

Dominguez-Gil, Carlota; Contreras, Ignacio

## Article

# A DEA-inspired model to evaluate the efficiency of education in OECD countries

Revista de Métodos Cuantitativos para la Economía y la Empresa

## Provided in Cooperation with:

Universidad Pablo de Olavide, Sevilla

*Suggested Citation:* Dominguez-Gil, Carlota; Contreras, Ignacio (2021) : A DEA-inspired model to evaluate the efficiency of education in OECD countries, Revista de Métodos Cuantitativos para la Economía y la Empresa, ISSN 1886-516X, Universidad Pablo de Olavide, Sevilla, Vol. 31, pp. 329-346,  
<https://doi.org/10.46661/revmetodoscuanteconempresa.4318>

This Version is available at:

<https://hdl.handle.net/10419/286232>

### Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

### Terms of use:

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by-sa/4.0/>



## A DEA-inspired model to evaluate the efficiency of education in OECD countries

DOMÍNGUEZ-GIL, CARLOTA

Department of Economics, Quantitative Methods and Economic History  
Pablo de Olavide University (Spain)  
Correo electrónico: [cdomgil@alu.upo.es](mailto:cdomgil@alu.upo.es)

CONTRERAS, IGNACIO

Department of Economics, Quantitative Methods and Economic History  
Pablo de Olavide University (Spain)  
Correo electrónico: [iconrub@upo.es](mailto:iconrub@upo.es)

### ABSTRACT

In this paper empirical application to the study about the efficiency of the performance of the educational systems across countries is developed. With the information published in the PISA 2015, Data Envelopment Analysis methodology is considered to evaluate the efficiency in the use of the resources devoted to education by OECD countries. Similar to previous studies, the main resources needed for learning, financial, human resources, material and time have been considered. Alternatively to previous proposals, the mean scores have not been included as the output of the process. Instead of that, to quantify the results of the learning process, the percentages of students in each proficiency level of the PISA test have been computed.

An ad hoc model based on the Additive DEA-model is proposed, adapting the formulation to the particular features of the vector of outputs considered. Considering that the aggregate value of output is fixed and that the relative weight of the outputs differs, inefficient units improve their performance by reallocating that fixed value among different outputs, moving units from the less valued to the most valued ones.

**Keywords:** education, PISA, DEA.

**JEL classification:** I21; C61.

**MSC2010:** 90B90.

# Un modelo inspirado en DEA para evaluar la eficiencia de la educación en los países de la OCDE

## RESUMEN

En el presente trabajo se presenta un modelo para el estudio de la eficiencia de los sistemas educativos de los países de la OCDE. A partir de la información publicada en el informe PISA 2015, se utiliza la metodología del Análisis Envolvente de Datos (DEA) para analizar la eficiencia en el uso de los recursos destinados a la educación en los países OCDE. Siguiendo la línea de anteriores estudios, se consideran los principales recursos destinados a la educación, esto es, recursos materiales, recursos humanos y tiempo dedicado a la enseñanza. De manera alternativa a los estudios anteriores, no se consideran las puntuaciones medias de los exámenes como las salidas del sistema. En nuestro estudio, la cuantificación de los resultados se realiza a través de los porcentajes de estudiantes que alcanzan cada nivel de desempeño en las pruebas normalizadas realizadas en PISA.

Se desarrolla un nuevo modelo de evaluación basado en el modelo aditivo dentro de la metodología DEA, en el que tanto la formulación como los objetivos se adaptan a las características de las variables propuestas. Considerando que el valor agregado de las salidas está fijado y que los pesos que deben asignarse a cada output deben estar ordenados, el modelo evalúa los posibles movimientos de outputs desde las categorías menos valoradas a las más valoradas.

**Palabras clave:** educación, PISA, DEA.

**Clasificación JEL:** I21; C61.

**MSC2010:** 90B90.



## 1. Introduction.

There is a recent and increasing debate at the developed countries about the relevance of controlling public expenses in education. On the one hand, based on the correlation between the economic growth and social development with the level of human capital, there is a clear incentive for an increasing investment in education. On the other hand, the economic crisis and public deficit in almost all countries impose the necessity of a best use of every coin invested in the educational system.

In this context, the concept of efficiency of educational systems becomes crucial. That is, governments are required to provide educational services by minimizing the amount of public resources (money) devoted to them. Or equivalently, they are required to obtain good results in terms of educational outputs with the available (fixed) resources.

From the point of view of the economics of education, education is seen as a production process in which diverse inputs are used to obtain multiple outputs for a given production technology. The theoretical approach of linking resources to educational outcomes at school level is based on the production function proposed in Levin (1974) and Hanushek (1979). For a particular school  $s$  the function is defined as follows:

$$A_s = f(B_s, S_s); \quad (1)$$

where  $A_s$  represents the educational output, normally measured through scores on standardised tests. It is clear that it is not an easy task to quantify the education received by an individual, due to its inherent intangibility and necessity to consider the quality beyond several years of study. However, there is a consensus in the literature about considering the results from a standardised test as educational outputs. They are difficult to forge and they are taken into account by policy makers and families when making decisions in education.

In (1) the inputs are divided into  $B_s$  and  $S_s$ , which denote the average student's family background and the educational resources assigned to school  $s$  respectively. Classically, they consider the main inputs required to carry out the learning process: raw material, physical and human capital.

Nevertheless, unlike other industries, education presents certain characteristics that hinder the estimation of a production function. Mancebón and Bandrés (1999) stress the intangible and multiple nature of the output, the time-lag in achieving its results, its cumulative nature and that the educational process is carried out by the customers themselves. This is why non-parametric techniques such as Data Envelopment Analysis (DEA) are so convenient to measure the efficiency in this context. They allow the assessment of the efficiency of the different units without having to estimate a production function.

DEA is a statistical technique used to evaluate the relative efficiency of a set of units developed in Charnes et al. (1978). By using linear programming a frontier of best-practice units is constructed based in observed data. The efficient frontier is used as a benchmark against which the performance of less efficiency units can be assessed. The estimated frontier envelops all the available observations, and each deviation from that frontier is interpreted as a measure of the inefficiency of the units. The DEA methodology has been widely used to analyse efficiency in several areas of public expenditure. The main reason for its widespread application is its flexibility, the fact that it accounts for multiple outputs, the uncertainty about true production technology and the lack of price information; making it well suited to the peculiarities of the public sector Santin and Sicilia (2015).

In DEA, efficiency is defined in a technical sense. That is to say, as the ability of transforming inputs into outputs for a given technology. The concept of efficiency was first contextualized in the field of education by Levin (1974) and has been widely used in the literature to evaluate efficiency in education. Although a complete literature review would require a specific research paper, some of the previous studies about the efficiency in education must be cited. In any case, a more detailed revision can be seen in Worthington (2001) and Johnes (2006).

This family of studies starts with Charnes et al. (1981), where the authors of the DEA methodology investigate the efficiency of an educational program in the USA. Since then, several works have continued the study of efficiency in the field of education. Afonso and Aubyn (2006a; 2006b), Sutherland et al. (2009) or Agasisti (2011), among others, considered international data to assess a comparison across countries. Examples of studies for a particular country are, for instance, Bessent and Bessent (1980), Bessent et al. (1982) or Agasisti (2011); in particular, Mancebón and Bandrés (1999) or Cordero-Ferrera et al. (2011) developed studies of the different types of school across the regions in Spain.

Works like Clemens (2002), Aristovnik (2011) or Agasisti (2014) apply DEA for the study of efficiency placing the emphasis on the educational spendings. Other related papers, introduced new elements into the analysis. That is the case of Portela and Thanossoulis (2001), which analysed the efficiency of English secondary schools by decomposing them into the efficiency depending on the centre and on the individual students themselves. In a similar way Mancebón et al. (2012) studied the results for Spain, in an attempt to differentiate between the effects of the type of school, the school, and the students in the efficiency; and Giménez et al. (2007) which introduced the concept of managerial efficiency.

In the aforementioned studies, diverse inputs are considered: measures of schools' resources like expenditure per student, eventually articulated in subcategories, student/teacher ratios, facilities, contextual variables to measure the student-family's background...

With respect to the outputs, although different measures can approximate the results of the educational process (success rates, grades assigned by teachers,...), there exists a consensus about the use of indicators derived by standardised test scores as they are homogeneous, comparable across countries and more difficult to manipulate. In this point, the Programme for International Assessment (PISA programme), launched in 2000 and carried out every three years, constitutes an important source of information to study the competencies acquired by the students and to make comparison across economies.

The PISA programme, initiated in 2000 and carried out every three years, has experimented a constant increase in the number of participating schools and countries. In the first edition of the programme, 265,000 students from 32 countries were evaluated. The last edition of this report in 2015 covered 540,000 students from 72 countries. The main target of the programme is to evaluate educational systems worldwide by testing the skills and knowledge of 15-year-old students in mathematics, science and reading (and, since 2012 in financial literacy as an option for each country).

In addition to academic achievement data, summarizing the results on the test about different topics, the PISA database contains a vast amount of information about students, their households and the schools they attend; as well as synthetic indexes, elaborated by OECD experts, by clustering responses to related questions provided by students and school authorities.

In this paper an alternative DEA-inspired model is proposed in order to assess the efficiency of the educational systems in the OECD (The Organisation for Economic Co-operation and Development) countries, using the information included in the PISA database. In particular, we are interested in the consideration of the number of students that achieve each proficiency level as the output of the system. To this end, an innovative model based in DEA methodology is developed.

The rest of the paper is organized as follows. Section 2 introduces DEA methodology and a new model for the evaluation of the efficiency, in a situation in which the output represents percentages of different categories is studied. In Section 3 the problem of measuring the efficiency of the educational systems across economies through PISA dataset is introduced and the dataset is described. Section 4 contains the discussion of results and Section 5 is devoted to the conclusions.

## 2. Methodology: Data Envelopment Analysis.

Data Envelopment Analysis (DEA) is a technique originally proposed in Charnes et al. (1978) as a methodology to evaluate the relative efficiency of a set of units, referred to as Decision Making Units (DMUs) in DEA terminology, involved in a production process or in public services. This methodology formalizes the original ideas proposed in Farrell (1957) of measuring efficiency of the production. In DEA models, the technical efficiency is defined as the relative ability of each DMU to produce outputs from several inputs. The basic efficiency of each unit is evaluated through the ratio of outputs over inputs. That is to say, the measurement of efficiency is defined as a ratio of weighted outputs over weighted inputs. Consider a set of  $n$  DMUs to be evaluated. Each DMU consumes  $m$  inputs to produce  $s$  outputs. By  $x_{ij}$  and  $y_{rj}$  are denoted, respectively, the amount consumed of input  $i$  ( $i = 1, \dots, m$ ) and the amount produced of output  $r$  ( $r = 1, \dots, s$ ) by the  $j$ th DMU (with  $j = 1, \dots, n$ ). The efficiency of unit  $j$  is defined as follows:

$$\text{Efficiency of unit } j = \frac{\sum_{r=1}^s v_r \cdot y_{rj}}{\sum_{i=1}^m u_i \cdot x_{ij}}; \quad (2)$$

where  $v_r$  and  $u_i$  denotes the weights assigned to output  $r$  and input  $i$  respectively.

DEA models determine those DMUs that constitute the efficiency frontier (efficient units) and the distance of the remaining DMUs (inefficient units) from the frontier. This distance, which represents a measure of the inefficiency of the units, will depend on the DEA model considered. The main characteristic of DEA methodology is that each unit can freely select the weighting vector, (i.e., each DMU can select their own vectors of weights  $u$  and  $v$  so that its own efficiency measurement is optimized), with a common set of constraints that limit this value for the complete set of units, usually equal to or lower than unity. Therefore, each DMU can select its own vector of weights to optimize its individual efficiency measurement. Hence, if a unit fails to achieve the maximum value of efficiency, this failure cannot be attributed to an arbitrary selection of the weighting factors.

Mathematically, the evaluation of unit  $o$  is determined as the solution of the following model:

$$\begin{aligned} \text{Max } \theta_o &= \frac{\sum_{r=1}^s v_r \cdot y_{ro}}{\sum_{i=1}^m u_i \cdot x_{io}} \\ \text{s. t. } \theta_j &= \frac{\sum_{r=1}^s v_r \cdot y_{rj}}{\sum_{i=1}^m u_i \cdot x_{ij}} \leq 1 \quad j = 1, \dots, n \\ u_i, v_r &\geq 0 \quad i = 1, \dots, m; r = 1, \dots, s. \end{aligned} \quad (3)$$

Note that model (3) determines the efficiency of unit  $o$ , with its own vector of weights (these ones that maximizes the efficiency ratio) subject to a common set of constraints such that the efficiency score is not greater than the unity for the  $n$  DMUs. Model (3) must be computed  $n$  times, one for each DMU. An efficient unit is characterized by an efficiency score ( $\theta_o$ ) equal to the unity. The remaining units, which achieve a value lower than the unity, are considered inefficient.

Model (3) can be transformed into a linear programming model with some algebraical transformations Charnes et al. (1978). The previous model is equivalent to the following expression:

$$\begin{aligned} \text{Max } \sum_{r=1}^s v_r \cdot y_{ro} \\ \text{s. t. } \sum_{i=1}^m u_i \cdot x_{io} &= 1 \\ \sum_{r=1}^s v_r \cdot y_{rj} - \sum_{i=1}^m u_i \cdot x_{ij} &\leq 0 \quad j = 1, \dots, n \\ u_i, v_r &\geq 0 \quad i = 1, \dots, m; r = 1, \dots, s. \end{aligned} \quad (4)$$

Model (4) is referred as CCR-model (in reference to the initial of its authors: Charnes, Cooper and Rhodes).

Two different specifications of DEA models can be considered: output-oriented, in which each units tries to maximizes its vector of output for a given amount of input; and input-oriented, in which the units tries to optimizes the amount of consumed inputs to produce a given amount of output. Note that the objective of the model implies respectively the determination of the maximum radial (proportional) reduction of inputs and the expansion of the outputs, such that the unit under evaluation is included in the production possibility set, constructed as a linear hull of the observed values of the  $n$  DMUs. Efficient units, since they are located at the efficiency frontier, do not admit any reduction of the vector of inputs, which is reflected by an efficiency score equal to the unity.

DEA-models can deal with both constant returns to scale (CRS) and variable returns to scale (VRS). Model CCR considers that all the units operate under constant return of scale. In Banker et al. (1984) the model with VRS assumption is proposed (commonly referred as BCC model). The model includes a convexity condition in the construction of the production possibility set. An interested reader can find a more extended explanation about the DEA methodology in Banker et al. (1984) or Cooper et al. (2000) among others.

Nevertheless, the application of DEA and the development of models has vastly exceeded its initial objectives, by generating a wide number of models and procedures, all of which are characterized by an endogenous determination of weights. That is, the weighting vectors are determined as a variable of the problem and are not externally fixed by the decision makers.

Those extensions includes the development of alternative models and the inclusion of variables which initially do not fit with the methodology. Among the models proposed as an alternative to the radial measures, one of the most applied is the additive model. This model was initially proposed in Charnes et al. (1985). In contrast to standard CCR and BCC models, which consider a radial measure to compute the distance to the efficient frontier, the additive model considers the maximization of the distance to the efficient frontier to evaluate the performance of each DMU. The basic expression of the additive model with VRS is:

$$\begin{aligned}
 &Max \quad \sum_{i=1}^m s_{i0}^- + \sum_{r=1}^s s_{r0}^+ \\
 &s. t. \quad \sum_{j=1}^n \lambda_j \cdot x_{ij} \leq x_{i0} - s_{i0}^- \quad i = 1, \dots, m \\
 &\quad \quad \sum_{j=1}^n \lambda_j \cdot y_{rj} \geq y_{r0} + s_{r0}^+ \quad r = 1, \dots, s \\
 &\quad \quad \sum_{j=1}^n \lambda_j = 1 \\
 &\quad \quad \lambda_j, s_r^+, s_i^- \geq 0
 \end{aligned} \tag{5}$$

This family of models deals directly with input excesses and output shortfalls (proposing a slack-based efficiency measure). Although this model can discriminate between efficient and inefficient DMUs by the existence of slacks, it has no means of gauging the depth of inefficiency, as can the efficiency measure in the CCR and BCC models. For a detailed discussion of the features of additive models, see, for instance, Tone (2001) and Lovell and Pastor (1995). This last paper is particularly interesting for the sake of this paper as the authors develop a model in which weights to differentiate between the factors (inputs and outputs) are included.

Note that model (5) is a non-oriented model; both inputs and outputs can be modified by inefficient units to reach the efficient frontier. The projections of observed values (denoted respectively by to the efficient frontier are determined as:

$$\begin{aligned}
 \hat{y}_{r0} &= y_{r0} + s_r^{*+} \\
 \hat{x}_{i0} &= x_{i0} - s_i^{*-};
 \end{aligned} \tag{6}$$

where  $s_r^{*+}$  and  $s_i^{*-}$  denote the optimal values of the slacks determined when model (5) is computed. These quantities represent the differences between the observed values and the corresponding reference point. The projected efficient point is reached by reducing inputs and/or increasing outputs so as to maximize the sum of the slacks in the objective function (this is why models the objective functions of this class of models are also referred to as slack-based measures). The original non-oriented model can alternatively be transformed to an input-oriented or output-oriented model whereby only the corresponding slack variable is considered in the objective function. Also the CRS model can be considered just eliminating the convexity constraint  $\sum_{j=1}^n \lambda_j = 1$ .

In Thrall (1996), a weighted additive model is proposed. The model includes a vector of weights for the slacks in the objective, respectively  $g^+ = (g_1^+, \dots, g_s^+)$  and  $g^- = (g_1^-, \dots, g_m^-)$  for outputs and inputs slacks, which may be determined either subjectively or objectively in a separate procedure. The model with the assumption of VRS is transformed into the following,

$$\begin{aligned}
 \text{Max} \quad & \sum_{i=1}^m g_i^- \cdot s_{io}^- + \sum_{r=1}^s g_r^+ \cdot s_{ro}^+ \\
 \text{s. t.} \quad & \sum_{j=1}^n \lambda_j \cdot x_{ij} \leq x_{io} - s_{io}^- \quad i = 1, \dots, m \\
 & \sum_{j=1}^n \lambda_j \cdot y_{rj} \geq y_{ro} + s_{ro}^+ \quad r = 1, \dots, s \\
 & \sum_{j=1}^n \lambda_j = 1 \\
 & \lambda_j, s_r^+, s_i^- \geq 0
 \end{aligned} \tag{7}$$

These weighting factors can be utilized in order to ensure that the units of measure associated with the slack variables do not affect the optimal solution. Note that the original additive model fails to satisfy the property of unit invariance. That is, the projections of the inefficient units on the efficient frontier depend on the scales used to measure each variable, which implies that the efficient measure does not have an intuitive interpretation Gouveia et al. (2008) since the objective is a sum of incommensurable slacks. It is necessary, therefore, to pre-standardize the original dataset when the variables are measured in diverse units. In contrast, the additive model is translation invariant, which renders it an optimal option to handle with negative values (since they can be transformed to positive values by adding an adequate positive quantity).

From an economic point of view, these weighting factors represent the marginal worth of the corresponding slack. Weights are associated with unit cost and unit prices of excess and shortfall slack variables. Hence the sum of weighted slack represents an approximation of the total cost of inefficiencies Bardhan et al. (1996).

For both radial and additive models, standard DEA-models assume certain basic features. Among others, must be cited the consideration of positive real values for variables (inputs and outputs); that all the outputs are desirable (in the sense that more is always preferred to less); the assumption that all the variables are controllable by DMUs (i.e. all variables, inputs and outputs can be modified by the units to achieve the efficient frontier); and that, once the efficient frontier is identified, inefficient DMUs reach this frontier by increasing the observed output values, decreasing the observed input values or by simultaneously modifying both variables. This depends on the orientation of the model: output-oriented, input-oriented, or non-oriented models, respectively.

However, many real-world situations can be found in which these assumptions are not verified. For those situations, a number of variations over original DEA models have been developed. Among others, for those cases in which real values do not fit the data available, several proposals can be found. See, among others, Cook et al. (1993) and Cook et al. (1996), where the inclusion of ordinal data and data on categorical variables is studied; Lozano and Villa (2006) where integer values are considered or Färe et al. (1989) and Scheel (2001) in which the inclusion of undesirable outputs is studied.

In this paper, a new model for the evaluation of the efficiency is proposed, which takes ideas from additive models and the consideration of non-standard variables. In particular, we consider situations in which only the redistribution of the observed output values will be permitted for the



efficiency to be attained, and not the incorporation of new units to increase the value of the output vector.

## 2.1. A DEA-inspired model to evaluate the efficiency in the presence of percentages.

In this section, a variation of additive model that permit to include percentages as values is developed. Let consider that the outputs represents percentages of categories of the same variable. This supposes that in every case, for both observed and projected values, the sum is equal to 100. We consider that the categories are ranked from the less to most valued ones.

Both features have important implications for the benchmarks and the way in which the inefficient units are projected to the efficient frontier. Necessarily, the improvement of the observed value of outputs must be carried out by a reallocation of the units from the less valued categories to the most ones. This is the unique alternative to improve the value of the outputs since increasing the value of the observed output (without reducing any other) is not a feasible option.

For this task, we propose a model inspired in the additive model described previously. Consider a set of  $n$  DMUs which are being evaluated with respect to the  $m$  inputs and one output separated in  $s$  categories. It is interesting to bear in mind that this supposes to consider in practice  $s$  outputs (each  $y_{rj}$  represents the values observed for DMU  $j$  in category  $r$ , with  $r = 1, \dots, s$ ).

Starting with the weighted additive model (7), consider a weighted output-oriented model. The evaluation of the DMU  $o$  is carried out by computing:

$$\begin{aligned}
 \text{Max} \quad & \sum_{r=1}^s g_r^+ \cdot s_r^+ \\
 \text{s. t.} \quad & \sum_{j=1}^n \lambda_j \cdot x_{ij} \leq x_{io} - s_i^- \quad i = 1, \dots, m \\
 & \sum_{j=1}^n \lambda_j \cdot y_{rj} \geq y_{ro} + s_r^+ \quad r = 1, \dots, s \\
 & \sum_{j=1}^n \lambda_j = 1 \\
 & \lambda_j, s_r^+, s_i^- \geq 0.
 \end{aligned} \tag{8}$$

In the context described above, the only way to improve the efficiency for an inefficient unit is to reallocate units across categories. That is to say, if one output is increased in one unit then it necessarily implies a reduction by the same amount in one or more than one of the remaining outputs. We propose the following variation regarding the output-oriented weighted additive model:

$$\begin{aligned}
 \text{Max} \quad & \sum_{r=1}^s g_r \cdot (s_r^{++} - s_r^{+-}) \\
 \text{s. t.} \quad & \sum_{j=1}^n \lambda_j \cdot x_{ij} \leq x_{io} - s_i^- \quad i = 1, \dots, m \\
 & \sum_{j=1}^n \lambda_j \cdot y_{rj} \geq y_{ro} + (s_r^{++} - s_r^{+-}) \quad r = 1, \dots, s \\
 & \sum_{j=1}^n \lambda_j = 1 \\
 & \sum_{r=1}^s s_r^{++} - \sum_{r=1}^s s_r^{+-} = 0 \\
 & \lambda_j, s_i^-, s_r^{++}, s_r^{+-}, g_r \geq 0;
 \end{aligned} \tag{9}$$

where  $g_r$  represents the weighting factor assigned to the the  $r$ th category. The vector  $g = (g_1, \dots, g_s)$  has to be constructed in order to assure that the relative importance of the categories are well represented. This can be a set of incomplete information, represented by a set of constraints with  $g_r$  variables (in this case, model (9) is not a linear model) or it contains a numerical value, objectively or subjectively determined. In that case, it is easy to see that model (9) is a linear programming model. In any case, considering that latter levels are better than prior ones, the relation between components of vector  $g$  must hold:  $g_r \leq g_{r+1}$ , for every  $r = 1, \dots, s - 1$ .

It is interesting to note that the slack variables of the output have been divided into two separated variables denoted by  $s_r^{++}$  and  $s_r^{+-}$ . The outputs represent percentages so both observed and projected

values must verify that the sum is equal to 100. This implies that any modification of the observed value must be carried out by a reallocation. That is to say, if one output increase (this raise is measured by variable  $s_r^{++}$ ) this necessarily implies that other(s) is(are) reduced (denoted by variables  $s_r^{+-}$ ) in order to assure that the sum of the  $s$  outputs is equal to 100.

The objective function of model (9) implies that the projected efficient point is reached by increasing certain levels (the most valued ones) and reducing others (the least valued), obtained from the maximization of augmentations and reductions through the objective function of (9). Note that the projections only affect the observed output values (output-oriented model), such that  $\hat{y}_{ro} = y_{ro} + s_{ro}^{++} - s_{ro}^{+-}$ .

The first and second set of constraints includes the classic DEA production structure and therefore all the units have to verify that  $\sum_{j=1}^n \lambda_j x_{ij} \leq x_{io}$  and that  $\sum_{j=1}^n \lambda_j y_{rj} \geq y_{ro}$ . Equivalent to said models, the condition of efficiency for the DMU  $o$  under model (7) is that the value of all slack variables is zero. That is to say, efficient units lead the constraints to the equality, and hence modifications are not possible. In any case, the observed values  $y_{ro}$  plus the optimum increase  $s_{ro}^{++}$  or minus the optimum decrease  $s_{ro}^{+-}$  will be compatible with the possibility production set. The output-orientation supposes that the modifications of inputs are not valuable. In this case, the target of the DMUs is the optimization of the observed outputs values (performance in mathematics test) for a given vector of inputs (resources assigned to the educational system).

The restriction  $\sum_{j=1}^n \lambda_j = 1$  is included in order to consider the VRS assumptions. By deleting that constraint, a model under constant return of scale would be constructed. Note that alternative assumptions over the returns of scale structure are also feasible.

The starting point of the DMU under evaluation  $o$  is its observed value of inputs and outputs. It is easy to see that a solution such that  $\lambda_o = 1$  and  $\lambda_j = 0$  for every  $j \neq o$ ; and  $s_r^{++} = s_r^{+-} = s_i^- = 0$  for every  $i, r$  always exists, therefore the model is feasible.

The model proposes a modification of the output vector only if it involves a positive value of the objective function of (9). This is equivalent to a new distribution of the values  $y_{or}$  which implies the movement of units from the less valued outputs to the most valued outputs. The improvement of the output is measured through the weighted sum of the differences ( $s_r^{++} - s_r^{+-}$ ). It is important to highlight how constraint  $\sum_{r=1}^s s_{1ro}^+ - \sum_{r=1}^s s_{1ro}^- = 0$  guarantees that only reallocations of units across the  $s$  outputs are permitted, and not an increment of total output, to improve the efficiency of DMU  $o$  is feasible.

By considering a vector of weights such that  $g_r \leq g_{r+1}$ , an inverse distribution (in which the worst categories are globally increased at the expense of the best categories) is not considered by the model. Bearing in mind that if one level is increased in a unit, this necessarily implies that any other (considering the simplest case) decreases by the same quantity this modification only holds in those cases in which the objective function is positive (which only occurs if the difference between better and worse levels is positive). Otherwise, the result of the objective is negative and does not improve the initial valuation of the unit.

Note that if no modifications on the outputs are carried out, then the value computed by the model is zero. Movements across outputs will only be carried out if the vector of outputs is improved (the aggregated measure) such that the values of better variables increase at the expense of a decrease in the values of the worse variables.

### 3. Evaluating the efficiency of educational systems of OECD countries.

In this Section, a proposal for the evaluation of the efficiency of the educational systems of OECD countries is studied. Similar to the main papers revised in Section 1, this study is based on the information of the PISA programme. Several studies about efficiency in education are based on the information contained in the PISA database. Some of these studies were referred in the literature review in Section 1. We consider a set of 34 OECD countries (all the OECD countries included in PISA except Greece, since the data of one of the input considered is not available).

With regard to the inputs variables, although each proposal consider a particular set of variables, most of them try to include the classical division of inputs: raw material, physical and human capital. In this paper, we consider, as does the PISA report itself, four types of resources needed for learning: financial resources, human resources, material resources, and resources of time.

As an indicator of the intensity of financial resources invested by each country in education, we use the cumulative expenditure by educational institutions per student from 6 to 15 years old measured in equivalent USD converted using purchasing-power parities (this input is labelled as  $I_1$ ). We consider it a very convenient proxy for the financial inputs as it takes into account the long-term nature of the learning process. Moreover, it uses a converted unit that enables various countries to be compared regardless of their cost of living.

With respect of human resources, teachers represent the most important part, and hence we use the student-teacher ratio. PISA provides the average number of students per teacher in every country. In order to use it as an input in the DEA model, the inverse of this ratio is calculated, that is, the number of teachers divided by the number of students ( $I_2$ ).

The third kind of input PISA identifies in the learning process is that of material resources. Schools need certain resources such as facilities, classrooms, heating,... Currently, countries are also making a special effort to provide students with technological material, such as access to the Internet and computers. Following Agasisti (2011), technological material is used here as a proxy for the material resources. Specifically, we use the number of computers available for educational purposes in the school divided by the number of students ( $I_3$ ).

The last type of resource that education requires is time. This variable measures the time per week spent in school in regular mathematics lessons, expressed in hours ( $I_4$ ). It is important to highlight that the selection of the time in mathematics is justified by the selection of the outputs. The evaluation is focused is the performance in this topic.

Table 1 summarizes the main characteristics of the variables described above.

**Table 1. Descriptive statistics of input variables.**

	$I_1$	$I_2$	$I_3$	$I_4$
<b>Max</b>	187,458.81	0.14	1.52	7.20
<b>Min</b>	27,848.44	0.42	0.16	2.40
<b>Average</b>	90,293.98	0.27	0.78	3.66
<b>Std. Dev.</b>	34,585.41	1.20	0.31	0.84

Source: Own elaboration.

Most authors, when choosing the output for DEA-models, use the mean scores of the topics evaluated in PISA. These scores are determined based on the so-called plausible values. These are found within the probability distribution estimated for a student's score in each test. Therefore, for every student's test, PISA provides five plausible values, where these are the probabilities for the student to obtain each of the values.

The PISA mean scores are based on the Rasch model, see Rasch (1960) and Wright and Masters (1982), which uses plausible values instead a particular mean value for each student's knowledge. These values are random values obtained from the distribution function of the results estimate from the results obtained in each test. They can be interpreted as a representation of the ability range of each student Wu and Adams (2007) The determination of plausible values can be seen in detail in OECD (2012).

The main reason stated in the report for using the plausible values is the necessity for the transformation of a continuous variable (e.g. student's ability) into a discrete variable (e.g. the scores). In this process, the plausible values have proved themselves as unbiased measures for the variable. They reduce the errors both from measuring and from the omission of underlying aspects that have not been considered specifically in the test.

However, the computation of these plausible values presents numerous disadvantages for researchers since it is necessary to calculate any given statistic, e. g. the mean, for every plausible value and then to compute the average for every individual student, which renders this method cumbersome. If the investigator were to omit this procedure, then the results could be biased.

In order to avoid all these problems, we propose an alternative vector of outputs: the percentage of students of each country in the different proficiency levels. As an alternative way of measuring the results in every subject, specifically mathematics, the PISA report classifies the students depending on their achievement in seven categories, called proficiency levels. The way these proficiency levels are constructed take into account not only the abilities of the students but also the difficulty of the items, thereby constituting a scale of literacy. In doing so, every proficiency level can be described as a group of abilities we can expect from the students contained within this level. According to the PISA report, its aim is to provide useful information for decision-making and predictions about education policies. This is why, in a complementary way, various related reports published by the OECD provide the percentage of students in each level. Working with these results enables any problems regarding plausible values to be avoided.

To consider the vector of proficiency levels as the output of the model, a modified efficiency evaluation model is required. It is important to bear in mind the characteristics of these values and to adapt the existing models to these particularities.

It is interesting to point out how these values can be easier for policy makers to evaluate and interpret. With these variables, the benchmark of the efficiency model are represented by the percentages of students that must be in each level for an observed vector of resources. The improvement is measured through the number of students that must achieve a particular level of proficiency in the test.

Therefore, the results in each topic in the PISA are standardized with a mean of 500 and a standard deviation of 100. Seven proficiency levels are constructed in which the students are allocated depending on their results in the topic (see OECD (2013)). In this paper, the results obtained in Mathematics have been considered.

The first level comprises the students with scores below 357.77 points. The following levels includes students with scores included in the following intervals: second level from 357.77 to 420.07 points, third level from 420.07 to 482.38, fourth level 482.38 to 544.68 points, fifth from 606.99 to 669.3 points. The last level, the most valued one, includes the students with more than 669.3 points. The data considered here includes the percentages of students that achieve each proficiency level, as a means

of reflecting the performance of the educational system. The outputs vector contain the seven level described above, labelled from  $O_1$  (percentage of students with scores below 355.77 points) to  $O_7$  (percentage of students with scores over 669.3 points). The data of the statistics of the seven outputs considered (proficiency levels) are summarized in Table 2.

**Table 2. Descriptive statistics of output variables.**

	$O_1$	$O_2$	$O_3$	$O_4$	$O_5$	$O_6$	$O_7$
<b>Max</b>	25.53	31.11	28.27	29.98	25.92	15.05	6.58
<b>Min</b>	2.22	7.75	17.23	12.89	3.20	0.31	0.02
<b>Average</b>	8.47	14.89	22.55	24.81	18.60	8.37	2.31
<b>Stand. Dev.</b>	5.68	5.11	2.55	3.59	5.11	3.46	1.44

Source: Own elaboration.

The special characteristics of these proficiency levels require an adaptation of the model for the evaluation of the efficiency. The model developed in Section 2.1, model (9), is fitted to these particular features. The consideration of an additive model (versus a radial model) is based on the characteristics of the feasible variations for the outputs. In order to reach the frontier by modifying the percentages in each level, these outputs could increase in different quantities and not radially. Note that levels denote different importance; obviously higher scores imply a larger importance. An increase of the number of students in the latter categories requires, from the educative system, a greater effort than an increase of the number of students in the previous ones. From this point of view, the efficient countries would be those that have larger percentages of students in the better proficiency levels and smaller percentages in the worse categories. It is important to bear in mind that the outputs represent percentages; consequently the sum for each DMU has to be equal to 100 not only for the observed values but also for the projection in the efficiency frontier. That is, if the country achieves more students in better categories this is because it has fewer students in the worse categories.

To compute the model, a set of 34 units (OECD countries, all the OECD countries included in PISA except Greece), are being evaluated with respect to the four inputs previously and the vector of output which represents the seven proficiency levels. To mitigate the effect of outliers and/or the existence of errors, the models has been robustified using the concepts proposed in Cazals et al. (2002). To this end, 2,000 computation rounds of each model are obtained with a sub-sample of 28 randomly selected units.

### 3.1. Discussion of results.

Table 3 summarizes the results obtained for a weighting vector such that  $g = (1,2,4,8,16,32,64)$ . Note that the particular value assigned to  $g$  could be done in several ways, in a subjective way (like the one we propose) or by means of an additional procedure that measure the relative importance of each level. Note that each component  $g_r$  of vector  $g$  tries to approximate the marginal worth of the corresponding slack (in relative terms). Thus, the objective function of model (9) approximate the total cost of the inefficiency on the unit. The selection of  $g$  can proceed from a political decision in order to emphasise the relative importance or effort in the reallocation of one unit from one category to the other. Or alternatively, the determination of  $g$  may result from a technical analysis. In any case, the consideration of alternative values for vector  $g$  affects to inefficient units since the sum of slacks are weighted in a different way. Therefore, the construction of a ranking of unit based on the optimal value of the model would be affected. But note that those unit characterized as efficient are not affected by any modification in vector  $g$ .

**Table 3. Main results: slack variables results.**

DMU	$O_1$	$O_2$	$O_3$	$O_4$	$O_5$	$O_6$	$O_7$
Australia	-3.68	-4.35	-2.69	1.31	4.57	3.48	1.36
<b>Austria</b>	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Belgium	-2.85	-2.87	0.16	2.25	1.65	0.53	1.13
Canada	1.57	0.07	-2.18	-2.86	-0.32	2.02	1.71
<b>Chile</b>	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<b>Czech Republic</b>	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<b>Denmark</b>	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<b>Estonia</b>	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<b>Finland</b>	0.00	0.00	0.00	0.00	0.00	0.00	0.00
France	-5.18	-4.97	-0.72	3.06	2.97	2.69	2.14
Germany	-0.17	-0.37	-0.31	0.02	0.25	0.32	0.26
<b>Hungary</b>	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Iceland	-4.69	-6.64	-6.40	0.38	7.35	6.59	3.42
<b>Ireland</b>	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<b>Israel</b>	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Italy	-2.80	-2.89	-2.27	1.26	2.85	2.55	1.30
<b>Japan</b>	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<b>Korea</b>	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Latvia	-0.22	-4.02	-6.23	-1.69	5.01	5.23	1.92
Luxembourg	-3.67	-6.88	-4.69	0.63	5.04	6.06	3.52
<b>Mexico</b>	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<b>Netherlands</b>	0.00	0.00	0.00	0.00	0.00	0.00	0.00
New Zealand	-1.22	-2.66	-1.46	0.83	2.07	1.87	0.57
Norway	0.01	-1.72	-4.94	-3.02	2.22	4.53	2.92
<b>Poland</b>	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Portugal	-3.09	-4.69	-3.84	0.07	3.53	4.78	3.24
<b>Slovak Republic</b>	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<b>Slovenia</b>	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Spain	-3.62	-5.07	-4.04	0.09	5.17	5.10	2.38
Sweden	-2.38	-3.20	-3.90	-0.51	3.61	3.95	2.27
<b>Switzerland</b>	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<b>Turkey</b>	0.00	0.00	0.00	0.00	0.00	0.00	0.00
United Kingdom	-2.52	-4.01	-4.98	-1.83	4.17	5.60	3.56
United States	-5.07	-6.70	-5.30	1.87	6.94	5.98	2.28

Source: Own elaboration.

In Table 3, the most relevant results of the application of our model are shown. For every country, the net (positive or negative) variation for each proficiency level is provided. It represents the amount by which that specific country must increase or decrease the percentage of students in that category to become efficient, calculated as the difference between the  $s_r^{++}$  and the  $s_r^{+-}$  variables. Those countries that lie on the efficiency frontier show a 0 in all the slacks.

From this analysis, the countries can be classified into two different groups, efficient (denoted in bold) and inefficient. In the first group, we find Austria, Chile, Czech Republic, Denmark, Estonia, Finland, Hungary, Ireland, Israel, Japan, Korea, Mexico, Netherlands, Poland, Slovak Republic, Slovenia, Switzerland and Turkey. The way in which these countries achieve efficiency differs greatly. Certain countries, such as Chile, Turkey and Mexico, despite their low results, have an efficient educational system, because their investment in education is comparatively smaller.

The results in Table 3 must be interpreted as follows. Each value represents the percentage of the net variation of the corresponding level. Let consider the case of Spain. With the resources considered the system is characterized as inefficient. The improvement proposed implies the raise of the outputs from  $O_4$  (propose an increase of the percentages of students with scores between 482.28 and 544.68 of

0.48 points) to  $O_7$  (the feasible increase in the percentage of students with a score over 669.30 is 2.44 percentage points) at expenses a reduction of the remaining ones. The excess in the first level, students under 357.77 points, is 4.25 percentage points. For the following two levels, the reduction is 5.55 and 4.15 respectively. It is clear that this reallocation of students from the worst valued level to the best ones would suppose a improvement in the aggregated value of the output and in the results of the systems (better students' results with a given amount of resources). But also the model guarantees that the proposed reference value is feasible in the sense that it is included in the possibility production set constructed with the observed units. This feature explains movements like the one proposed for Norway, in which a raise in the first level (the one with student with lowest scores) is proposed. This is explained by the requirement of the classic constraints of DEA for being enveloped by the efficiency frontier. Even so, the aggregated value of the projected output vector would increase.

A subset of countries found in the first group involve those systems that obtain good results in PISA but need to invest resources above the mean. Austria and Finland are found in this subset. Finally, there are countries that achieve great levels of proficiency but employ fewer resources than the rest of the members of the OECD. This is the case of Estonia, Korea, Netherlands and Poland. The inefficient countries are those which, given their available resources, should obtain better results in PISA. Among these, we can mention Portugal, Sweden and Italy as the countries that are farthest from the efficiency frontier.

Another important result provided by the model is the units of reference for each country. In order to become efficient, the inefficient units have to increase their outputs until they reach the efficient frontier. The inefficient countries should modify their outputs until they reach the levels of those efficient countries that have a similar structure of inputs. In Table 4 (see Annex), the inefficient countries can be seen in the first column and the countries which they should imitate appear in the following columns with their corresponding lambda value. In this case, the reference sets have been obtained by computing model (9) for the complete set of units.

From this point of view, and given that these countries are efficient, we can consider the best countries in terms of educational efficiency to be those which constitute a reference for other countries. Since these units not only are located in the efficiency frontier but also there are certain units with a similar combination of inputs and outputs that are revealed as less valued. In this respect, Korean, as the reference for 14 countries, Switzerland for 9 and Estonia, Ireland and Netherlands for 5 constitute the reference set for the inefficient countries.

On the other hand, we found efficient countries which do not constitute the reference set for any inefficient unit. This is the case of Austria, Czech Republic, Israel, Mexico and Turkey. In brief, this set of countries constitutes *extreme* cases, since their combination of inputs and outputs are characterized as efficient, they are quite different of the remaining countries under evaluation. That is, the observed values of inputs and output of these units are quite different of the other ones and this could be the cause to be part of the efficiency frontier (and not a good performance).

#### **4. Concluding remarks.**

In this paper, an study on the efficiency of the educational systems of the OECD countries has been developed. The study is based on the application of the Data Envelopment Analysis (DEA) methodology and the dataset provided by the PISA report. The analysis has been done considering the resources of each system and results of each economy in the mathematical test in PISA 2015.

The PISA report assesses the learning achievement of the students and classifies them into seven level of proficiency, depending on their abilities. Therefore, for each country, the percentage of students in every proficiency level is available. We propose to use these values instead of the mean score on a particular topic to evaluate each country. Using this variable as an output permits to avoid the consideration of plausible values and a straight interpretation of the benchmarks but requires a specific model, such as that developed in this paper. It is easy to see that the total amount of the different outputs cannot increase, but can only be reallocated, since we are dealing with percentages.

A variation of the weighted additive DEA model to reallocate outputs has been proposed. Contrary to radial models, the strategy to achieve efficiency of additive models allows each variable (inputs and outputs) to be modified by a particular quantity. By including a vector of weights, the relative importance of each variable can be suitably represented. This feature allows us to take into account the differences in cost or the effort the units must exert to increase or reduce, respectively, the diverse outputs and inputs.

We develop a model for this particular context. We consider a situation in which the aggregate value of the output is a fixed value and the strategy to improve the performance is a reallocation of units across the outputs. Increasing a particular output necessarily implies a reduction of any other output (in order to maintain the aggregate value constant and equal to 100) and the DMUs are interested in moving units from the least to the most valued level.

As a result, the countries have been classified into two different groups: the efficient and inefficient units. The first group is identified by null values in all the slacks. Additionally, the model provides an efficient reference country with similar input structure for every inefficient country in order to improve their results. The countries that serve as a reference for the greatest number of educational systems are Korea and Japan. For the inefficient units, the values of slacks and projected outputs can serve as an accurate guide to political actors. These values represent a target for the number of students that achieve each proficiency level, and have been obtained considering other countries with a similar structure of inputs and outputs.

Future lines of research could carry out an in-depth study into the causes that provoke the differences between the educational systems, and could analyse how to make reforms that would solve the problems of the inefficient countries, since DEA models can only offer general guidelines. Improvements in the theoretical model are also possible, among others, the inclusion of additional information (complete or incomplete) or the consideration of additional procedures in order to determine the vector of weights associated to each level.

## Acknowledgments

This work has been partially funded by the Ministry of Science, Research and Universities, project PGC2018-095786-B-I00.

## References

- Agasisti, T. (2011). Performances and spending efficiency in higher education: A European comparison through non-parametric approaches. *Education Economics*, 19(2), 199-224.
- Agasisti, T. (2014). The efficiency of public spending on education: An empirical comparison of EU countries. *European Journal of Education*, 49(4), 543-557.
- Afonso, A., & St. Aubyn, M. (2006a) Non-parametric approaches to education and health efficiency in OECD countries. *Journal of Applied Economics*, 8(2), 227-246.
- Afonso, A., & St. Aubyn, M. (2006b). Cross-country efficiency in secondary education provision: a semi-parametric analysis with non-discretionary inputs. *Economic Modelling*, 23(3), 476-491.
- Aristovnik, A. (2011). An analysis of the efficiency of education spending in central and eastern Europe. *EconPapers*. Retrieved from <http://EconPapers.repec.org/RePEc:isv:mklp11:277-286>.



- Banker, R.D., Charnes, A., & Cooper, W.W. (1984). Some models for estimating technical and scale inefficiency in data envelopment analysis. *Management Science*, 30, 1078-1092.
- Bardhan, I., Bowlin, W.F., Cooper, W.W., & Sueyoshi, T. (1996). Models for efficiency dominance in data envelopment analysis. Part I: Additive models and MED measures. *Journal of the Operational Research Society of Japan*, 39, 322-332.
- Bessent, A.M., & Bessent, E.W. (1980). Determining the comparative efficiency of schools through data envelopment analysis. *Educational Administration Quarterly*, 16(2), 57-75.
- Bessent, A.M., Bessent, E.W., Kennington, J., & Reagan, B. (1982). An application of mathematical programming to assess productivity in the Houston independent school district. *Management Science*, 28, 1355-1367.
- Cazals, C., Florens, J.P., & Simar, L. (2002). Nonparametric frontier estimation: A robust approach. *Journal of Econometrics*, 106(1), 1-25.
- Charnes, A., Cooper, W.W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operations Research*, 2, 429-444.
- Charnes, A., Cooper, W.W., & Rhodes, E. (1981). Evaluating program and managerial efficiency: an application of Data Envelopment Analysis to Program Follow Through. *Management Science*, 27, 668-697.
- Charnes, A., Cooper, W.W., Golany, B., & Seiford, L. (1985). Foundations of Data Envelopment Analysis for Pareto-Koopmans efficient empirical production functions. *Journal of Econometrics*, 30(1), 91-107.
- Clemens, B. (2002). How efficient is education spending in Europe?. *European Review of Economics and Finance*, 1, 3-26.
- Cook, W.D., Kress, M., & Seiford, L.M. (1993). On the use of ordinal data in Data Envelopment Analysis. *Journal of the Operational Research Society*, 44 (2), 133-140.
- Cook, W.D., Kress, M., & Seiford, L.M. (1996). Data Envelopment Analysis in the presence of both quantitative and qualitative factors. *Journal of the Operational Research Society*, 47, 945-953.
- Cooper, W.W., Seiford, L.M., & Tone, K. (2000). *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*. New York: Springer.
- Cordero-Ferrera J.M., Crespo-Cebada, E., Pedraja-Chaparro, F., & Santin-Gonzalez, D. (2011). Exploring educational efficiency divergences across Spanish regions in PISA 2006. *Revista de Economia Aplicada*, 19(57), 117-145.
- Färe, R., Grosskopf, S., Lovell, C.A.K., & Pasurka, C. (1989). Multilateral productivity comparison when outputs are undesirable: A non-parametric approach. *The Review of Economic and Statistics*, 71, 90-98.
- Farrell, M.J. (1957). The measurement of the efficiency. *Journal of the Royal Statistical Society Series A*, 120(3), 253-290.
- Gimenez, V., Prior, D., & Theme, C. (2007). Technical efficiency, managerial efficiency and objective setting in the educational system: An international comparison. *Journal of the Operational Research Society*, 58, 996-1007.

- Gouveia, M.C., Dias, L.C., & Antunes, C.H. (2008). Additive DEA based on MCDA with imprecise information. *Journal of the Operational Research Society*, 59(1), 54-63.
- Hanushek, E. (1979). Conceptual and empirical issues in the estimation of educational production functions. *Journal of Human Resources*, 14(3), 351-288.
- Johnes, J. (2006). Data Envelopment Analysis and its application to to the measurement of efficiency in higher education. *Economics of Educations Review*, 25(3), 273-288.
- Levin, H. (1974). Measuring the efficiency in educational production. *Public Finance Quarterly*, 2, 3-24.
- Lovell, C.A.K., & Pastor, J.T. (1995). Units invariant and translation invariant DEA models. *Operations Research Letters*, 18, 147-151.
- Lozano, S., & Villa, G. (2006). Data Envelopment Analysis of integer-values of inputs and outputs. *Computers and Operational Research*, 33, 3004-3014.
- Mancebón M.J., & Bandrés, E. (1999). Efficiency evaluation in secondary schools: The key role of model specification and of ex post analysis of results. *Education Economics*, 7(2), 131-152.
- Mancebón, M.J., Calero, J., Choi, A. & Ximénez-De-Embún, D.P. (2012). The efficiency of public and publicly subsidized high schools in Spain: Evidence from PISA-2006. *Journal of the Operational Research Society*, 63(11), 1516-1533.
- OECD (2012). *PISA 2009 Technical Report*. PISA, OECD Publishing. DOI: 10.1787/9789264167872-en.
- OECD (2013). *PISA 2012 Results: What Makes Schools Successful? Resources, Policies and Practices (Volume IV)*. PISA, OECD Publishing. DOI: 10.1787/9789264201156-en.
- Portela, M.C.A.S., & Thanassoulis, E. (2001). Decomposing school and school type efficiency. *European Journal of Operational Research*, 132(2), 114-130.
- Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen, Denmark: Danmarks Paedagogiske Institut.
- Santin, D., & Sicilia, G. (2015). Measuring the efficiency of public schools in Uruguay: main drivers and policy implications. *Latin America Economic Review*, 24(5), 1-28.
- Scheel, H. (2001). Undesirable outputs in efficiency evaluation. *European Journal of Operational Research*, 132, 400-410.
- Sutherland, D., Price, R., & Gonand, F. (2009). Improving public spending efficiency in primary and secondary education. *OECD Journal of Economic Studies*, 1, 1-30.
- Thrall RM. (1996). Duality, classification and slacks in DEA. *Annals of Operations Research*, 66, 109-138.
- Tone, K. (2001). A slacks-based measure of efficiency in data envelopment analysis. *European Journal of Operational Research*, 130 (3), 498-509.
- Wright, B.D., & Masters, G.N. (1982). *Rating scale analysis*. Chicago: MESA Press.

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

Wu, M., & Adams, R. (2007). *Applying the Rasch model to psycho-social measurement: A practical approach*. Melbourne: Educational Measurement Solutions.

## Annex 1

**Table 4. Main results. Benchmarks.**

DMU	References												
	Chile	Denmark	Estonia	Finland	Hungary	Ireland	Japan	Korea	Netherlands	Poland	Slovak	Slovenia	Switzerland
Australia		0.244	0.273					0.363					0.119
Belgium			0.065	0.167				0.467					0.301
Canada								0.824	0.176				
France			0.591					0.205					0.205
Germany						0.447		0.349	0.107			0.086	0.010
Iceland							0.800	0.200					
Italy						0.212		0.310		0.380			0.098
Latvia			0.762		0.238								
Luxembourg								0.889	0.111				
New Zealand	0.026					0.387		0.487			0.101		
Norway								0.278	0.222				0.500
Portugal							0.258	0.742					
Spain		0.146	0.652			0.007		0.174					0.022
Sweden				0.065				0.337				0.326	0.272
United Kingdom							0.095	0.905					
United States						0.247		0.042	0.458				0.253

Source: Own elaboration.