & \underset{s^+,\gamma}{Max} \quad \displaystyle\sum_{r=1}^{s} s_r^+ & & (3.7)\\
 & s.t. & & \\
 & \quad \displaystyle\sum_{j \in E_{VRS}} \gamma_j x_{ij} \leq x_{i0}, & i = 1, \ldots, m & (3.8)\\
 & \quad \displaystyle\sum_{j \in E_{VRS}} \gamma_j y_{rj} = y_r + t_r + s_r^+, & r = 1, \ldots, s & (3.9)\\
 & \quad \displaystyle\sum_{j \in E_{VRS}} \gamma_j = 1, & & (3.10)\\
 & t_r,\ \lambda_j,\ y_r,\ s_r^+,\ \gamma_j \geq 0, & \forall r, j & (3.11)
\end{array}
\quad (3)
$$

Regarding the computation of (3), even for the Bilevel Programming problem where all the functions are linear, as in (3), the model to be solved is non-convex and NP-hard [27]. In this paper, as in Aparicio et al. [12], we resort to reformulate (3) as a one-level model, substituting the lower level problem by its corresponding Karush–Kuhn–Tucker (KKT) optimality conditions [72]. Accordingly, (3.7)–(3.10) must be substituted by (4.1)–(4.9).

$$
\begin{array}{lll}
\displaystyle\sum_{j \in E_{VRS}} \gamma_j x_{ij} + l_i = x_{i0}, & i = 1, \ldots, m & (4.1)\\
\displaystyle\sum_{j \in E_{VRS}} \gamma_j y_{rj} = y_r + t_r + s_r^+, & r = 1, \ldots, s & (4.2)\\
\displaystyle\sum_{j \in E_{VRS}} \gamma_j = 1, & & (4.3)\\
-\displaystyle\sum_{i=1}^{m} \eta_i x_{ij} + \sum_{r=1}^{s} \mu_r y_{rj} + \psi + \tau_j = 0, & j \in E_{VRS} & (4.4)\\
-\eta_i + e_i = 0, & i = 1, \ldots, m & (4.5)\\
\mu_r \geq 1, & r = 1, \ldots, s & (4.6)\\
\gamma_j \tau_j = 0, & j \in E_{VRS} & (4.7)\\
l_i e_i = 0, & i = 1, \ldots, m & (4.8)\\
l_i,\ \eta_i,\ \mu_r,\ \tau_j,\ e_i \geq 0, & \forall i, j & (4.9)
\end{array}
\quad (4)
$$

Constraints (4.7)–(4.8) are not linear. Nevertheless, constraints of this nature are not difficult to be implemented through Special Ordered Sets (SOS)[4] [17].

Finally, as we are interested in exploring potential contextual or environmental variables that might affect the performance of the evaluated schools, in a second stage we estimate different Tobit regression models. This approach has been extensively used by in previous literature to analyze dependent variables subject to a known upper or lower bound like our non-radial measures. We are aware that this conventional regression model has been criticized by Simar and Wilson [73], who argue that it might yield biased estimations because the efficiency scores estimated in the first stage are serially correlated. They address this issue by proposing two algorithms that incorporate the bootstrap procedure in a truncated regression model that allow for valid inference while simultaneously generating standard errors and confidence intervals for the efficiency estimates conventional regression methods like Tobit.[5]

The main problem of this approach is that it only considers the radial term of the efficiency in the second stage, thus we decided to use Tobit regressions in our empirical analysis. This choice is supported by two main reasons. First, the consistency of the estimation in a second-stage regression increases as the sample size enlarges, thus the potential bias disappears in large-size samples [88]. Since in our empirical application we have more than 10,000 observations, we consider that our estimates can be consistent. Second, our dataset contains five different plausible values drawn from the estimated distribution of results of each school. Therefore, we are already working with a certain confidence interval from the beginning and there would be no need to apply methods based on resampling like bootstrapping [34]. The estimated results from the Tobit regressions are extremely useful to identify what exogenous variables have a significant influence on both the average efficiency measure of performance and the output-specific score as well as whether the direction of this influence is positive or negative.

3. Data and variables

This section includes an empirical illustration with real data by applying the methodology proposed in the previous section. In particular, we use comparative data about schools operating in the 34 OECD countries participating in the OECD's PISA (*Programme for International Student Assessment*) 2012 survey. The survey takes place every three years, starting in 2000, thus PISA 2012 represents the fifth wave of this study. For each assessment, one of reading, mathematics and science is chosen as the major domain and given greater emphasis. In 2012, the major domain was mathematics (as well as in 2003) and two additional competences were also assessed for the first time (problem solving and financial literacy).

This survey uses a two-stage stratified design sampling [83]. In the first stage of sampling, schools having age-eligible students are sampled systematically with probabilities proportional to the school size. A minimum of 150 schools is selected in each country. Subsequently, 35 15 year-old students are randomly selected from each school to participate in the survey. Data were collected between March and May 2012 for countries in the northern hemisphere and May-August 2012 for countries in the southern hemisphere.

One of the main advantages of using PISA data is that this study does not evaluate cognitive abilities or skills through using one single score but each student receives five different scores (plausible

[4] SOS is a way to specify that a pair of variables cannot take strictly positive values at the same time and is a technique related to using special branching strategies. Traditionally, SOS was used with discrete and integer variables, but modern optimizers, like for example CPLEX, use also SOS with continuous variables.

[5] See Simar and Wilson [73,74] for a technical description of the model.

Table 2
Dataset composition: number of schools in each country.

Country	Schools	Country	Schools
Australia	775	Japan	191
Austria	191	Korea	156
Belgium	287	Luxembourg	42
Canada	885	Mexico	1471
Chile	221	Netherlands	179
Czech Republic	297	New Zealand	177
Denmark	341	Norway	197
Estonia	206	Poland	184
Finland	311	Portugal	195
France	226	Slovak Republic	231
Germany	230	Slovenia	338
Greece	188	Spain	902
Hungary	204	Sweden	209
Iceland	134	Switzerland	411
Ireland	183	Turkey	170
Israel	172	United Kingdom	507
Italy	1194	USA	162
TOTAL		11,767	

values) that represent the range of abilities that a student might reasonably have (see [65] for details). Specifically, the dataset provides measures on students' performance based upon pupils' responses to different test booklets, each of each of which includes only a limited number of test questions. Thus, it is difficult to make claims about individual performance with great accuracy. Using a complex process based on item response theory model, the survey organizers produce test scores for participants taking into account the difficulty of each test question.[6] Plausible values can therefore be defined as random values drawn from this distribution of proficiency estimates [60,86].

In addition, the survey collects a great volume of data about other factors potentially related to those results, such as variables representing student's background, school environment or educational provision. This information comes from the responses given to different questionnaires completed by students and school principals. From these data, it is possible to extract a great amount of information referred to the main determining factors of educational performance.

Our final dataset comprises a total number of 11,767 schools distributed across countries as reported in Table 2. As explained above, the minimum number of participating schools in each country must be 150, although in our sample we have the exceptional case of Luxembourg, where there are only 42 participating schools. Likewise, in several countries the sample is very large due to the existence of representative samples for different regions within the country (e.g. Australia, Canada, Italy, Mexico, Spain).

The output variables are represented by the average test scores achieved by students belonging to the same school in the two most relevant competences: reading and mathematics. Those test scores are obtained from the five plausible values, which have an international mean of 500 and standard deviation of 100. Following the recommendations made by survey organizers [63], our empirical analyses have been performed independently on each of these five plausible values. Due to space restrictions, we only present the estimations obtained with that first value, although the results of the empirical analysis are quite similar to the others.

The selection of inputs is a tough decision, since in the dataset there is 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 measures of human and capital resources [35,84]. In our empirical study we use the inverse of the student–teacher ratio, i.e., the number of teachers per (hundred) students (TEACHERS) and an index representing the quality of school resources (SCMATEDU) created by PISA analysts from the responses given by school principals regarding several aspects (computers, educational software, calculators, books, audio visual resources or laboratory equipment). Moreover, we also consider the average socio-economic status of students in the school (ESCS) as an additional input, since students are the "raw material" to be transformed through the learning process.[7] ESCS stands for the Economic, Social and Cultural Status, and provides a measure of family background that includes the highest levels of parents' occupation, educational resources and cultural possessions at home. Since the original values of SCMATEDU and ESCS presented positive and negative values, all of them were rescaled to show positive values.[8]

Finally, we have also selected some contextual variables in order to explore whether the performance of the evaluated units may be affected by the educational environment or the type of school management. In the following lines we provide a brief explanation about each of the variables selected and the expected direction of its influence according to previous literature.

- School ownership. Recent literature provides some empirical studies assessing whether the public or private nature of the school may affect their level of efficiency. Regarding this issue, in the literature we can find evidence that supports the idea of better performance in private schools [33,59] while others do not find enough evidence to justify this superiority [46,58]. In our case, we have included this information using a dummy variable taking a value of one if the school is private and zero if it is public.
- School location. This is a common factor usually considered as a potential factor affecting the efficiency of schools (see [35]), since whether a school is located in a main urban area or in a less densely populated area may affect its scale of operations and/or its ability to attract teaching staff. In order to take this into account we have defined a dummy variable that takes value one if the institution is located in a small town with less than 15,000 inhabitants.
- Percentage of girls. Various studies have demonstrated that girls usually perform better than boys in reading [62], while in mathematics the situation is the opposite [54]. Since we are interested in exploring whether some schools may concentrate their effort on one competence, this percentage might be a relevant factor in order to explain those potential behaviors.
- Percentage of repeaters. Different meta-analyses and literature reviews about the practice of retaining students have concluded that it has a negative effect on achievement (e.g. [52,53,87]). In our framework, we explore whether schools having a higher proportion of retained students may be more efficient.
- Competition with (at least) another school in the same neighborhood. The reference framework is the idea that the presence of more schools in a certain area should raise the performance of schools operating in that area as a response to pressures from nearby competitors [2,61], thus we expect a positive influence of this dummy variable on efficiency levels.

Moreover, we also take into account several variables representing students´attitudes toward mathematics, the domain about which there is more available information, in order to test whether the orientation of students toward one subject might be relevant in

[6] See Von Davier and Sinharay [82] for further details.

[7] This is a common practice in several recent papers attempting to measure the efficiency of schools (e.g. [3,4,33,70,77]).

[8] The rescaling process was made by adding up the minimum value to all the original values of the variables. This transformation does not alter the efficient frontier (or empirical production function) and hence the associated DEA model is translation invariant.

Table 3
Descriptive statistics of variables included in the analysis.

Variable	Type	Mean	Std. dev.	Min	Max
READING	Output	480.15	68.49	98.23	782.37
MATHS	Output	482.00	70.06	158.39	734.68
ESCS	Input	4.27	0.76	0.01	6.09
SCMATEDU	Input	3.59	1.04	0.008	5.576
TEACHERS	Input	9.76	16.01	0.098	1075.27
School factors					
PRIVATE	Contextual	0.169	0.374	0	1
RURAL	Contextual	0.328	0.469	0	1
PCGIRLS	Contextual	0.479	0.179	0	1
REPEATERS	Contextual	0.160	0.230	0	1
COMPETITION	Contextual	0.725	0.447	0	1
MATHHOMEWORK	Contextual	0.391	0.150	0	1
MATHEXAMS	Contextual	0.440	0.142	0	1
MATHCLASSES	Contextual	0.519	0.134	0	1
ENJOYMATHS	Contextual	0.261	0.136	0	1
READPLEASURE	Contextual	0.323	0.072	0.207	0.582

order to explain potential behaviors of schools concentrating their effort in that subject or another one. Specifically, we select four representative variables: (i) proportion of students that work hard on math homework[9]; (ii) proportion of students that declare to be prepared for maths exams; (iii) proportion of students that usually pay attention in maths classes[10]; and (iv) proportion of students that declare to enjoy maths. Finally, in order to supplement the extensive information available about activities related to mathematics, we have also retrieved some information about reading habits represented by the percentage of students reading for enjoyment at least 30 min a day.[11] Table 3 reports the descriptive statistics for all these variables (outputs, inputs and contextual factors).

4. Results

In this section we report the results obtained applying the closest target approach to the entire sample of schools across all countries. As explained in Section 2, this methodology allows us to determine projection points on the frontier for each output without assuming that they should be equiproportional. However, initially we present them in the form of the traditional non-radial Russell measures, i.e. as the average of the proportional rates of output expansion, so that they can be easier to interpret and compare with other empirical studies using traditional radial measures. Specifically, in Table 4 we show the average values of those "artificial" measures summarized by country. It is worth noting that those scores are presented in the form of Farrell output-oriented efficiency measures, thus values equal to one indicate that a unit is efficient, while values greater to one reflect the percentage of inefficiency.

The mean efficiency score *a la* Russell of the entire sample is 1.367, although the mean values by countries vary substantially, ranging from an inefficiency level of 20% for Korean schools to 52% for Slovenian ones. If we compare those values with average test scores in maths and reading for each country, also shown in Table 4, we observe that there is not always a straightforward relationship between results and efficiency. For instance, Turkey is among the best performers in terms of (aggregate) efficiency

Table 4
Efficiency scores versus average results in maths and reading.

Country	Maths	Reading	Efficiency score
Korea	551	534	1.2032
Japan	534	535	1.2574
Turkey	437	461	1.2758
Poland	527	527	1.2833
Estonia	518	514	1.3003
Ireland	498	520	1.3034
Netherlands	517	505	1.3082
New Zealand	497	511	1.3106
Canada	507	509	1.3193
Germany	507	500	1.3300
Mexico	410	419	1.3433
USA	479	496	1.3433
Czech Rep.	500	496	1.3446
Spain	491	490	1.3487
Finland	508	511	1.3491
Switzerland	514	492	1.3500
UK	488	496	1.3552
Australia	492	500	1.3683
Portugal	479	480	1.3690
Belgium	508	500	1.3763
France	489	498	1.3768
Norway	491	505	1.3911
Austria	488	475	1.4051
Denmark	484	482	1.4054
Luxembourg	490	486	1.4085
Sweden	483	486	1.4106
Chile	432	446	1.4118
Hungary	461	472	1.4167
Italy	476	476	1.4306
Slovak Rep.	473	450	1.4336
Israel	464	484	1.4383
Iceland	488	478	1.4386
Greece	440	461	1.4985
Slovenia	461	438	1.5223
MEAN	488	489	1.3668

despite their results being relatively poor. Something similar also occurs for Mexico, which is placed in a high position in the rank of countries according to the efficiency of their schools despite their schools have the worst results in both competences. In contrast, some countries with relatively good results in both reading and math are placed in a worst position in terms of efficiency (e.g. Finland or Norway), which indicates that schools operating in those countries are not sufficiently exploiting their resources.

Figs. 2 and 3 illustrate that there is a negative relationship between efficiency and performance in maths and reading across countries. In general terms, countries where the average PISA score is higher, also tend to show lower levels of inefficiency. Likewise, it is also worth noting that some countries with similar average efficiency scores have very different levels of performance (see the case of Portugal and Finland).

Even more interesting than this evidence, which can also be derived using a traditional DEA approach, we would like to explore the projection (ratio between the target value and the actual value) for each output calculated with the proposed model developed by Aparicio et al. [12] before calculating the average output expansions. According to the values shown in Table 5, there are divergences between reading and maths projection values that are not visible when only average values are reported. Thus, for example, schools operating in Turkey, Chile and Mexico present significantly higher values in maths, which indicates that they need to make a greater effort to improve results in this area, since they are performing relatively well in reading. Turkish schools in fact present the lower averaged projection values among all the OECD countries in reading. The opposite occurs in Slovenia, Slovak Republic or Switzerland, where schools present significantly higher

[9] A large number of studies have generated conflicting findings about effects of homework on educational achievement [29,80].

[10] These indicators can be interpreted as proxies of student engagement, which has been traditionally identified as a key factor associated with better student achievement [26].

[11] Since PISA 2012 does not include specific information about reading activities and attitudes, we have drawn this information from PISA 2009. Given that the participating schools were not the same in that year, we have calculated those indicators aggregated at country level.

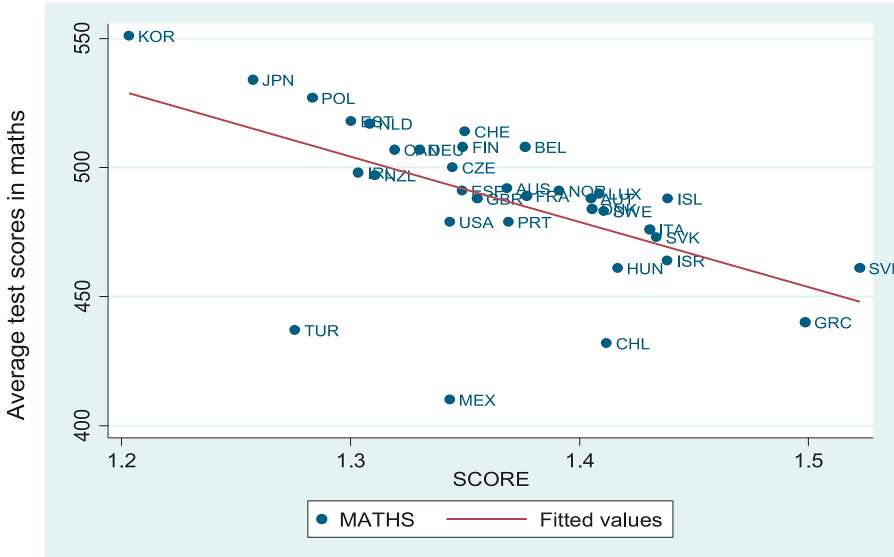

Fig. 2. Relationship between efficiency and maths across countries.

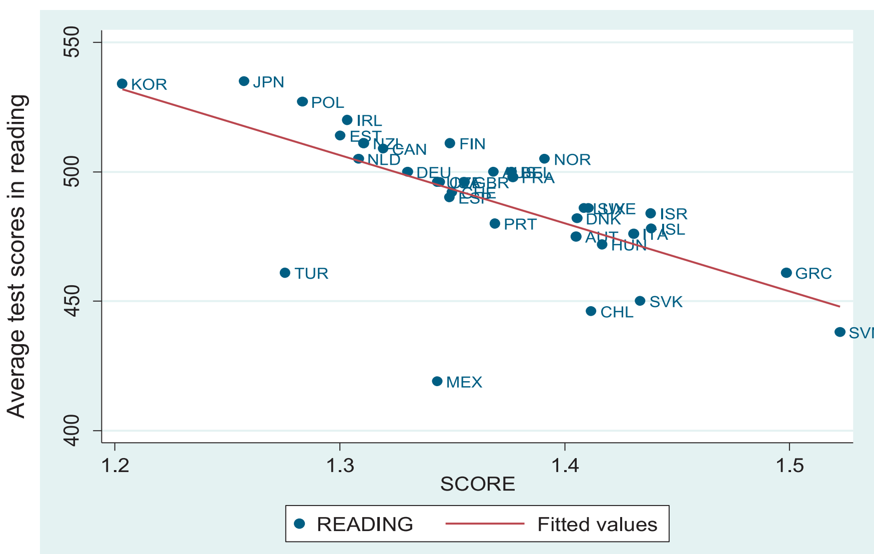

Fig. 3. Relationship between efficiency and reading across countries.

values in their projections for reading than in maths. Actually, this divergence is above 10 points, so the distance to the frontier for schools operating in those countries are, on average, 10% further in the reading competency.

The possibility of measuring the efficiency of schools using an international benchmark and considering both radial and non-radial displacements to measure the distance to the frontier allows for identifying that some schools might focus more on one dimension of the educational outcome. In particular, according to shape of the graph shown in Fig. 4, in global terms there are higher inefficiencies in reading. Indeed, 26 out of the 34 OECD countries present higher values in this competency and, in most cases, they are higher than 5%. Therefore, it seems that most part of schools around the world present higher efficiency levels in math proficiency. Nevertheless, there are some exceptions. In addition to the countries mentioned above (Turkey, Chile and Mexico), which are the poorest countries in the sample, some of the richest countries such as Norway or United States also present higher values of inefficiency in maths.

Finally, with the aim of exploring some potential characteristics of countries or schools that might explain why some schools present higher levels of inefficiency in maths or reading we have estimated two different Tobit regression models, one for each output.[12] Specifically, Table 6 reports the estimation results from the regression using math expansion as dependent variable and Table 7 reports the estimates for reading expansion as dependent variable. Since both measures are bounded from below at the value of one, a positive (negative) coefficient suggests a negative (positive) impact of the corresponding exogenous variable. Moreover, a coefficient is only statistically significant when both the lower

[12] In our estimation we used data about the five efficiency scores obtained with each of the five plausible values of the output variables in order to obtain more consistent results. Specifically, we follow the procedure recommended by survey organizers, which consists of estimating each regression model five times (once using each of the plausible values), thus we have five separate parameter estimates and five estimates of the sampling error and, subsequently, we take the average of the five estimates (See [65], p. 44 for details).

Table 5
Projections for each output dimension.

Country	Average efficiency score	Maths (M)	Reading (R)	Difference (R-M)
Korea	1.2032	1.1702	1.2362	6.6010
Japan	1.2574	1.2311	1.2838	5.2645
Turkey	1.2758	1.3206	1.2311	−8.9429
Poland	1.2833	1.2569	1.3097	5.2763
Estonia	1.3003	1.2709	1.3297	5.8840
Ireland	1.3034	1.3049	1.3020	−0.2934
Netherlands	1.3082	1.2738	1.3426	6.8833
New Zealand	1.3106	1.3022	1.3190	1.6772
Canada	1.3193	1.3029	1.3357	3.2730
Germany	1.3300	1.3004	1.3596	5.9190
Mexico	1.3433	1.3778	1.3087	−6.9126
USA	1.3433	1.3517	1.3349	−1.6784
Czech Rep.	1.3446	1.3144	1.3748	6.0459
Spain	1.3487	1.3163	1.3810	6.4710
Finland	1.3491	1.3467	1.3516	0.4875
Switzerland	1.3500	1.2974	1.4026	10.5229
UK	1.3552	1.3486	1.3617	1.3109
Australia	1.3683	1.3534	1.3832	2.9825
Portugal	1.3690	1.3285	1.4095	8.1024
Belgium	1.3763	1.3410	1.4115	7.0555
France	1.3768	1.3505	1.4031	5.2527
Norway	1.3911	1.4110	1.3712	−3.9798
Austria	1.4051	1.3777	1.4324	5.4627
Denmark	1.4054	1.3801	1.4306	5.0557
Luxembourg	1.4085	1.3814	1.4356	5.4181
Sweden	1.4106	1.3879	1.4333	4.5357
Chile	1.4118	1.4545	1.3692	−8.5327
Hungary	1.4167	1.3981	1.4352	3.7096
Italy	1.4306	1.3957	1.4655	6.9814
Slovak Rep.	1.4336	1.3649	1.5024	13.7524
Israel	1.4383	1.4506	1.4260	−2.4580
Iceland	1.4386	1.4433	1.4339	−0.9408
Greece	1.4985	1.4915	1.5055	1.4084
Slovenia	1.5223	1.4488	1.5957	14.6966
TOTAL	1.3655	1.3435	1.3780	

Table 6
Determinants of total inefficiency in maths (Tobit regression).

Variable	Coefficient	SE	(95% Confidence interval)	
			Low	High
PRIVATE	−0.0252***	0.0040	−0.0421	−0.0261
RURAL	0.00288	0.0034	0.0037	0.0170
PCGIRLS	−0.0237***	0.0083	−0.1091	−0.0762
REPEATERS	0.282***	0.0066	0.3049	0.3312
COMPETITION	−0.0177***	0.0037	−0.0281	−0.0135
MATHHOMEWORK	0.0871***	0.0131	0.0145	0.0661
MATHEXAMS	−0.199***	0.0135	−0.2277	−0.1742
MATHCLASSES	−0.134***	0.0149	−0.2037	−0.1450
ENJOYMATHS	−0.0444***	0.0128	−0.0514	−0.0008
READPLEASURE	−0.0378	0.0265	0.0110	0.1156

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Table 7
Determinants of total inefficiency in reading (Tobit regression).

Variable	Coefficient	SE	(95% Confidence interval)	
			Low	High
PRIVATE	−0.0431***	0.0047	−0.0523	−0.0339
RURAL	0.0180***	0.0039	0.0103	0.0257
PCGIRLS	−0.162***	0.0097	−0.1807	−0.1427
REPEATERS	0.355***	0.0077	0.3393	0.3697
COMPETITION	−0.0239***	0.0043	−0.0323	−0.0155
MATHHOMEWORK	0.0053	0.0152	−0.0351	0.0245
MATHEXAMS	−0.204***	0.0158	−0.2347	−0.1728
MATHCLASSES	−0.216***	0.0173	−0.2495	−0.1816
ENJOYMATHS	−0.0099	0.0149	−0.0392	0.0193
READPLEASURE	0.1458***	0.0308	−0.0393	0.0816

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

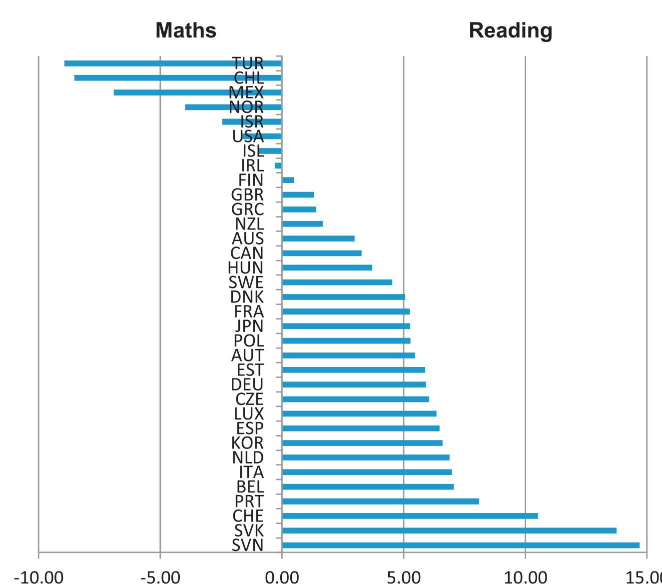

Fig. 4. Average levels of inefficiency per competency (R – M).

bound and upper bound of the confidence interval have the same sign.

The results shown in Tables 6 and 7 reveal that school factors are almost identically associated with both dependent variables, thus it seems that the aforementioned existing divergences between reading and mathematics across schools are not related to the school environment. Thus, in both cases being a private school, having a higher proportion of girls and having competition are positively associated with efficiency measures, whereas schools placed in a rural area and, more importantly, having a higher proportion of repeaters are negatively related to efficiency in both competences.[13]

In contrast, we observe some differences in variables representing attitudes toward math and reading. For instance, reading for pleasure is a key factor as a determinant of the reading expansion, but we find no relationship between this variable and math efficiency. Nevertheless, these results should be interpreted cautiously because this information was retrieved at a country level due to the lack of information related to reading activities in the original dataset. Similarly, we find that math enjoyment affects the expansion in maths, but it is not related to reading efficiency. These results reveal the existence of a certain substitution effect between subjects with regard to the preferences of students that might affect their performance and, subsequently, the efficiency demonstrated by schools with regard to each dimension of the output. Another divergence arises for working hard on math homework, which is found to be negatively associated with math efficiency, but it is not a significant factor for reading. One explanation of the negative relation is that students with lower levels of proficiency in reading tend to put more time into homework because of necessity and pressure from parents and teachers [50]. This result suggests that assigning more homework to students does not seem to be the appropriate strategy that school should apply to achieve higher levels of efficiency neither in mathematics nor in reading.

Finally, we find that both indicators related to school engagement, i.e. being prepared for math exams and paying attention in maths classes, are positively and significantly related to both dependent variables. Although the degrees of engagement depends in a certain extent of intrinsic motivation of students, schools have the capacity of influencing this type of behaviors, thus promoting student engagement appears to be a good way to be more efficient.

5. Concluding remarks

In this paper we have used data from OECD countries participating in PISA 2012 to assess the efficiency of schools in a cross-country framework. In our empirical analysis we consider that schools might concentrate their efforts on improving more the results in one dimension of the educational output than in other. To do this, we rely on non-radial efficiency measures of performance, which are particularly interesting in the context of education, since they allow for identifying different levels of (in) efficiency for each output analyzed. In particular, we apply a methodology recently developed by Aparicio et al. [12] that determines the closest targets and the least distance to the strongly efficient frontier in DEA based on Bilevel Linear Programming.

Although it is true that many alternative non-radial measures exist in DEA, those based on the determination of the least distance yield useful benchmarking information. In particular, this type of measures can be useful from the point of view of practice, for example to managers in their decision making. It is especially important for firms/units which seek to achieve superior performance results as soon as possible. Indeed, the distance which forms the basis of the technical inefficiency measure in this paper generates targets that are easily achievable by units. Among the set of non-radial DEA measures associated with the determination of the least distance, we opt for the output-oriented Russell measure. Nevertheless, we recognize that other measures based on least distances could be also used in the empirical application. The evaluation of the impact of the utilization of different measures is outside the scope of this paper, although it could be a good avenue for further research.

Our findings indicate that larger potential improvements may be achieved in reading proficiency than in maths. In global terms we detect higher levels of efficiency in the reading competency, thus it seems that schools around the world are concentrating their efforts in maths, maybe because math proficiency is generally considered as one of the strongest predictors of positive outcomes for young adults, such as their ability to participate in post-secondary education and their expected future earnings [64]. This might entail neglecting reading. Actually, around three quarters of the studied countries present greater inefficiency values in this competency and, in most cases, this divergence exceeds 5%. Even so, there are some exceptions, such as schools operating in poor countries like Turkey, Chile or Mexico, which perform relatively better in reading than in maths.

We have also examined some potential determinants associated with each dimension of efficiency by estimating one regression for each output projection. The results reveal that divergences detected between the two dimensions of the output are mainly explained by students´attitudes, while school characteristics do not appear to be differential factors. In particular, we identify divergences with regard to the reading habits of students as well as homework assignments, whereas variable representing student engagement has a positive effect on both educational outcomes. However, this approach does not allow for a causal interpretation of results, but it allows a future line of research based on the search of the causes of inefficiency to be flagged up.

These findings provide some interesting insights into the analysis of determinants of educational attainment using a cross-country approach. However, further future research will be needed to explore some of the results discussed here in greater depth. For instance, the proposed analysis could be replicated using data at pupil level in order to test whether the origin of the existing divergence in terms of inefficiency between subjects might come from intrinsic characteristics of the students that affect their performance instead of the activities carried out by schools.

Acknowledgments

Juan Aparicio and Martin Gonzalez thank the financial support from the Spanish Ministry for Economy and Competitiveness (Ministerio de Economía, Industria y Competitividad), the State Research Agency (Agencia Estatal de Investigacion) and the European Regional Development Fund (Fondo Europeo de DEsarrollo Regional) under grant MTM2016-79765-P (AEI/FEDER, UE). Additionally, Jose M. Cordero also acknowledges the support from the Spanish Ministry for Economy and Competitiveness through grant ECO2014-53702-P and Junta de Extremadura through grant IB16171.

References

[1] Afonso A, Aubyn MS. Cross-country efficiency of secondary education provision: a semi-parametric analysis with non-discretionary inputs. Econ Model 2006;23(3):476–91.
[2] Agasisti T. The efficiency of Italian secondary schools and the potential role of competition: a data envelopment analysis using OECD-PISA2006 data. Educ Econ 2013;21(5):520–44.
[3] Agasisti T. The efficiency of public spending on education: an empirical comparison of EU countries. Eur J Educ 2014;49(4):543–57.
[4] Agasisti T, Zoido P. *The efficiency of secondary schools in an international perspective: preliminary results from PISA 2012* (No. 117). OECD Publishing; 2015.
[5] Ando K, Kai A, Maeda Y, Sekitani K. Least distance based inefficiency measures on the pareto-efficient frontier in DEA. J Oper Res Soc Jpn 2012;55:73–91.
[6] Ando K, Minamide M, Sekitani K. Monotonicity of minimum distance inefficiency measures for data envelopment analysis. Eur J Oper Res 2017;260(1):232–43.
[7] Aparicio J, Ruiz JL, Sirvent I. Closest targets and minimum distance to the pareto-efficient frontier in DEA. J Prod Anal 2007;28:209–18.

[13] The negative relationship of repeaters might explain why countries like Japan or Korea, where the practice of retention is almost inexistent (see [68]), are among the top performers in terms of both achievement and efficiency.

[8] Aparicio J, Pastor JT. A well-defined efficiency measure for dealing with closest targets in DEA. Appl Math Comput 2013;219:9142–54.
[9] Aparicio J, Mahlberg B, Pastor JT, Sahoo BK. Decomposing technical inefficiency using the principle of least action. Eur J Oper Res 2014;239:776–85.
[10] Aparicio J, Pastor JT. Closest targets and strong monotonicity on the strongly efficient frontier in DEA. Omega 2014;44:51–7.
[11] Aparicio J, Pastor JT. On how to properly calculate the Euclidean distance-based measure in DEA. Optimization 2014;63(3):421–32.
[12] Aparicio J, Cordero JM, Pastor JT. The determination of the least distance to the strongly efficient frontier in data envelopment analysis oriented models: modelling and computational aspects. Omega 2016. http://dx.doi.org/10.1016/j.omega.2016.09.008i.
[13] Aparicio J. A survey on measuring efficiency through the determination of the least distance in data envelopment analysis. J Centr Cath 2016;9(2):143–67.
[14] Aristovnik A, Obadić A. Measuring relative efficiency of secondary education in selected EU and OECD countries: the case of Slovenia and Croatia. Technol Econ Dev Econ 2014;20(3):419–33.
[15] Asmild M, Hougaard JL, Kronborg D, Kvist HK. Measuring Inefficiency via potential improvements. J Prod Anal 2003;19(1):59–76.
[16] Banker RD, Charnes A, Cooper WW. Some models for estimating technical and scale inefficiencies in data envelopment analysis. Manage Sci 1984;30:1078–92.
[17] Beale EML, Tomlin JA. Special facilities in a general mathematical programming system for non-convex problems using ordered sets of variables. In: Lawrence J, editor. Proceedings of the fifth international conference on operational research. Tavistock Publications; 1970. p. 447–54.
[18] Bogetoft P, Hougaard JL. Efficiency evaluations based on potential (non-proportional) improvements. J Prod Anal 1999;12:233–47.
[19] Bray M, Thomas RM. Levels of comparison in educational studies: different insights from different literatures and the value of multilevel analyses. Harvard Educ Rev 1995;65(3):472–90.
[20] Briec W. Minimum distance to the complement of a convex set: duality result. J Optim Theory Appl 1997;93:301–19.
[21] Briec W. Hölder distance function and measurement of technical efficiency. J Prod Anal 1998;11:111–31.
[22] Chambers R, Chung Y, Färe R. Benefit and distance functions. J Econ Theory 1996;70:407–19.
[23] Chambers RG, Mitchell T. Homotheticity and non-radial changes. J Prod Anal 2001;15 31-19.
[24] Charnes A, Cooper WW, Rhodes E. Measuring the efficiency of decision making units. Eur J Oper Res 1978;2(6):429–44.
[25] Charnes A, Cooper WW, Golany B, Seiford L, Stutz J. Foundations of data envelopment analysis for Pareto–Koopmans efficient empirical production functions. J Econometr 1985;30:91–107.
[26] Christenson SL, Reschly AL, Wylie C. Handbook of research on student engagement. Springer Science & Business Media; 2012.
[27] Colson B, Marcotte P, Savard G. An overview of bilevel optimization. Ann Oper Res 2007;153:235–56.
[28] Cook WD, Hababou M. Sales performance measurement in bank branches. Omega 2001;29(4):299–307.
[29] Cooper H, Robinson JC, Patall EA. Does homework improve academic achievement? A synthesis of research, 1987–2003. Rev Educ Res 2006;76:1–62.
[30] Cordero JM, Muñiz M, Polo C. The determinants of cognitive and non-cognitive educational outcomes: empirical evidence in Spain using a Bayesian approach. Appl Econ 2016;48(35):3355–72.
[31] Cordero JM, Santín D, Simancas R. Assessing European primary school performance through a conditional nonparametric model. J Oper Res Soc 2017;68(4):364–76.
[32] Creemers B, Kyriakides L. The dynamics of educational effectiveness: a contribution to policy, practice and theory in contemporary schools. Abingdon, Oxon: Routledge; 2008.
[33] Crespo-Cebada E, Pedraja-Chaparro F, Santín D. Does school ownership matter? An unbiased efficiency comparison for regions of Spain. J Prod Anal 2014;41(1):153–72.
[34] De Jorge J, Santín D. Determinantes de la eficiencia educativa en la Unión Europea. Hacienda Pública Española 2010(193):131–55.
[35] De Witte K, López-Torres L. Efficiency in education. A review of literature and a way forward. J Oper Res Soc 2017;68(4):339–63.
[36] Deprins D, Simar L, Tulkens H. Measuring labor inefficiency in post offices. In: Marchand M, Pestieau P, Tulkens H, editors. The performance of public enterprises: concepts and measurements. Amsterdam, North-Holland; 1984. p. 243–67.
[37] Deutsch J, Dumas A, Silber J. Estimating an educational production function for five countries of Latin America on the basis of the PISA data. Econ Educ Rev 2013;36:245–62.
[38] Färe R, Grosskopf S, Lovell CAK. The measurement of efficiency of production. Kluwer Nijhof Publishing; 1985.
[39] Farrell MJ. The measurement of productive efficiency. J R Stat Soc, Ser A: Gen 1957;120:253–81.
[40] Fukuyama H, Maeda Y, Sekitani K, Shi J. Input-output substitutability and strongly monotonic p-norm least-distance DEA measures. Eur J Oper Res 2014;237:997–1007.
[41] Fukuyama H, Masaki H, Sekitani K, Shi J. Distance optimization approach to ratio-form efficiency measures in data envelopment analysis. J Prod Anal 2014;42:175–86.
[42] Fukuyama H, Hougaard JL, Sekitani K, Shi J. Efficiency measurement with a nonconvex free disposal hull technology. J Oper Res Soc 2016;67(1):9–19.
[43] Giambona F, Vassallo E, Vassiliadis E. Educational systems efficiency in European Union countries. Stud Educ Eval 2011;37(2):108–22.
[44] Giménez V, Prior D, Thieme C. Technical efficiency, managerial efficiency and objective-setting in the educational system: an international comparison. J Oper Res Soc 2007;58(8):996–1007.
[45] Giménez V, Thieme C, Prior D, Tortosa-Ausina E. An international comparison of educational systems: a temporal analysis in presence of bad outputs. J Prod Anal 2017;47(1):83–101.
[46] Goldhaber DD. Public and private high schools: Is school choice an answer to the productivity problem? Econ Edu Rev 1996;15(2):93–109.
[47] Grifell-Tatje E, Lovell CAK, Pastor JT. A quasi-malmquist productivity index. J Prod Anal 1998;10:7–20.
[48] Grosskopf S, Hayes KJ, Taylor LL. Efficiency in education: research and implications. Appl Econ Perspect Policy 2014;36(2):175–210.
[49] Gustafsson JE. Effects of international comparative studies on educational quality on the quality of educational research. Eur Educ Res J 2008;7(1):1–17.
[50] Gustafsson JE. Causal inference in educational effectiveness research: a comparison of three methods to investigate effects of homework on student achievement 1. School Eff School Improv 2013;24(3):275–95.
[51] Hanushek EA. Conceptual and empirical issues in the estimation of educational production functions. J Human Resour 1979;14:351–88.
[52] Holmes CT. Grade-level retention effects. A meta-analysis of research studies. In: Shepard L, Smith ML, editors. Flunking grade: research and policies on retention. London: The Falmer Press; 1989. p. 16–33.
[53] Huddleston AP. Achievement at whose expense? A literature review of test-based grade retention policies in U.S. school. Educ Policy Anal Arch 2014;22(18):1–31.
[54] Hyde JS, Fennema E, Lamon SJ. Gender differences in mathematics performance: a meta-analysis. Psychol Bull 1990;107(2):139.
[55] Johnes J. Operational research in education. Eur J Oper Res 2015;243(3):683–96.
[56] Koopmans TC. Analysis of production as an efficient combination of activities. Activity analysis of production and allocation. Koopmans TC, editor. New York: John Wiley; 1951.
[57] Levin H. Measuring the efficiency in educational production. Public Finance Q 1974;2:3–24.
[58] Mancebon MJ, Muñiz MA. Private versus public high schools in Spain: disentangling managerial and programme efficiencies. J Oper Res Soc 2008;59(7):892–901.
[59] McEwan PJ. The effectiveness of public, catholic, and non-religious private schools in Chile's voucher system. Educ Econ 2001;9(2):103–28.
[60] Mislevy RJ, Beaton AE, Kaplan B, Sheehan KM. Estimating population characteristics from sparse matrix samples of item responses. J Educ Meas 1992;29(2):133–61.
[61] Misra K, Grimes PW, Rogers KE. Does competition improve public school efficiency? A spatial analysis. Econ Educ Rev 2012;31(6):1177–90.
[62] Mullis IM, Kennedy A, Foy P. IEAś progress in international reading literacy study in primary schools in 40 countries PIRLS 2006 International Report. TIMMS and PIRLS International Study Center, Lynch School of Education, Boston College; 2007.
[63] OECD. PISA data analysis Manual, SPSS. 2nd ed. Paris: OECD Publishing; 2009.
[64] OECD. PISA 2012 Results: what students know and can do – student performance in mathematics, reading and science, vol. I. Paris: PISA, OECD Publishing; 2013.
[65] OECD. PISA 2012 technical report. Paris: PISA, OECD Publishing; 2014.
[66] Pastor JT, Ruiz JL, Sirvent I. An enhanced DEA Russell graph efficiency measure. Eur J Oper Res 1999;115(3):596–607.
[67] Pastor JT, Aparicio J. The relevance of DEA benchmarking information and the least-distance measure: comment. Math Comput Model 2010;52:397–9.
[68] Portela MCAS, Castro P, Thanassoulis E. Finding closest targets in non-oriented DEA models: the case of convex and non-convex technologies. J Prod Anal 2003;19:251–69.
[69] Ruiz JL, Sirvent I. Common benchmarking and ranking of units with DEA. Omega 2016;65:1–9.
[70] Santín D, Sicilia G. Measuring the efficiency of public schools in Uruguay: main drivers and policy implications. Latin Am Econ Rev 2015;24(1):1–28.
[71] Shephard RW. Cost and production functions. Princeton, N.J.: Princeton University Press; 1953.
[72] Shi C, Lu J, Zhang G. An extended Kuhn–Tucker approach for linear bilevel programming. Appl Math Comput 2005;162:51–63.
[73] Simar L, Wilson PW. Estimation and inference in two-stage, semi-parametric models of production processes. J Econometr 2007;136(1):31–64.
[74] Simar L, Wilson PW. Two-stage DEA: caveat emptor. J Prod Anal 2011;36(2):205–18.
[75] Sutherland D, Price R, Gonand F. Improving public spending efficiency in primary and secondary education. OECD J: Econ Stud 2009;2009(1):1–30.
[76] Thieme C, Giménez V, Prior D. A comparative analysis of the efficiency of national educational systems. Asia Pacific Educ Rev 2012;13:1–15.
[77] Thieme C, Prior D, Tortosa-Ausina E. A multilevel decomposition of school performance using robust nonparametric frontier techniques. Econ Educ Rev 2013;32:104–21.
[78] Todd PE, Wolpin KI. On the specification and estimation of the production function for cognitive achievement. Econ J 2003;113(485):3–33.

[79] Tone K. A slacks-based measure of efficiency in data envelopment analysis. Eur J Oper Res 2001;130(3):498–509.
[80] Trautwein U, Köller O. The relationship between homework and achievement: still much of a mystery. Educ Psychol Rev 2003;15:115–45.
[81] Verhoeven M, Gunnarsson V, Carcillo S. Education and health in G7 countries: achieving better outcomes with less spending; 2007. IMF Working Papers p. 1–50.
[82] Von Davier M, Sinharay S. Analytics in international large-scale assessments: item response theory and population models. In: Rutkowski L, Von Davier M, Rutkowski D, editors. Handbook of international large-scale assessment: background, technical issues, and methods of data analysis. London: CRS Press; 2013. p. 155–74.
[83] Willms JD, Smith T. A manual for conducting analyses with data from TIMSS and PISA; 2005. Report prepared for UNESCO Institute for Statistics.
[84] Worthington AC. An empirical survey of frontier efficiency measurement techniques in education. Educ Econ 2001;9(3):245–68.
[85] Wu DD. Bilevel programming data envelopment analysis with constrained resource. Eur J Oper Res 2010;207:856–64.
[86] Wu M. The role of plausible values in large-scale surveys. Stud Educ Eval 2005;31(2-3):114–28.
[87] Xia N, Kirby SN. Retaining students in grade: a literature review of the effects of retention on students' academic and nonacademic outcomes; 2009. Technical Report No. 678.
[88] Xue M, Harker PT. Overcoming the inherent dependency of DEA efficiency scores: a bootstrap approach. Wharton Financial Institutions Center, University of Pennsylvania; 1999. *Unpublished Working Paper*.