



JRC TECHNICAL REPORTS

# **Efficiency of investment in compulsory education: An Overview of Methodological Approaches**

Tommaso Agasisti and Giuseppe Munda

2017

# **Efficiency of investment in compulsory education: An Overview of Methodological Approaches**

This publication is a Technical report by the Joint Research Centre, the European Commission's in-house science service. It aims to provide evidence-based scientific support to the European policy-making process. The scientific output expressed does not imply a policy position of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use which might be made of this publication.

#### **Contact information**

Tommaso Agasisti, Politecnico di Milano School of Management  
Department of Management, Economics and Industrial Engineering  
[tommaso.agasisti@polimi.it](mailto:tommaso.agasisti@polimi.it)

Giuseppe Munda, European Commission, Joint Research Centre  
Directorate B – Innovation and Growth, Unit JRC.B.4 – Human Capital and Employment  
TP 361 – Via E .Fermi 2749 – I-21027 – Ispra (Va) – ITALY  
[giuseppe.munda@ec.europa.eu](mailto:giuseppe.munda@ec.europa.eu)

#### **JRC Science Hub**

<https://ec.europa.eu/jrc>

JRC106681

EUR 28608 EN

PDF ISBN 978-92-79-68864-5 ISSN 1831-9424 doi:10.2760/140045

---

Luxembourg: Publications Office of the European Union, 2017

© European Union, 2017

The reuse of the document is authorised, provided the source is acknowledged and the original meaning or message of the texts are not distorted. The European Commission shall not be held liable for any consequences stemming from the reuse.

All images © European Union 2017

How to cite: Tommaso Agasisti and Giuseppe Munda; Efficiency of investment in compulsory education: An Overview of Methodological Approaches; EUR 28608 EN. Luxembourg (Luxembourg): Publications Office of the European Union; 2017. JRC106681; doi:10.2760/140045

## Table of contents

Acknowledgements.....	4
Abstract .....	5
1. Introduction .....	6
2. The concept(s) of efficiency in education .....	9
2.1. Three baseline concepts: technical, allocative and overall efficiency.....	9
2.2 Two additional definitions: spending and scale efficiency .....	11
3. Measuring educational efficiency in practice: the selection of inputs and outputs ..	13
4. Methodological approaches for assessing efficiency in education: non-parametric methods vs stochastic frontier models.....	20
4.1 Non-parametric methods: Data Envelopment Analysis.....	20
4.2. Parametric methods: Stochastic Frontier Analysis (SFA).....	26
4.3. Some recent advancements in methodology.....	28
5. Methodological approaches for assessing efficiency in education: Multi-Criteria Evaluation .....	31
5.1 What is multi-criteria evaluation?.....	31
5.2 Why discrete approaches can be useful for efficiency analyses?.....	32
6. Conclusions.....	35
ANNEX. ....	37
References .....	52

## **Acknowledgements**

Comments by colleagues at JRC.B4 and DG EAC on previous versions of this document have been very useful for its improvement.

## **Note**

This technical report is part of CRELL VIII Administrative Arrangement agreed between DG EDUCATION and CULTURE (EAC) and DG JOINT RESEARCH CENTRE (JRC). In particular it refers to point 2.1 of the Technical Annex accompanying CRELL VIII.

## Abstract

The policy discourses often refer to the term “efficiency” for indicating the necessity of reducing resources devoted to interventions and whole sub-sectors, while keeping the output produced constant. In this technical report, we review the theoretical and empirical foundations of efficiency analysis as applicable to the educational policy. After introducing some key concepts and definitions (technical, allocative, spending and scale efficiency), the report illustrates which variables of inputs, outputs and contextual factors are used in applied studies that assess efficiency in compulsory education. Then, an explanation of methods for conducting efficiency studies is proposed; in particular frontier methods such as non-parametric approaches (as Data Envelopment Analysis) and parametric models (as Stochastic Frontier Analysis) and multi-criteria approaches (such as Multi-Objective Optimisation and Discrete Methods) are reviewed. The main objective of this report is to present to the interested reader the main technical tools which can be applied for carrying out real-world efficiency analyses. A tween report presents an application of efficiency analysis for European compulsory education, at country level.

## 1. Introduction

The educational policies, in the last decades, have been characterized by a growing attention to the role that skills and educational results exert on the economic and social development of countries and communities. Since the literature outlined the potential role of human capital (HC) in the process of economic growth, policy makers have been more and more interested in understanding those factors that are correlated with the creation and development of people's HC (Benhabib & Spiegel, 1994; Romer, 1990; Barro, 2001; Hanushek & Woessmann, 2008; 2010 and 2012). In this context, the main practical aim of educational policy makers is to create the opportunities for maximizing student results (as for instance, achievement or test scores). The result of improving students' results can be obtained (beyond the approaches based on teaching quality, such as improvement of teachers' quality, innovation in teaching, and the use of digital technologies) by means of specific interventions on various aspects of the educational process, such as:

- Intervening in the *system-level arrangements* about the level of autonomy granted to educational institutions, implementing policies for accountability, selecting the optimal degree of competition between schools, etc. – see Woessmann, 2007;
- Qualifying the *management and governance of schools* (Bloom et al., 2015; Di Liberto et al., 2015), making principals and managers more skilled on the technical and leadership grounds;
- Providing *incentives* to schools and to staff through performance-based funding systems and reforms (Ladd, 1996; Jongbloed & Vossensteyn, 2001);
- Increasing the resources available, although the academic literature debates on the actual link between resources invested and results obtained (Hanushek, 1986; 1997; 2006; Krueger, 2003) reaching inconclusive findings, and a clear link between quantities of resources and educational results is still to be demonstrated.

Whatever the tools that are used for improving students and institutions' results<sup>1</sup>, the debate on the determinants of educational performance is vivid and relevant both between and within countries. On one side, the availability of international standardized tests allows **benchmarking educational systems across countries**, with the aim of understanding the determinants of student achievement, as measured by test scores – see, for instance, the international analyses such as Programme for International Student Assessment (PISA), Trends in International Mathematics and Science Study (TIMSS) and Progress in

---

<sup>1</sup> With this general expression, we mean the broader array of performances that can be considered as objective function for schools and universities, among which: achievement scores, retention, non-cognitive skills, research quality, knowledge dissemination, etc. In this context, we want to be open in discussing the various areas of educational performance that can be inserted as outputs in the context of efficiency analyses, without being forced to limit the analysis to easily-measurable variables.

International Reading Literacy Study (PIRLS). Several authors have used information from these ‘internationally-standardized’ test scores to derive lessons about national-level outcomes – see, for example, the conclusions drawn by Hanushek & Woessmann (2010) or the indications from the OECD’s reports (OECD, 2014). Alternatively, one could analyse what determines the fact that, within the same scholastic system, some schools obtain better educational results than others, and how test scores depends on various students’ personal characteristics and background (as noted since Coleman et al., 1966), besides schooling. There exist many papers that conduct empirical estimates about the determinants of such within-country differences in educational results between-schools and across individuals (Greenwald et al., 1996), and many of them obtain similar findings, such as the role of individual and schools’ socioeconomic status (SES) (Perry & McConney, 2010; Haveman & Wolfe, 2005), teachers’ quality (Darling-Hammond, 2000), peer effects (Sacerdote, 2011), etc.

A parallel stream of the literature is the one that discusses the **efficiency of educational systems and organizations**, not their absolute performance (i.e. test scores); in other words, the analysis is focused not on the overall results obtained by students, schools (on average) or education system (as a whole), but on the ability of reaching such results by using the least amount of possible resources – or, conversely, of maximizing the educational results with the available resources (Johnes, 2004). In this type of analysis, then, the inputs enter into the picture – i.e. the empirical study specifically intends to consider how many resources are employed for obtaining those results, and not only the level of educational outcomes. This way, the empirical analysis must also deal with the collection of data about the inputs, and it should model the process of transformation of the inputs (resources) into outputs (educational results). Two levels of analysis can be considered (more precise definitions are provided in the Section §2 of this Report):

- one that poses its attention on the *‘spending efficiency’ at country level* (how the financial resources allocated to education are used, and which average educational performance are able to ‘produce’?), and
- one that looks at *‘technical efficiency’ of each single school/university*, considered as an organization that uses financial and human resources, besides managerial techniques and technology, to produce (average) educational achievement of its students<sup>2</sup>.

Why is the analysis of efficiency in education important for policy making, beyond measuring and investigating educational performance? In our opinion, there are three aspects that deserve specific attention:

---

<sup>2</sup> It is also possible to measure “technical” efficiency at country level, although this technical measure then loses its ability to describe the “educational process” which is better conceptualized at institutional level, see Gimenez et al. (2007).

1. Efficiency encompasses the concept of educational performance, but puts its interpretation within the area of *feasibility*. Specifically, the framework behind efficiency analysis considers the amount of resources as limited, and so focuses on the maximum gains of performance that can be achieved, given the resources available. This is strikingly different from traditional analyses of educational performance assuming that students and schools can obtain the level of performance observed in other contexts/situations, which are instead very different in terms of resources employed.
2. Efficiency measurement is intrinsically *context-specific*. In particular, the inputs and outputs that are used are somehow dependent upon the characteristics of students who are attending the institutions (i.e. different socio-economic background), the values and human capital stocks of families and communities living in the areas where the school operates, etc. In this perspective, efficiency measurements must try to disentangle effects on performances that are due to managerial activities from those that are attributable to contextual factors. Failing such an objective will result in biased, unfair and misleading, measures of efficiency.
3. Efficiency analyses can inform policy-makers also about the *combination of inputs* that can result in output-maximization.

The importance of improving efficiency is not confined to single countries or specific grades, but is instead central in the modern studies about educational challenges as an imperative for the future (Hanushek & Luque, 2003; Sutherland et al., 2010) and the discussion about means for improving efficiency is faced by several governments – also in an international perspective. These aspects explain why the European Commission has underlined the importance of efficiency considerations for shaping educational policy.

In the first part of this technical report, we describe the academic literature that defines and measures the efficiency in the field of education. In so doing, we pay particular attention to the selection of the relevant variables (i.e. inputs, outputs, and contextual factors that affect efficiency) and to the empirical approaches that can be used for efficiency measurements, such as frontier methods and multi-criteria evaluation. Frontier methods are the traditional efficiency assessment approach in education economics and management. Multi-criteria evaluation has been widely used in various fields since the sixties both at micro at macro levels of analysis (see e.g. Figueira *et al.*, 2016); common applications in public policy refer to energy, finance, sustainable development, land use, regional planning, ... . In the framework of education policy, the desirability of the peculiar characteristics of multi-criteria evaluation has been advocated by various authors (e.g. Dill, & Soo, 2005; Guskey, 2007; Ho *et al.*, 2006; Malen and Knapp, 1997; Nikel & Lowe, 2010; Rossell, 1993; Stufflebeam, 2001; Tzeng *et al.*, 2007).

In the context of frontier methods, we describe non-parametric methods such as Data Envelopment Analysis or DEA and parametric methods like Stochastic Frontier Analysis or SFA. Multi-criteria evaluation approaches are reviewed by considering both continuous (i.e. extensions of traditional linear programming methods) and discrete (i.e. the case where the number of options is finite in number) approaches. While continuous approaches are still related to frontier methods (in particular they can be considered an attempt of improving DEA techniques), discrete multi-criteria methods are based on complete different assumptions (and can hence be considered a complementary approach).

We use such in-depth review to propose lines of research that can increase the awareness about educational spending efficiency in Europe, and potential indications to policy makers (and institutions' managers).

## 2. The concept(s) of efficiency in education

### 2.1. Three baseline concepts: technical, allocative and overall efficiency

In this report, the concept of efficiency that is adopted is derived by the pioneering work by Farrell (1957), in which the author develops a general framework for defining, analysing and measuring efficiency. The three main operative definitions that we use for efficiency are:

- *Technical efficiency*, that is defined as the lowest amount of input(s) that can be used for the production of a given level of output(s) – or, conversely, the highest amount of output(s) that can be produced, given the available level of input(s);
- *Allocative efficiency*, which is the best combination of input(s) that can be used, given their relative price, for producing a given level of outputs;
- *Overall (total economic) efficiency*, which measures the best combination of inputs that can be used for producing a technically efficient amount of outputs.

A graphical illustration can help in describing the different senses of these definitions (see Figure 1).

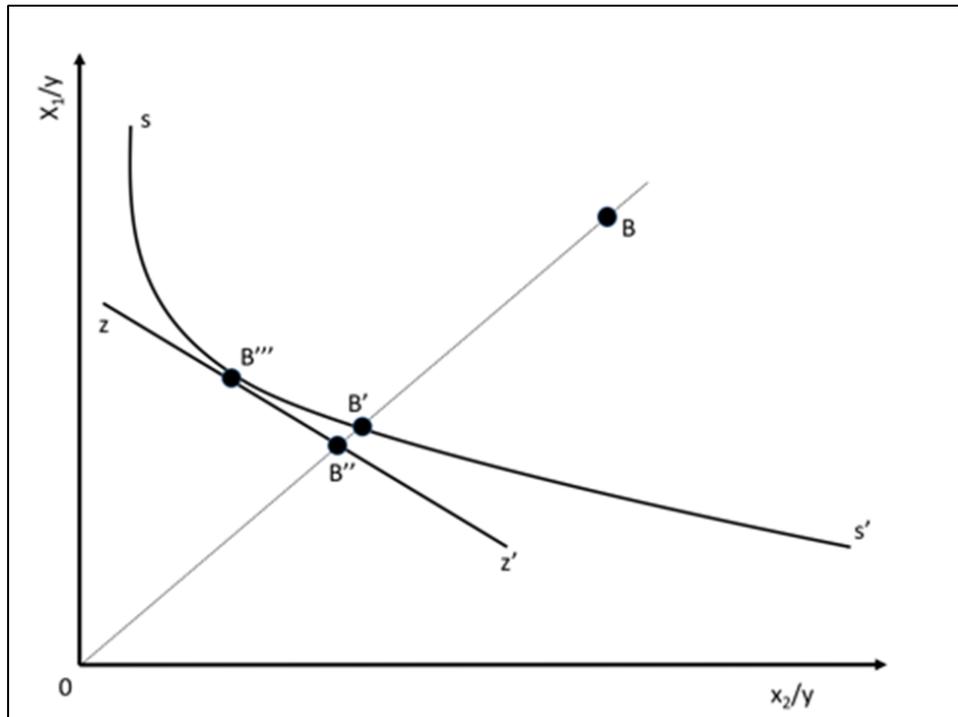
Let us consider a sector where educational institutions produce one type of output (for instance: the number of formative credits offered to their students) with two inputs: academic staff,  $x_1$  and administrative (support) staff,  $x_2$  – it is important to note, here, that inputs are expressed in physical units, not spending levels. The line  $ss'$  identifies the isoquant of efficient production, that is the set of efficient combinations of the two inputs  $x_1$  and  $x_2$  that can be used to produce a given (maximum) level of outputs<sup>3</sup>. The institution B is deemed to be inefficient because it does not lie on the isoquant; instead, it is using an excessive amount of inputs for the production of the given level of output. Assuming that B is producing the same level of output  $y$  that lies in the efficient isoquant, the technical

---

<sup>3</sup> In this example, an input oriented approach is used (i.e. the level of outputs is fixed, and the analyst measures the potential reduction of inputs for producing it). Therefore, the interested reader can refer to Johnes (2004) for a formal illustration of the analogous problem in an output-oriented framework.

efficiency of the unit B can be measured as the distance from the isoquant, where the point B' illustrates which is the level of inputs that is really necessary to produce – efficiently – the amount of outputs observed in B. In formal terms:

$$TE_B = OB' / OB \quad (1)$$



**Figure 1. Measuring technical efficiency in an input-oriented framework**

Thus, the degree of technical inefficiency (which is calculated as  $1-TE_B$ ) is the amount of inputs that can be reduced for obtaining the same level of output that is now (inefficiently) produced. It is important to note here that the estimation of technical efficiency is made by assuming that all units (schools or universities) are experiencing the same returns to scale for the various inputs – that is to say, no scale effects are present for any input; this is quite a heroic assumption, that will be relaxed when introducing the alternative viewpoint, in the section § 2.2, where we define the concept of “scale” efficiency.

To define the concept of allocative efficiency, the relative price of the two inputs  $x_1$  and  $x_2$  should be introduced, and in the figure this is represented by the line  $zz'$ . As can be observed, whilst both B' and B''' are technically efficient solutions, only the latter is allocative efficient, in the sense that the level of output can be obtained with the best combination of inputs – in other words, with the combination of inputs that minimize their

(relative) prices. The degree of allocative efficiency is measured by the distance between the isoquant  $ss'$  and the line  $zz'$ ; mathematically, allocative efficiency  $AE_B$  is measured as:

$$AE_B = OB''/OB' \quad (2)$$

Lastly, the overall efficiency  $OE_B$  combines the information derived from technical  $TE_B$  and allocative efficiency  $AE_B$ ; in mathematical terms:

$$OE_B = OB''/OB \quad (3)$$

The measure of overall inefficiency  $(1-OE_B)$  quantifies how much of input(s) can be reduced to produce the same level of outputs; and provides information about how the mix of inputs can be changed to minimize the relative cost of employing them.

An aspect specifically related with the measurement of efficiency in education must be discussed here. The information about prices is seldom present in educational studies, for various reasons: the lack of schools' autonomy in deciding teachers' salaries (i.e. regulations in salaries), absence of precise data about facilities' and furniture's prices, etc. As a consequence, the vast majority of studies focuses on various versions of technical efficiency, and the number of studies that deal with allocative efficiency is still very limited – notable exceptions are some studies that focus on relative prices of productive factors, such as Grosskopf et al. (2001), Banker et al. (2004) on Texas public school districts, Haelermans & Ruggiero (2013) on Dutch public schools.

## **2.2 Two additional definitions: spending and scale efficiency**

In addition to these baseline definitions, we will use also two variations of the concepts of efficiency, which should be interpreted as ancillary to the three main ones listed above, originally provided by Farrell (1957). The two additional definitions are:

- Spending efficiency, which is analogous to the concept of technical efficiency, but using expenditures as inputs instead of physical units;
- Scale efficiency, which focuses on comparing units with similar levels of inputs.

If we consider a setting where inputs are not measured in physical terms, but instead in expenditure terms, the information that can be derived from the efficiency analysis is an estimate of '*spending efficiency*'. This concept can be defined as the institution's ability to minimize the amount of expenditures for producing a given level of output(s), and/or to maximize the amount of output(s) produced with a certain level of expenditures. If the amount of spending is only available in aggregate, then the very concept of allocative efficiency loses its sense, because it is not possible to distinguish between the different

types of inputs. Instead, if different categories of spending can be identified (for example, instructional, human resources, facilities, etc.)<sup>4</sup>, then spending allocative efficiency can be estimated through the computation of elasticities between inputs and the ratios of their prices. In other words, the main difference between allocative and spending efficiency stems from a different point of view regarding the inputs: in the former case, each input can be employed together with its relative price, while in the latter the inputs are considered together as a sum of various expenditure categories (thus, the distinction between quantities and prices is somehow difficult to be assessed, and no indications about the technical optimality of input mixes can be derived).

A final concept that deserves attention is that of *scale efficiency*, that must be formulated when estimating technical efficiency, by relaxing the constant returns to scale (CRS) assumption made until this moment – in other words, assuming that the ratio between output(s) and input(s) can be different at different levels of production. For instance, in the Figure 2 a typical production function is illustrated (in the simplistic case of one input and one output), where (increasing and decreasing) returns to scale vary according to the input level  $x$  – thus, the frontier (optimal) line of production is  $OF_{VRS}$ ; in contrast, the frontier estimated under the assumption of constant returns to scale is  $OF_{CRS}$ . Taking the incorrect frontier into consideration (i.e.  $OF_{CRS}$  instead of  $OF_{VRS}$ ) would lead to underestimating the (technical) efficiency for school A; indeed as indicated mathematically:

$$TE_{CRS} = A_0A / A_0A'$$

$$TE_{VRS} = A_0A / A_0A'' \tag{4}$$

The measures of efficiency under the two different assumptions can be combined to obtain an estimate of the scale efficiency  $SE$ :

$$SE = A_0A' / A_0A'' \tag{5}$$

In this example, an  $SE$  below 1 indicates that school A operates at non-optimal scale. In fact, an optimal scale of operations, for the specific production process depicted in the figure, is that of school B (with a level of input= $x_B$ ). In other terms, scale efficiency  $SE$  can be interpreted as the distance of the present level of inputs used from the optimal one, once technical efficiency in production is assumed.

---

<sup>4</sup> Especially in the USA, there are some academic studies that classify expenditures along categories and types, and relate them empirically to various measures of educational outputs; see, for instance: Ryan (2004) and Webber & Ehrenberg (2010). For a country level approach, see Gundlach et al. (2001).

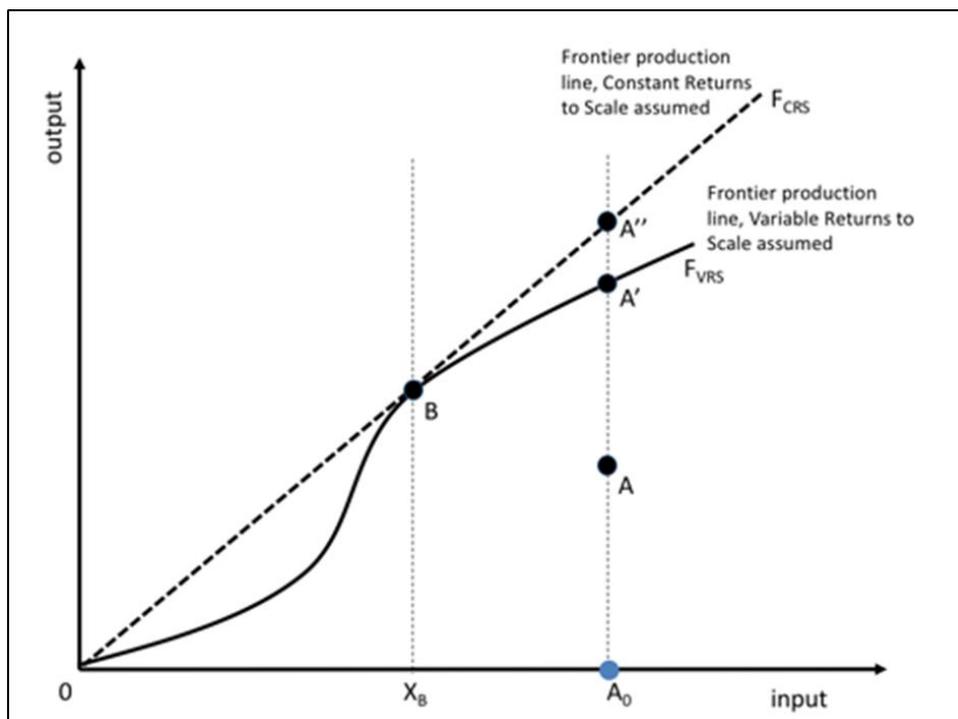


Figure 2. Measuring scale efficiency in a simplified setting with one input and one output

### 3. Measuring educational efficiency in practice: the selection of inputs and outputs

In this section, we provide some discussion about the selection of variables that are relevant for the measurement of efficiency: inputs, outputs and contextual variables. With the latter group of indicators, the analyst aims at describing which factors are statistically associated with higher/lower scores of efficiency (i.e. after they are calculated). While efficiency scores (*eff\_scores*) represent the ability of transforming inputs (for instance, resources) into outputs (for instance, test scores), this second level of analysis investigates whether there are recurrent factors statistically correlated with such scores<sup>5</sup>. As will be explained later, the contextual factors<sup>6</sup> considered as potentially correlated with *eff\_scores* include both (i) descriptions of educational processes (i.e. selected by schools/universities) and (ii) the so

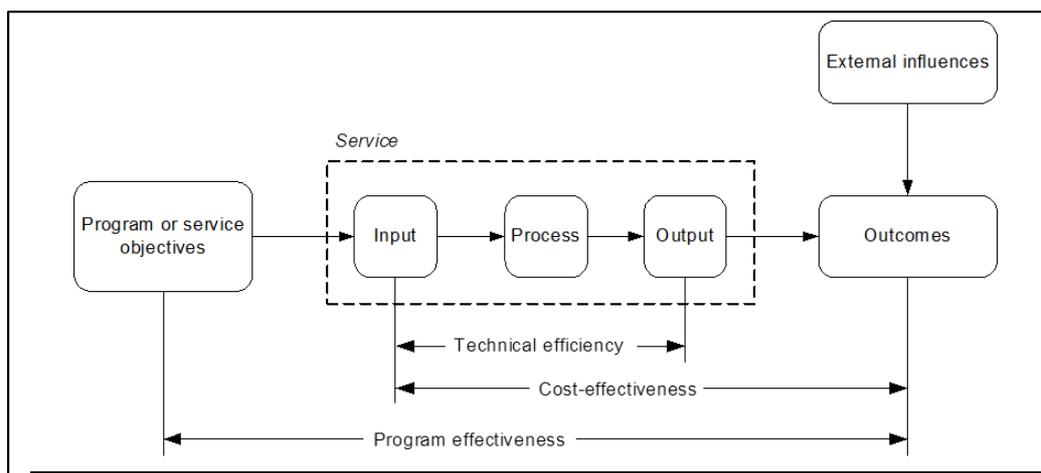
<sup>5</sup> It is important to remark here that such an analysis is correlational in nature, because no causal inference can be realized about how such contextual variables are having a “causal impact” on the efficiency of the organization.

<sup>6</sup> Certain literature labels these contextual variables as non-discretionary factors (in this sense, see Cordero-Ferrera et al., 2008). We prefer the definition of contextual variables (as indicated by Worthington & Dollery, 2002 in their broader discussion about the public sector), because we argue that some of these variables are indeed non-discretionary (i.e. they are ‘external’ in a pure sense), while others (such as managerial and educational processes) can be influenced by schools’ decisions and actions.

called (purely) external variables (i.e. features that are beyond the schools/university control, as for example the socio-economic characteristics of the student population served by the institution).

The selection of inputs and outputs is a crucial task in efficiency studies (Coelli *et al.*, 2005; Cooper *et al.*, 2011); indeed, the ability of describing actual efficiency differentials stems from the precision of the production process. The lack of detailed information about the process itself (efficiency studies do not include a description about how heterogeneous are the educational and managerial processes used by the institutions) poses all the empirical evidence on the shoulders of the relationship between inputs and outputs, and the selection of how defining (and measuring) them is decisive.

Inputs are those factors that are used by the institutions for producing educational services. They can be classified in three broad groups: (i) financial resources (of various types, and with various destinations), (ii) human resources (those devoted to educational activities, and support personnel), and (iii) facilities – that can be consumables or use of infrastructures. Outputs should measure the results of the educational services offered by the institutions. Ideally, such measures should include together indications of quality and quantity of the services produced, and should refer exclusively to the output (i.e. the service produced by the institution) – and not the outcome (i.e. the impact of the output produced). The public management literature, indeed, associates the concept of effectiveness to the comparison between outcomes and inputs (see Figure 3; interesting discussions in Golany & Tamir, 1995; Moore, 1995).



Source: SCRGSP (Steering Committee for the Review of Government Service Provision) 2006, Report on Government Services 2006, Productivity Commission, Canberra.

**Figure 3. The Report of Government Services framework**

Nevertheless, in the educational literature outputs are usually measured as achievement, test scores, graduation rates, etc. – something that is more similar to the effects of the educational services, than to the quantities produced. In this Report, we do not consider the difference between efficiency and effectiveness in this respect, and we acknowledge that the efficiency literature normally considers only outputs into the analyses.

The contextual variables can be divided into sub-groups:

- those that are contextual characteristics of the educational institutions (*features and processes* set by the institution itself). Thus, the institution can indeed modify its efficiency by acting on these levers. In this specific sense, exploring the correlations between efficiency scores and these contextual variables can be useful, as evidence can be used (with caution) to understand which recurrent factors can be found in institutions with higher/lower levels of efficiency.
- those that describe the *external context* in which the institution operates (i.e. the wealth of a territory, the proportion of immigrants residing there, etc.). This second sub-group of variables can be broadly considered as related to factors that are external to efficiency measurement (i.e. the school/HEI cannot modify the features of the place in which it operates, although they have an effect on their operations). Considering this group of factors as a separate group is important to calculate efficiency of schools/universities without the risk of taking external influences into the picture.

Indeed, sometimes the analyst desires analyses of institutions' efficiency net of the impact of contextual variables, that is to say to explore only how efficient the educational 'production' process in the hands of the institutions' managers is. The problem of considering the influence of external variables on efficiency results of educational institutions has been specifically introduced to take into account that inputs are often "non-discretionary", in the sense that schools/HEIs cannot always select their inputs – for example, students – many times because of equity/ethical reasons. Several methods have been proposed to estimate the impact of non-discretionary inputs and/or external contextual variables on outputs' production, and consequently on efficiency (Cordero-Ferrera, *et al.*, 2008). What is important here, beyond the technical aspects, is that it would be unfair to benchmark institutions against each-others, without "levelling the playing field" by considering the heterogeneous environmental harshness that they face (Ruggiero, 2004). Indeed, evaluations that do not consider the role of external variables would have misleading conclusions (Agasisti, *et al.*, 2014). Although the methodological debate about these aspects did not conduct to a conclusive agreement, for sake of simplicity three widely-adopted approaches are mentioned here:

- *the specification of non-discretionary inputs in efficiency estimations*; they are considered as a "constraint" in deriving efficiency scores, so those units that benefit from more convenient conditions do not receive higher scores because this –

adopting adequate procedures for this purpose, as suggested by Ruggiero (2004b) and Estelle et al. (2010);

- the use of second-stage regression to assess the impact (correlation) of contextual variables on efficiency scores, and then use them to “*adjust*” the measures of efficiency for taking the exogenous variables into account – see, for example, the procedure used by De Witte & Moesen (2010) with data at country level;
- the employment of measures of “*conditional efficiency*”, as suggested by Daraio & Simar (2015), in which efficiency scores are “*conditioned*” by external factors which are neither inputs nor outputs under the control of the organization.

An example of the necessity of taking external variables into account is provided here. Should the socioeconomic status of the students (SES) be included as one external (conditional) variable (as in Ray, 1991), or instead as one of the inputs? The advantage of the former solution is that efficiency measurements are not affected by the different composition of students who attend the institution; however, it implicitly assumes that the ratio of transformation of inputs into outputs is independent from students’ SES (which is a heroic assumption). This is relaxed through the second approach, which however comes at the price of considering students’ SES as modifiable by the unit of observation, which is obviously not true (unless the schools can select their students). A method to incorporate SES among inputs in a more credible way is to consider it as a non-discretionary input, which actually seems easier in the context of non-parametric methodologies (see Johnson & Ruggiero, 2014; see section §4 for a presentation of techniques). The problem of considering students’ socioeconomic background in the evaluation of educational institutions’ efficiency appears as more cogent in the context of primary and secondary education, as it is assumed that achievement gaps would be (at least partially) filled in higher levels of education.

While the example just discussed refers to students’ socioeconomic characteristics, there are other variables that deserve the same attention for assessing the institutions’ efficiency, as for instance: school composition (proportion of girls, immigrants, etc.), institution location in urban/rural area, degree of competition, etc. (an attempt of a complete list is contained in De Witte & Lopez-Torres (2015)).

In the remainder of this paragraph, we discuss the selection of inputs and outputs for primary/secondary education<sup>7</sup>. When considering primary and secondary schools as unit of

---

<sup>7</sup> Although the methodological and theoretical framework within which efficiency analyses are conducted is similar across primary/secondary and high education, two main reasons justify the choice of discussing the selection of variables separately. First, standardized test scores are well rare in HE, while they are pretty much diffused in the context of primary and secondary education – and this leads to differences in the way outputs are defined, and consequently efficiency is operationalized. Second, HEIs (and especially universities) are often multi-product organizations, which produce not only (higher) educational services, but also research. In this context, the selection of outputs should reflect this diversity of missions

analysis, the literature converged to the use of some groups of input variables (De Witte & Lopez-Torres, 2015): student-related, family-related, school-related and community-related. Focusing our attention to the studies that consider the institution (and not the individual student) as a unit of analysis, students' own inputs – as well as family ones – are usually averaged by-school.

Students' features usually include psychological and behavioural aspects, among which innate ability would be helpful but is rarely included, because no reliable measures of it are easily available. When at disposal, prior academic achievement is included among inputs, so that the resulting efficiency estimate is a value-added improvement of output, given the existent input. Some surveys use students' questionnaires where questions about motivation etc. are included (because of a lack of available and reliable data), and school averages can help in showing how schools differ in terms of their available raw inputs (i.e. students' human capital). Usually, a set of students' demographic information is usable for describing school inputs: proportion of males/females, immigrant students, students with disabilities, students who were retained in previous years, etc.

The most important variable at family-level is the description of the average socioeconomic status (SES) of students attending the school. There are several ways of measuring students' average SES: parental occupation, family's income, parental education, resources available at home, eligibility for free meals or economic benefits, etc. An alternative approach, when multiple sources of information can be complemented, is to calculate composite indicators about families' socioeconomic and cultural status. The most popular index of this type is the one proposed by the Organisation for Economic Co-operation and Development (OECD), which calculates the index named ESCS (Economic, Social and Cultural Status) of students and schools according to the following framework: "The Programme for International Student Assessment (PISA) index of economic, social and cultural status was created on the basis of the following variables: the International Socio-Economic Index of Occupational Status (ISEI); the highest level of education of the student's parents, converted into years of schooling; the PISA index of family wealth; the PISA index of home educational resources; and the PISA index of possessions related to "classical" culture in the family home" (OECD, 2002).

School-level input variables can both reflect available physical resources (books, building, computers, class, bus, grants, etc.) and expenditures (teaching, research, administrators, supporting staff), and to the extent that prices are accounted for, they represent two sides of the same coin. The number of teachers is a key input employed in several studies, as expressed in various ways – frequently, in the form of students:teachers ratio. As a mean to

---

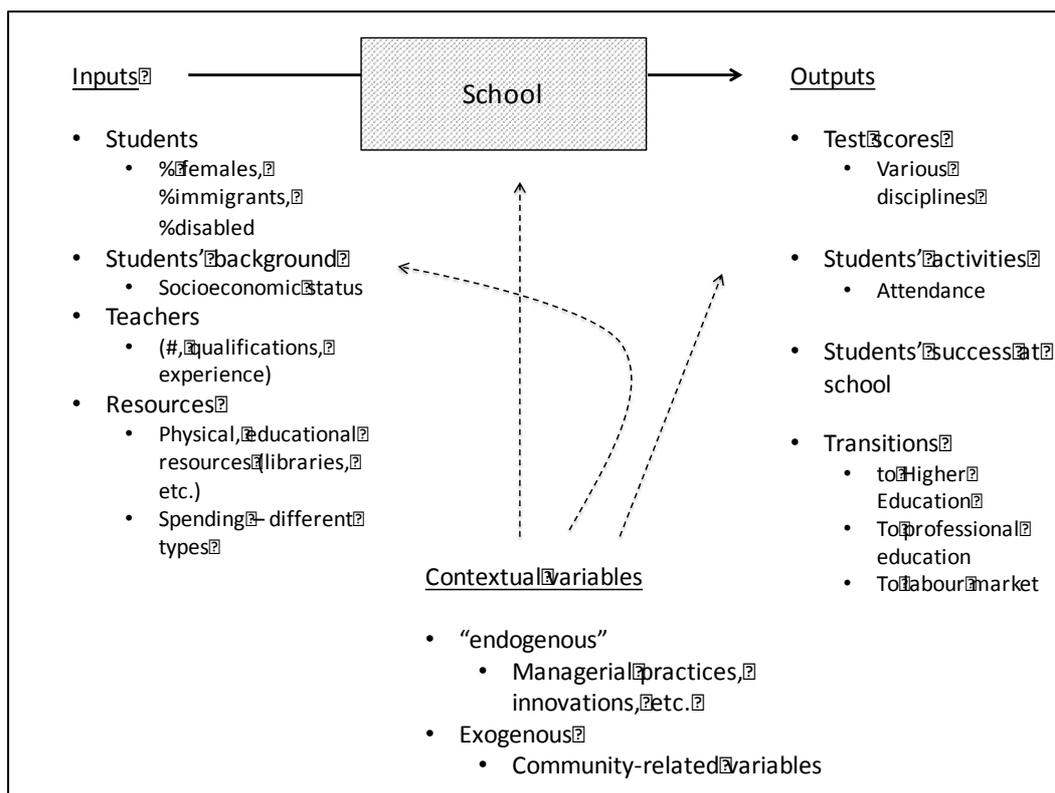
and operations, and the eventual trade-offs and complementarities (i.e. scope economies) between missions and activities, following the methodological indications by Cohn et al. (1989).

control for differences in inputs' quality, sometimes proxies for teachers' experience or qualification are included in the vector of inputs themselves (among others, as in Sarrico, *et al.*, 2010). A growing body of the literature is also paying attention to the role that certain managerial practices, and/or innovations, and/or specific educational processes, can play on affecting outputs (see, for example: Haelermans & De Witte, 2012; Mancebon *et al.*, 2012). Therefore, following the reasoning proposed in the introduction to this §3, these elements are much more classifiable among the contextual variables than among inputs – in other words, they deal more with the use of inputs, and not with the inputs' quantities or qualities. The information about the governance of the school (if it is public or somehow private) is frequently used for comparing the efficiency of public and private schools – paralleling the literature that compares raw performances between these types of schools, see Dronkers & Robert (2008) for an international comparison. Also includible in the group of contextual variables are those that reflect the community in which the schools operate: indicators for competition among schools, neighbourhood characteristics, urban/rural areas, educational level of the population in the area.

Outputs are typically measured through test scores in standardized evaluation of achievement. Some studies, however, also consider other output measures, such as the drop-out rates (Alexander *et al.*, 2010), or the attendance rate (Grosskopf & Moutray, 2001).

Figure 4 graphically represents the educational production process of a primary or secondary school, as potentially considered in the framework of efficiency analysis, while Table 1 reports the main inputs, outputs and contextual factors described in this paragraph.

After having discussed concepts of efficiency and the main issues related with the choice of relevant inputs and outputs, and before entering into details about the techniques available for efficiency measurement, a general point must be clarified here. The study of efficiency is essentially a comparison exercise, which considers the transformation of inputs (resources) in outputs (educational results) as a block box. No clues about the more productive processes are directly provided, and even the analyses about the determinants of efficiency scores provide just indirect information about the solutions to be adopted for improving productivity (i.e., they do not identify causal relationships between certain factors and efficiency itself). In this sense, results from efficiency analyses must be always interpreted as exploratory in nature, and do not support any specific organizational setting or “best solution” to be adopted. The correct perspective of analysis, then, should accompany efficiency analyses with other econometric and statistical techniques which corroborate the findings with a more analytical identification of mechanisms behind the efficiency of educational activities.



**Figure 4. Inputs and outputs of the educational production process (Primary and secondary schools)**

Variable's group	List of potential indicators
<b>Inputs (student-related)</b>	Innate ability, prior achievement, gender, age, disability, immigrant status
<b>Inputs (family-related)</b>	Parental occupation, education; socio-economic and cultural status,
<b>Inputs (school-related)</b>	Physical resources (books, facilities, ICT instruments); human resources (teachers and their characteristics and qualifications)
<b>Outputs</b>	Test scores, drop-out rates, success (attrition) rates
<b>Contextual variables (external)</b>	Public/private status; Socio-economic variables of the territory where the school operates.
<b>Contextual variables (internal)</b>	Educational processes (for instance, structure of curriculum, use of ICT) and managerial practices (for instance: actors involved in decision-making). School and class sizes

**Table 1. Examples of inputs, outputs and contextual variables for analysing schools' efficiency**

## 4. Methodological approaches for assessing efficiency in education: non-parametric methods vs stochastic frontier models

In this Section, we review the main frontier methods, i.e. non-parametric and parametric<sup>8</sup> approaches. Section 5 will be devoted to multi-criteria evaluation, which can be considered a complementary approach, thus particularly useful for robustness analyses. Among the non-parametric methods, we indicate Data Envelopment Analysis (DEA) as the most popular, while Stochastic Frontier Analysis (SFA) is indicated as the most used approach within the group of parametric ones. It is again important to recall, here, that we are considering mainly technical efficiency, when not differently indicated.

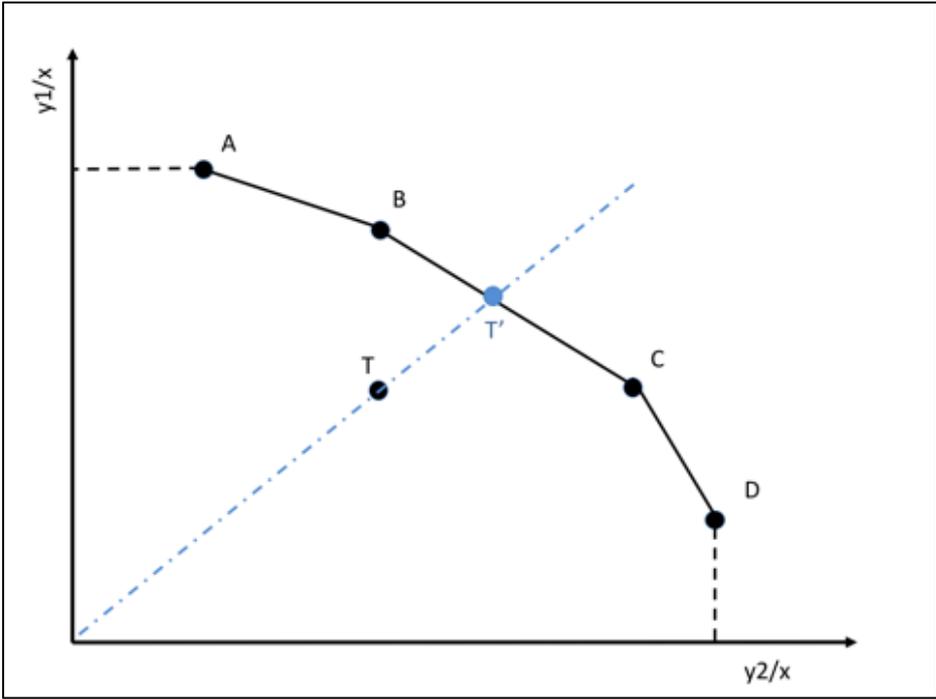
### 4.1 Non-parametric methods: Data Envelopment Analysis

The basic idea at the core of the Data Envelopment Analysis (hereafter, DEA) is to assess by how much output can be increased, given the available inputs (output-oriented models) or, conversely, by how much inputs can be reduced given the produced output (input-oriented models). The method is very useful in a multi-input / multi-outputs context, because the technique can handle several inputs and outputs at the same time, collapsing the judgment about the efficiency in production in single-number indicator. Also, the method is completely non-parametric, because it does not employ any functional form of the production process – this is also a nice property, given that the knowledge about the educational production function is still very limited, and assumptions about the relationships between inputs and outputs can be sometimes non-verifiable.

A graphical illustration of the DEA functioning is useful here. Let us consider a simplified setting where five schools are operating (A, B, C, D and T), which produce two outputs (for instance, reading test scores  $y_1$  and mathematics test scores  $y_2$ ), using a single input  $x$  (for instance, measured through the inverse of students:teachers ratio). By computing two ratios ( $y_1/x$  and  $y_2/x$ ), the positioning of each school can be reported in a Cartesian graph (see Figure 5). Four out of five schools, namely A, B, C and D can be deemed efficient, because there are no other schools able to produce more outputs (i.e. a higher combination of outputs, in this context), given the available input. Instead, T is an inefficient school, as it can produce a higher level of output(s) using the same amount of input. The degree of inefficiency can be obtained by projecting the level of production of T towards the frontier of efficient solutions, in a point that is indicated as  $T'$ , which measures the radial distance of T from the efficiency frontier, assuming that the frontier is convex – in other words, between B and C all efficient solutions of production do exist. As can be noted, the degree of inefficiency can thus be measured as  $OT/OT'$ , a number which is comprised between  $[0;1]$ .

---

<sup>8</sup> At the end of this paragraph, we also provide brief information dealing with some recent advancements in robust non-parametric techniques and semi-parametric techniques.



**Figure 5. A diagrammatic representation of DEA**

As can be noted, the measure of efficiency for T is a relative one, in the sense that it is not derived from a production function described a priori (i.e. in absolute terms), but instead as a comparison between T's actually performance and that one observed in the group of units to which T is compared against – they are the ones that are used for building the efficiency frontier that is the benchmark for each school. The same characteristics of the model illustrate why the model is deterministic in nature: any measurement error, as well as any change in the composition of the sample of units analysed, generates modifications in the calculation of efficiency frontier, and consequently alter the computation of each unit's efficiency score.

Mathematically, the DEA method can be illustrated as a problem of maximizing the ratio between the sum of outputs and the sum of inputs for each institution (sums of inputs and outputs are obviously standardized for accounting for different units of measurement). We first define the technical efficiency of each i-th institution ( $eff_i$ ) as follows, considering  $y_o$  outputs [with  $o=(1,...,s)$ ] and  $x_j$  inputs [ $j=(1,...,m)$ ], and  $w_o$  and  $v_j$  the weights for the  $o$  outputs and  $j$  inputs, respectively:

$$eff_i = \frac{\sum_{o=1}^s w_o y_{oi}}{\sum_{j=1}^m v_j x_{ji}} \tag{6}$$

Then, DEA efficiency score of each i-unit is the one that maximize the unit's efficiency score, by combining the weights in the optimal way:

$$\max \frac{\sum_{o=1}^s w_o y_{oi}}{\sum_{j=1}^m v_j x_{ji}} \quad (7)$$

In this sense, the resulting efficiency score is the one that sheds the best possible light on the i-th institution's performance. For obtaining the efficiency scores, the fractional problem illustrated in the equation (x) is transformed in the dual one, and then solved with linear programming. Specifically, a typical DEA formulation in one where:

$$\max \phi_i \quad (8)$$

subject to:

$$\phi_i y_{oi} - \sum_{z=1}^n \lambda_z y_{oz} \quad o = 1, \dots, s \quad (8a)$$

$$x_{ji} - \sum_{z=1}^n \lambda_z x_{jz} \quad j = 1, \dots, m \quad (8b)$$

$$\sum_{z=1}^n \lambda_z = 1 \quad (8c)$$

$$\lambda_z \geq 0 \quad \forall z = 1, \dots, n \quad (8d)$$

The value  $TE = 1/\phi_i$  represents the efficiency score of the i-th unit, and is constrained to be, mathematically, in the range [0;1]. The formulation above is about a model that is called "output-oriented", meaning that the main assumption is that the unit under observation (i.e. the school, the university) is trying to maximize the outputs (attainment, test scores, graduation rates, etc.) with the available resources (personnel, facilities, etc.). A converse problem can be specified, assuming that the unit is instead minimizing the used input for producing the given level of output(s); in some empirical exercises, it has been argued that such an approach (called "input-oriented") is more adequate for circumstances where input reductions (i.e. budget cuts) are in action (Cuhna & Rocha, 2012), such those of the recent financial crisis<sup>9</sup>.

Another important choice to be made about the specification of the DEA model is about constant or variable returns to scale (where the equation 8 illustrated above is considering Variable Returns to Scale, VRS model). The idea behind the different assumptions about the returns to scale is to compare each school/university with all the others in the sample (Constant Returns to Scale, CRS formulation) or instead more with those that have a similar level of output (VRS). A graphical illustration is presented in the Figure 6, where the simplest case of one input vs one output production is considered. As can be noted, while unit B is efficient whatever the assumption on returns to scale of operations, units A and C are

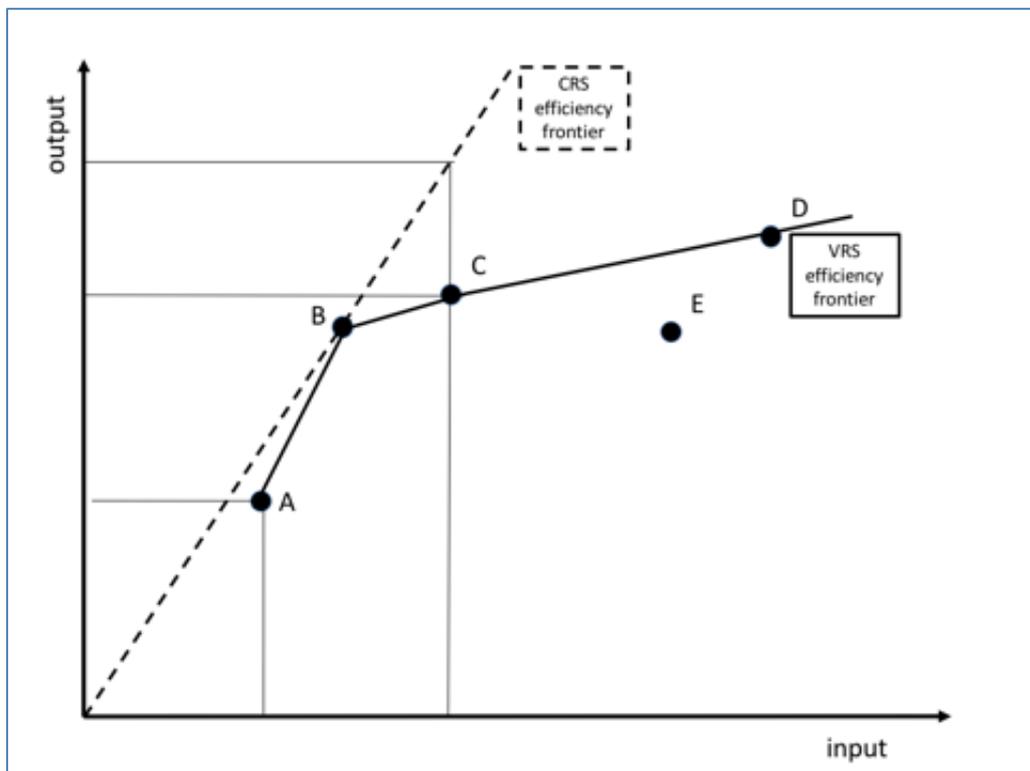
---

<sup>9</sup> A mathematical formulation of the input-oriented problem can be found in Johnes (2004). However, almost all the manuals that deal with DEA methods do discuss the differences between input and output orientation of the analysis. See, among others, Charnes et al. (2013) and Zhu (2015).

inefficient if benchmarked against the CRS frontier. The scale efficiency ( $TE_{scale}$ ) can be then considered an indicator about how far is the  $i$ -th unit from the optimal level of output that is expected to be produced, given the level of input(s) available, and can be computed as follows:

$$TE_{scale} = TE_{CRS} / TE_{VRS} \quad (9)$$

where  $TE_{CRS}$  is the (technical) efficiency score computed under the assumption of constant returns to scale, and  $TE_{VRS}$  is that computed under variable returns to scale. By construction,  $TE_{CRS} \geq TE_{VRS}$ , so that  $TE_{scale} \leq 1$ , in other words this measures how far is each unit from the segment of the efficiency frontier that includes all the units with similar level of inputs/outputs.



**Figure 6. DEA representation under Variable Returns to Scale (VRS) or Constant Returns to Scale (CRS) assumption**

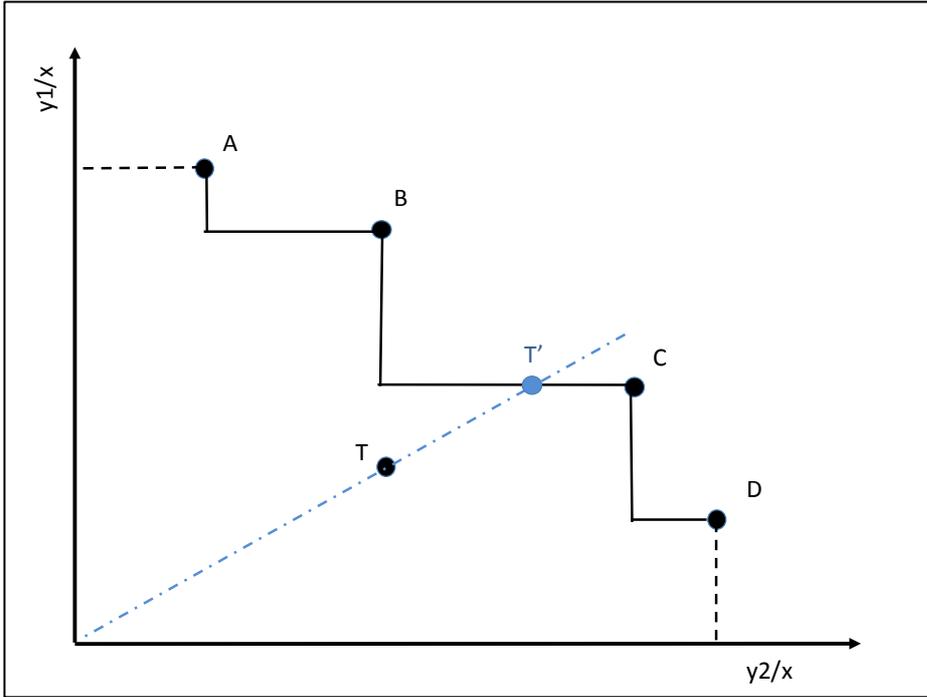
When compared to the parametric methods for evaluating the efficiency, DEA shows some important advantages<sup>10</sup>: (i) it can employ several inputs and outputs at the same time, (ii) it does not require a specification *a priori* of the functional form for the production function, (iii) it allows each unit to have its own objectives, through the free/automatic determination of weights for each input/output, and (iv) efficiency is determined by using observed performance levels, that is (linear combination of) real units operating in the sector, so that they constitute a real (achievable) reference point. These advantages come at a cost, nevertheless. First, the method is completely deterministic, that is any deviation of the units from the frontier is considered as fruit of inefficiency, whilst it can well be due to measurement errors and random noise – and there is no way to check this (as a consequence, efficiency scores cannot be considered in second stages for inferential analyses of their determinants). Second, although the method is good for incorporating multiple inputs and outputs simultaneously, the method does not consider the possibility for estimating economies of scope.

A related method for the evaluation of efficiency through a non-parametric approach is Free Disposal Hull (FDH). The intuition behind the approach is analogous to the one presented for DEA, but with the notable difference that the “convexity assumption” is relaxed. In other words, the method does not assume that linear combinations of inputs and outputs are possible, and the frontier is then estimated only by using existent units as a benchmark. A graphical representation is proposed in the Figure 7; it reproduces the same context of the Figure 5, but from it can be understood that efficiency of the inefficient unit is based on the frontier that is designed without the convexity assumption to connect efficient units (for a comparison of relative merits of DEA and FDH, see Worthington, 2001).

In general terms, DEA (and FDH) is intended to measure efficiency in a cross-section of data, that is to say this is not useful for the analysis of efficiency evolution over time. The literature about the use of parametric efficiency measurement attempted to solve this problem, however, and some approaches have been proposed. Among them, one of the most popular one is the use of the so called Malmquist Index – MI (Tone, 2004). The empirical setting starts by acknowledging that efficiency can vary over time (i) in a asymmetric way (that is to say, some units increase their efficiency, while others do not or even decrease it), and (ii) that efficiency variations in a non-parametric framework can derive by “pure” efficiency improvements (i.e. increasing the unit of outputs produced given the inputs) or by “frontier shifts”, which are technology shocks that affect all the units of the sample – although with different intensity and direction.

---

<sup>10</sup> A discussion of the relative advantaged and drawbacks of DEA and SFA can be found in Johnes et al. (2005), as well as in the literature review provided by Worthington (2001) and, in a more systematic fashion, in Fried et al. (2008).



**Figure 7. A diagrammatic representation of DEA**

For understanding these two different components, let us assume that a school produces two outputs  $y_1$  and  $y_2$  using a single input  $x$ . Let us assume that technical efficiency in time  $t$  can be measured with reference to the technology of production available at that time, so that  $TE_t = D_t(x_t, y_{1t}, y_{2t})$ , while the technical efficiency in a second time (denoted  $T$ ) can be calculated, as a cross-section, referring to the technology available at that time  $T$ , so that  $TE_T = D_T(x_T, y_{1T}, y_{2T})$ . Of course, simulations can be made to calculate technical efficiency at time  $t$  assuming the technology available in time  $T$ , that is  $D_t(x_T, y_{1T}, y_{2T})$ , and vice versa  $D_T(x_t, y_{1t}, y_{2t})$ . Combining these various measures, it is possible to derive an index that expresses variation of efficiency over time, by describing *how* efficiency varied between two periods  $T$  and  $t$ . This index can be then constructed as the product between two components, which are efficiency change ( $Eff\_change_{T,t}$ ) and frontier shift ( $Front\_shift_{T,t}$ ), so that:

$$Malmq_{T,t} = Eff\_change_{T,t} \times Front\_shift_{T,t} \quad (10)$$

The two components  $Eff\_change_{T,t}$  and  $Front\_shift_{T,t}$  are calculated as follows (see the name of inputs and outputs above):

$$Eff\_change_{T,t} = \frac{D_T(x_T, y_{1T}, y_{2T})}{D_t(x_t, y_{1t}, y_{2t})} \quad (10a)$$

$$Front\_shift_{T,t} = \left[ \frac{D_T(x_t, y_{1t}, y_{2t})}{D_t(x_t, y_{1t}, y_{2t})} \times \frac{D_T(x_T, y_{1T}, y_{2T})}{D_t(x_T, y_{1T}, y_{2T})} \right]^{\frac{1}{2}} \quad (10b)$$

The Malmquist index  $Malmq_{T,t}$  can then assume values higher or lower than 1, indicating that the resulting efficiency increased or decreased in the period under scrutiny, respectively; and such an index is determined by a product of the two components that also have values higher or lower than 1, to signal whether pure efficiency changed positively or negatively over time, and whether technology shocks did affect production and efficiency in a positive or negative manner. There are several recent applications of Malmquist indexes to the educational field, among which we highlight, in this Report, three recent examples. Parteka & Wolszczak-Derlacz (2013) applied a (statistically robust) MI to a sample of HEIs in a European comparison, to find that efficiency evolved very differently between HEIs of different countries. Essid et al. (2014) reveal that the productivity of Tunisian schools did not improve in the period of early 2000s that was analysed (2000-2004). Agasisti (2014) assessed the efficiency of public spending on education at country level (area of analysis: Europe), between 2006 and 2009, and found no evidence of any detectable, statistically significant efficiency change.

#### **4.2. Parametric methods: Stochastic Frontier Analysis (SFA)**

The parametric analysis of the efficiency is based on the assumption that it is possible to specify the production function of education, by individuating those factors that affect the performance of the  $i$ -th school/university, so that:

$$y_i = f(\bar{X}_i) + \varepsilon_i \quad (11)$$

where  $y_i$  is the measure of performance, and  $X_i$  is a vector of input characteristics. Particular attention is paid to the error  $\varepsilon_i$ ; indeed, in their seminal work, Aigner et al. (1977) suggest to decompose it to consider the possibility of inefficiency in production. Mathematically:

$$\varepsilon_i = v_i + u_i \quad (12)$$

where  $v_i$  is assumed to be the usual random noise with a distribution  $v_i \sim N(0; \sigma_v^2)$ , whilst  $u_i$  is one-sided: it represents the deviation from the frontier, and can be used for estimating the efficiency score of each  $i$ -th school/university,  $TE_i$ . The distribution of  $u_i$  must be defined by the analyst, and several hypotheses have been proposed in the literature, ranging from half-normal to exponential. The coefficients of the production function are estimated through maximum likelihood methods<sup>11</sup>.

---

<sup>11</sup> For an exhaustive and detailed treatment of the methods for measuring efficiency through a parametric approach, based on econometric theories and techniques, the interested reader should refer to Greene (2008).

In the baseline formulation, reported in the equation (11), the production function can accommodate only one output at a time, and this traditionally constituted a shortcoming, given the multi-output nature of the educational activities. This is particularly true for the case of Higher Education, as indicated in previous sections, and leads to many studies about HE based on cost functions (where costs are estimated to be function of output levels and input prices) instead of on production functions (where outputs are directly estimated to be function of inputs) – see, for example: Cohn et al. (1989); Izadi et al. (2002); Stevens (2005). The methodological problem is today solved by employing parametric distance functions, that can be used for employing several inputs and outputs simultaneously, maintaining the stochastic nature of the analysis – for an application in education, see Perelman & Santin (2011), who estimated the efficiency of educational production of Spanish students using OECD-PISA data.

At the same time, the efficiency scores obtained from a SFA have statistical properties, and can be used for inferential aims. Among the various models proposed for this purpose, the one developed by Battese & Coelli (1995) has been widely used for studying the determinants of schools/universities' (in)efficiency – see, for example, Kuo & Ho (2008), Kempkes & Pohl (2010), Cordero Ferrera et al. (2011). The idea is to regress the efficiency scores estimated for each unit  $TE_i$  on a set of so called “external” (environmental) variables  $z_i$ , that can be considered ‘explanatory factors’ of inefficiency in production; depending on exact specifications, they can be introduced directly in the parametric specification and jointly estimated when deriving efficiency indicators (for a deeper explanation of SFA, see Greene, 2008). In formal terms:

$$y_i = \alpha X_i + (v_i - u_i) \quad (13)$$

$$u_i \sim N(m_i, \sigma_u^2) \quad (13a)$$

$$v_i \sim N(0, \sigma_v^2) \quad (13b)$$

$$m_i = \beta z_i \quad (13c)$$

The (13c) illustrates how the mean of the distribution of the inefficiency term can be modelled as a function of a series of explanatory variables.

A further issue in estimating efficiency through the parametric approach is the choice of the functional form for the production (or cost) function. The choice of the best functional form for the Educational Production Functions (EPFs) is an evergreen in the economics of education literature, and many scholars have attempted to define EPFs both theoretically and empirically (Hanushek, 1979; Figlio, 1999; Todd & Wolpin, 2007). The problem is striking

especially when considering universities as units of analysis<sup>12</sup>, where the multiproduct nature should be considered for obtaining estimates of scale and scope effects; in their literature survey, Cohn & Cooper (2004), building on seminal work by Baumol *et al.* (1982) conclude that there is not a guideline theory to consider specific functional forms superior to others. In many cases, the trend is towards the use of more flexible forms, that allow to relax many of the assumptions behind the statistical relationships between inputs and outputs, such as quadratic forms or translog, as in Ruggiero & Vitaliano (1999) or Mensah & Werner (2003). Mathematically, a translog production function, for a process where an output  $y$  is produced using two inputs  $x_1$  and  $x_2$ , can be expressed as follows:

$$\ln(y) = \ln(\alpha_0) + \alpha_1 \ln(x_1) + \alpha_2 \ln(x_2) + \alpha_3 \ln(x_1) \ln(x_2) \quad (14)$$

The main interesting technical characteristics of SFA is that it allows formulation of hypotheses about the production function, and the findings can be used (in addition to efficiency considerations) to explore those topics that are traditionally interesting for economists who deal with production, such as unit and marginal costs, elasticity of output(s) to different inputs, returns to scale and – in multi-product settings – returns to scope. Examples about the traditional use of production paradigms in education are in Koshal & Koshal (1995; 1999 and 2000) or Laband & Lentz (2003) for studies about US colleges, Worthington & Higgs (2011) for Australia, Hashimoto & Cohn (1997) for Japan, and Glass *et al.* (1995) and Johnes (1997) for United Kingdom.

#### **4.3. Some recent advancements in methodology**

While the previous two sections §4.1 and §4.2 outline the most frequently used traditional frontier based tools for the measurement of efficiency, in this section, we list some interesting developments of the recent methodological literature advancements:

- The introduction of statistical properties into DEA, deterministic efficiency scores by means of bootstrapping procedures (see Simar & Wilson, 2000);
- The development of robust non-parametric estimates of efficiency, following the work by Daraio & Simar (2007);
- The use of advanced parametric methods for estimating efficiency in presence of heterogeneity across units in the way they realize the production (educational) process, as suggested by Tsionas (2002) and Greene (2005).

---

<sup>12</sup> The specific problem discussed here is relevant for universities, as they produce teaching and research jointly and simultaneously; of course, the same identical problem affects the analysis of schools' efficiency when considering their multiple outputs at the same time (i.e. the joint production of test scores in different domains/disciplines).

Given the highly technical content of these methodological discussions, we decided not to go into much detail, given that the main focus should be on policy-related aspects of efficiency analysis in education (and not technical refinements about the methods for estimating efficiency in itself). Thus, this report only introduces the main points about the current debates, and the interested reader should refer to the cited bibliography for more profound analyses of the technical, methodological aspects. Such advancements have been, however, already applied in some research about educational efficiency. In this perspective, introducing these discussions allows to derive practical information about new research approaches in this field.

The introduction of statistical properties into DEA has been justified for solving the problems related with the deterministic nature of the method. In the Simar & Wilson (2000)'s words: "(...) despite a small but growing literature on the statistical properties of DEA estimators, most researchers have used these methods while ignoring the sampling noise in the resulting efficiency estimators, and continue to do so" (p. 795). The method of bootstrapping the efficiency scores allows calculating confidence intervals around the estimated specific scores. This is primarily essential to judge the relative performance of the units adequately – that is, by clarifying which are the units that really outperform (or underperform) their counterparts in a statistically significant way. This bootstrapping approach is also helpful to derive information about the determinants of efficiency through second-stage regressions; while often academic studies run this type of second-stage regressions (where the dependent variable is the efficiency scores derived through DEA), the method is somehow questionable. As explained in Simar & Wilson (2007): "Since the DGP (Data Generation Process) has not been described, there is some doubt about what is being estimated in the two-stage approaches" (p. 32). As a consequence, the authors propose a novel method – based on a double-bootstrap procedure – that permits to derive consistent results of determinants of DEA efficiency scores. While the method has been advocated and used also in the recent literature about educational efficiency – see, for instance, Alberta Oliveira & Santos (2005), Afonso & St. Aubyn (2006) Alexander et al., (2010) – the methodological debate about validity and tools for second-stage regressions is still open (see McDonald, 2009). The methodological discussion is of primary interest for policy making and management in the educational field; indeed, the robustness of the findings about factors that correlate with efficiency in operations can suggest policy initiatives and/or managerial settings that promise superior results with same resources, or expenditure savings for the same level of outputs.

The book by Daraio & Simar (2007) describes solutions for developing robust non-parametric techniques for assessing efficiency. After having illustrated the steps proposed by Simar & Wilson (2000) for introducing statistical properties in non-parametric estimates of efficiency (through bootstrapping), the authors review three other ways for solving traditional drawbacks of the DEA approach. One is the adoption of order-m frontiers, that

use a robust approach for not using all the observations in deriving the frontier of efficient possibilities, and obtain this way efficiency estimates that are not influenced by outliers. Another method consists in calculating parametric approximations of the non-parametric frontier (an approach proposed by Daouia & Simar, 2005) – the aim of this technique is to obtain parameters' coefficients that can be used for statistical inference and economic considerations. A third innovative proposal consists in robust conditional (non parametric) frontier methods, as suggested by Daraio & Simar (2005); these frontiers can analyse and measure the effect of external environmental variables on the efficiency, in a way that overcomes the main problems associated with the traditional two stages.

Lastly, econometric methods based on stochastic frontier analysis for estimating efficiency have been recently advanced for disentangling various components that affect performance: heterogeneity in the production structure, efficiency and unobservable structural differences. In particular, Greene (2005) "(...) propose(s) specifications which can isolate firm heterogeneity while better preserving the mechanism in the stochastic frontier model that produces estimates of technical or cost inefficiency" (p. 270)<sup>13</sup>. The general idea behind these advancements is that the observed performance levels, as well as the estimated efficiency in production, can be determined not only by differences in operations, but also by (un)observable differences in the production technology and structure. If specific schools/universities are structurally different from those with which they are compared, then it is not legitimate to realize a straightforward benchmarking. Instead, the empirical modelling should aim at estimating production functions (and inefficiency) while separating out the structural differences that make the unit heterogeneous. The methods proposed by Tsionas (2002) and Greene (2005) pursue exactly this objective. Some examples of application in the educational field do already exist, as Johnes & Johnes (2009), Johnes & Schwarzenberger (2011) and Agasisti & Johnes (2010) employ these methodologies for studying the efficiency of universities in UK, Germany and Italy respectively. In general terms, also these methodological innovations can be grouped among those that intend to understand better which factors should be taken into account to avoid overestimating inefficiency, which instead is attributable to different (external, out-of-control or structurally determined) factors others than managerial decisions and operations.

---

<sup>13</sup> The most recent model that attempts at disentangling efficiency and heterogeneity is that proposed by Tsionas & Kumbhakar (2014), where "(...) a new panel data stochastic frontier model disentangles firm effects from persistent (time-invariant/long-term) and transient (time-varying/short-term) technical inefficiency. The model separates firm heterogeneity from persistent or time-invariant technical inefficiency" (p. 128).

## 5. Methodological approaches for assessing efficiency in education: Multi-Criteria Evaluation

In this Section we will first define the main concepts of multi-criteria evaluation, then it will be explained its relevance as a methodological tool for assessing efficiency of education systems. Multi-criteria evaluation approaches can be divided into continuous and discrete approaches. While continuous approaches are still related to frontier methods and they can be considered an attempt of improving traditional DEA techniques, discrete multi-criteria methods are based on complete different assumptions; from this point of view, they can be considered a complementary approach particularly useful to test robustness of results obtained by means of frontier based tools. More technical information is provided in the Annex.

### 5.1 What is multi-criteria evaluation?

Multi-criteria evaluation proceeds on the basis of defining four concepts, namely: objectives, evaluation criterion, goals and attributes (Figueira *et al.*, 2016). *Objectives* indicate the direction of change desired, e.g. growth has to be maximised, social exclusion has to be minimised, education performance has to be maximised. An *evaluation criterion* is the basis for evaluation in relation to a given objective (any objective may imply a number of different criteria). It is a function that associates alternative actions with a variable indicating its desirability according to expected consequences related to the same objective, a classical example in economics might be national income, savings and inflation rates under the objective of economic growth maximisation; in the framework of education policy, PISA scores can be used as criteria for evaluating outputs of an education system and so on. A *goal* is synonymous with a target and is something that can be either achieved or missed, e.g. at least 95% of children (from 4 to compulsory school age) should participate in early childhood education, the rate of early leavers from education and training aged 18-24 should be below 10%. If a goal cannot, or is unlikely to, be achieved, it may be converted to an objective. An *attribute* is a measure that indicates whether goals have been met or not, on the basis that a particular decision will provide the means of evaluating various objectives.

The number of alternatives may vary between 1, any discrete number and infinity. When the number of alternatives is not finite, there is a need to use *Multi-Objective Optimisation*, where the set of options is a continuous non-finite set. In practice these approaches are an extension of classical linear programming, where a plurality of objective functions has to be optimised instead of only one (for more details please see the Annex).

A “discrete multi-criterion problem” can be formally described as follows.  $A$  is a finite set of  $N$  feasible actions (or alternatives).  $M$  is the number of different points of view, or evaluation criteria,  $g_m$ , that are considered relevant to a specific policy problem. Where

action  $a$  is evaluated to be better than action  $b$  (both belonging to the set  $A$ ), by the  $m$ -th point of view, then  $g_m(a) > g_m(b)$ . In this way a decision problem may be represented in an  $N$  by  $M$  matrix  $P$  called an *evaluation or impact matrix*. In such a matrix, the typical element  $p_{ij}$  ( $i=1, 2, \dots, M; j=1, 2, \dots, N$ ) represents the evaluation of the  $j$ -th alternative by means of the  $i$ -th criterion (see Table 2). The impact matrix may include quantitative, qualitative or both types of information. In general, in a multi-criterion problem, there is no solution (ideal or utopia solution) optimising all the criteria at the same time, and therefore “*compromise solutions*” have to be found.

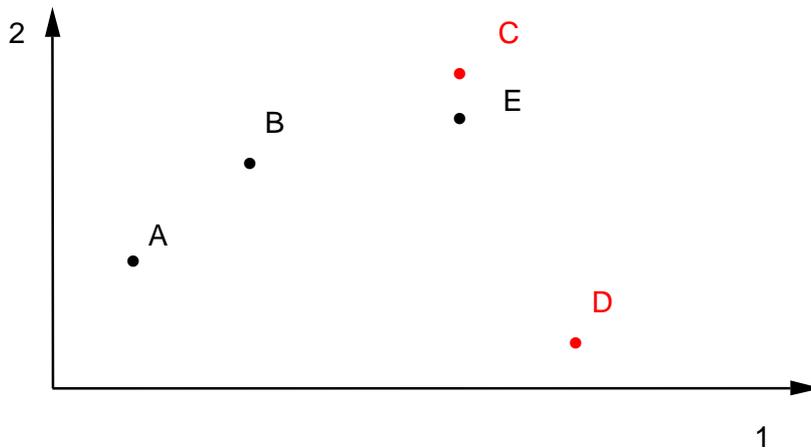
		Alternatives			
Criteria	Units	$a_1$	$a_2$	$a_3$	$a_4$
$g_1$		$g_1(a_1)$	$g_1(a_2)$	.	$g_1(a_4)$
$g_2$		.	.	.	.
$g_3$		.	.	.	.
$g_4$		.	.	.	.
$g_5$		.	.	.	.
$g_6$		$g_6(a_1)$	$g_6(a_2)$	.	$g_6(a_4)$

**Table 2. Example of an Impact Matrix**

### 5.2 Why discrete approaches can be useful for efficiency analyses?

As already noted in the Introduction, in the framework of education policy, the desirability of the peculiar characteristics of multi-criteria evaluation has been advocated by various authors (e.g. Dill, & Soo, 2005; Guskey, 2007; Ho *et al.*, 2006; Malen and Knapp, 1997; Nikel & Lowe, 2010; Rossell, 1993; Stufflebeam, 2001; Tzeng *et al.*, 2007). While continuous approaches are still related to DEA and can be considered an attempt of improving DEA techniques, discrete multi-criteria methods are based on complete different assumptions. From this point of view, they can be considered a complementary approach, particularly useful for testing robustness of DEA results. One of the main reasons of this relationship of complementarity can be found on the fact that the whole concept that *dominated alternatives* can be ignored and thus only *efficient alternatives* have to be taken into account is questioned. It has to be noted that this concept is the key assumptions of all frontier based approaches.

The concept of efficient alternatives can easily be illustrated graphically (see Figure 8 which refers to a 2-criteria state space). Alternative **C** performs better than **B** in all respects and hence **C** is preferred to **B**. The same can be said for **B** compared with **A**. Thus only **C** and **D** are efficient alternatives. It has to be noted that efficiency does not imply that every efficient solution is necessarily to be preferred above every non-efficient solution; e.g., the non-efficient alternatives **A** and **B** are preferable to the efficient alternative **D** if the second criterion would receive a high priority compared to the first criterion. The principle that inefficient solutions may be ignored (often presented as a simple technical step) needs the acceptance of the following assumptions:



**Figure 8. Graphical Representation of Efficiency in a Two-Dimensional Case**

- (1) The assumption that all the relevant criteria have been identified needs to be accepted. If relevant criteria are omitted, there are potential opportunity costs associated with assuming that it is safe to ignore dominated alternatives.
- (2) The assumption that only one alternative considered the best has to be identified needs to be accepted. Since the "second best" may have been eliminated during the technical screening, if more than one action has to be found, the elimination of the "inefficient" action may result in an opportunity loss (one has to note that if the best action is removed from the set of feasible alternatives, then the second best becomes a member of the non-dominated set). If one is interested in the  $\gamma$  problem formulation, then dominated alternatives cannot be eliminated. It has to be noted that in public policies, it is often much more useful to have a ranking of policy options than to select just one alternative.
- (3) A third problem is connected to the question: how relevant are "irrelevant" alternatives? Arrow's axiom of "the independence of irrelevant alternatives" states that the choice made in a given set of alternatives  $A$  depends only on the ordering made with respect to the alternatives in that set. Alternatives outside  $A$  (irrelevant since the choice must be made within  $A$ ) should not affect the choice inside  $A$ . Empirical experience does not generally support this axiom; thus to exclude some actions already inside  $A$  can have even less justification. However, the issue of the independence of irrelevant alternatives is particularly important and tricky when pair-wise comparisons are used. To clarify this point, let's imagine a football championship. To determine the winner all the teams have to compete pair-wise.

Then we need to know the performance of each team with respect to all the others, e.g., how many times a given team won, lost or was even. By using this information, we can finally determine who won the championship. Let's now imagine that when the championship is about to end and the team **X** is going to win (e.g. Barcelona), a new team **Y** is created (e.g. in Madrid). Would it be acceptable to allow this new team **Y** to play directly with **X**? Would the supporters of team **X** accept that if **Y** wins, then **Y** will also win the championship? Of course not!

This example seems to give a clear answer to our problem, but let's now imagine that instead of ranking football teams, our problem is to evaluate the performance of universities. Let's imagine that a study is almost finalized, and university **A** is going to be top ranked; however the study team discovers that an important university institution **Z** was not present in the original data set. Now the question is: can we just compare **A** with **Z** or do we have to make all the pairwise comparisons again? Now the answer is less clear cut. Moreover, let's imagine that the ranking at time  $T$  (without **Z**) ranks university **A** better than **B** and that at time  $T+1$  (when **Z** is considered in the pair-wise comparisons) **B** is ranked better than **A** just because **Z** is taken into consideration! Can this result be acceptable? To answer this question in a definitive manner is very controversial. What we can say for sure is that if pair-wise comparisons are used, it has to be accepted the assumption that the irrelevant alternative **Z** (irrelevant for the evaluation between **A** and **B**) can indeed change the relative evaluation of **A** and **B**. This phenomenon is called "rank reversal".

From these simple examples we can derive some conclusions:

- (1) When pair-wise comparisons are used, this information is not sufficient to derive a consistent ranking. It is necessary to exploit the relationships among all alternatives too. As a consequence no alternative is irrelevant.
- (2) If the set of alternatives is dynamic i.e. new alternatives enter the evaluation process all the pair-wise comparisons have to be done again. It is not possible just to compare the new alternative with the one that was first in the ranking.
- (3) The principle that the final ranking of all the alternatives depends on the relationship among the whole set of alternatives, may cause the effect of rank reversal.
- (4) Finally, a dominated action may be slightly worse than an efficient action, if indifference and/or preference thresholds are used, then the two actions could present an indifference relation (e.g., **C** and **E**).

As a conclusion of this discussion we can state then that, when the set of alternatives is a finite one, it makes sense the use of mathematical aggregation procedures that do not exclude dominated alternatives a priori. In the framework of efficiency analysis, this conclusion implies that results obtained through traditional frontier methods should always

be corroborated by also using non-frontier based mathematical approaches, such as multi-criteria methods. A numerical example is provided in the Annex.

## 6. Conclusion

It is widely understood that the learning process is a complex multi-dimensional issue and it is difficult to apply techniques that are able to capture this complexity and multidimensionality of the educational processes. For instance, it is not sufficient to assume that increasing expenditure will have a positive effect on student performance since what is vital is the way the additional budget is used and the accompanying complementary actions (i.e. if you are buying computers for the classroom you would also need to train the teachers and create platforms for the exchange of suitable academic material).

Also if the intention is to use evidence produced through such methods to enable policy-making, the robustness of the methods proposed and used should be assessed and the assumptions behind their empirical implementation ought to be clearly described. The choice of inputs (under and outside the school control) and contextual variables (such as family socio economic background, peer effects) should be motivated and related to the relevant literature. In addition, the use of aggregate level data of school performance is likely not enough to capture the complexity of the process since averaging does not capture the reality of the learning process.

This technical report looks into efficiency in compulsory education in a cross-country perspective from a methodological viewpoint and describes various methodologies and their relative advantages (i.e. DEA, SFA, MCE). In view of the need to support the policy makers in their difficult role, and as a result of recent advances in methodological issues which raise the robustness of the analysis, the report opens up the debate on the use of various techniques clearly suggesting what limitations are and the way such limitations may affect conclusions and also suggest a cautious manner to interpret the conclusions. It is expected that the study will inspire trust in the thorny process as it will enable all involved stakeholders to be informed of the use made of the methods by experts.

Efficiency analyses, as any other evaluation study, may present a number of risks, such as oversimplification, wrong policy conclusions due to model misspecification, and biased results caused by hidden subjective judgments in the design process. Uncertainty and sensitivity analyses can gauge the robustness of the results obtained and help the framing of the debate around the conceptual framework used, i.e. which representation of reality has been considered. Efficiency scores should be derived through a plurality of methodological approaches:

- Robust non-parametric methods, and stochastic frontier approaches, allow showing the statistical impact of contextual variables on production processes and efficiency. These methods should be employed, together with more traditional second-stage regressions and descriptive analyses, to reveal how efficiency estimates do indeed mask the influence of factors that are beyond the control of educational institutions' management. Such an attention would result in a clearer definition of good practices that are associated with better/higher educational performances – without creating

confusion with the role played by conditions that schools and universities must take as unchangeable, at least in the near future.

- A golden standard of any system for measuring efficiency should consist in defining whether results have the following two properties: (i) are robust to the selection of specific variables (inputs and outputs selection), and (ii) are robust to the selection of a method for efficiency analysis. Multi-criteria evaluation can be useful in tackling both questions since it allows dealing with many variables simultaneously and it is complementary in nature (since e.g. it considers dominated alternatives too, is non-compensatory and can use both qualitative and quantitative scores) with frontier based methods, such as DEA or SFA.
- Both Spearman and Pearson (ranking) correlation indexes can be computed, to give the reader a quantitative impression of the impact of each methodological change on the results. Also, graphical representations of the efficiency scores' distribution can be beneficial, especially if evidencing in histograms or boxplots the areas that are affected by the changes in variables and/or methods.
- Measuring efficiency should consider as much as possible the heterogeneity of institutions' missions and structural characteristics. In this direction, the use of advanced methods for estimating in a separate way the units' efficiency, their heterogeneity and the impact of external factors should be preferred to more traditional approaches of measuring efficiency assuming homogeneity across the various institutions' educational processes, practices and conditions.

To sum up, careful and extensive work in checking the results presented is really necessary, in our view, not only for the reasons of fairly confirming methodological robustness, but also for the practical implications that results can have on real, impactful policy-making inspired by the lessons learnt from efficiency analyses.

## ANNEX

### *Continuous multi-objective approaches useful for efficiency analyses*

The continuous multi-objective problem has been analysed by various authors who have developed a large number of theorems and algorithms (Steuer, 1986). Clearly, in this case the isolation of Pareto efficient solutions is the first step since in this way the number of options is reduced considerably in a meaningful way. However, there are two problems in operating only with efficient extreme point solutions:

1. in problems of a realistic size, the number of efficient extreme point solutions is very large;
2. the decision-maker is not necessarily satisfied with an extreme point solution as an approximation to the most preferred solution.

The above problems create a need to explore the efficient frontier carefully. This is done by means of interactive procedures, which need the following phases:

- search of a candidate for a compromise solution,
- communication to the decision-maker,
- reaction of the decision-maker.

Nijkamp and Voogd (1985) have expressed the following opinion with regard to interactive procedures: "interactive procedures have several benefits. They provide information to the decision committee in a stepwise way, they can easily be included in a dynamic decision environment, they lead to an active role of all participants involved, and a priori specification of preferences or weights is not strictly necessary, although they can be inferred ex post. A limitation of this approach is that the final solution can depend on the procedure followed and especially on the starting solution. In addition for several continuous evaluation methods there is no guarantee that the compromise solution can be obtained within a finite number of interactive cycles, unless it is assumed that the decision committee is acting in a consistent way".

In efficiency assessment studies, multi-objective linear programming (MOLP) has been recently applied for improving DEA techniques. MOLP and DEA share several concepts. Both approaches are grounded on the Pareto efficiency concept. To avoid confusion, the word weights here is used for the weighting coefficients of the objective functions in the multi-objective problem. For the input and output coefficients the word multiplier is used.

In many situations the classical DEA models do not have sufficient discriminant capabilities regarding the efficiency of the productive units (e.g. schools, factories, etc.) considered; in

particular when the number of units under evaluation is not large enough compared to the total number of inputs and outputs (Li and Reeves (1999)). Secondly, a unit can be efficient with non-zero multipliers in a few variables. Note that these two problems are interlaced.

Li & Reeves (1999) presented a multi-objective model (designated by MCDEA) with the aim of mitigating the above referred limitations of DEA. The authors propose a multi-objective approach for DEA in which additional objective functions are included (Charnes, Cooper, & Rhodes, 1978). In DEA, a given unit  $O$  is efficient when the constraint relative to that unit is active and, thus, its slack is nil (a slack is a variable that is added to an inequality constraint to transform it to an equality; it measures the amount of idle resources still remaining in stock at any point in time during the production process). The basic idea is to consider this slack as an efficiency measurement. To restrict multipliers freedom of choice, the MCDEA model takes into account two other objective functions. The first objective function is the classical efficiency maximization; the second is an equity function (a min-max deviation function); and the third one minimizes the sum of the deviations of all units' under analysis.

The solution method proposed by Li and Reeves uses the optimization of weighed sums of the three objective functions using the software ADBASE (Steuer, 1986). Another possibility is to use the TRIMAP package (Clímaco & Antunes, 1987, 1989; Antunes, Alves and Clímaco, 2016), which is an interactive environment dedicated to tri-objective linear programming models. It assists the decision-makers to perform a progressive and selective search of the feasible polyhedron non-dominated region. The aim is helping the decision maker to eliminate interactively the subsets of the non-dominated solutions which are of no interest to her/him, in order to concentrate the search of the preferred solutions. It combines three main procedures: weights space decomposition, introduction of constraints on the objective functions space and direct introduction of constraints on the weights space.

Why TRIMAP is a good interactive tool to analyse the MCDEA model? The TRIMAP package is intended for tri-objective problems. This fact allows for the use of graphical means particularly interesting in dealing with the Li and Reeves MCDEA Model. For instance, besides full numerical information on non-dominated solutions, the indifference regions on the weights space (which is a triangle for the tri-objective case) corresponding to the non-dominated bases previously calculated, are displayed graphically on the triangle. Regarding the MCDEA model, the knowledge of the weights space decomposition, as obtained through TRIMAP, allows the evaluation of the non-dominated solutions stability, being specially interesting the evaluation of DEA efficient solutions stability. Moreover, the eventual existence of optimal alternatives concerning the classical DEA function can also be verified.

Soares de Mello, Clímaco and Ângulo Meza (2009) proposed a TRIMAP-DEA index, in general enabling the complete ranking of the units. A full development of this outline explanation can be found in Clímaco, J. C. N., Soares de Mello, J. C. C. B. and Angulo- Meza, L. (2008).

Concerning applications of TRIMAP-DEA, for instance, an evaluation of highways performance can be found in Clímaco, J., Soares de Mello, J., Ângulo Meza, L. (2010).

### **Main approaches to the discrete multi-criterion problem**

In a discrete multi-criteria problem, there is a range of multicriteria problem formulations, which may take one of the following forms (Roy, 1996):

- ( $\alpha$ ) the aim is to identify one and only one final alternative;
- ( $\beta$ ) the aim is the assignment of each alternative to an appropriate predefined category according to what one wants it to become afterwards (for instance, acceptance, rejection or delay for additional information);
- ( $\gamma$ ) the aim is to rank all feasible alternatives according to a total or partial preorder;
- ( $\delta$ ) the aim is to describe relevant alternatives and their consequences.

Clearly the steps required by such a process need a number of arbitrary unavoidable subjective decisions. The degree of the subjective component may be higher or lower but it is always present. When different conflicting evaluation criteria are taken into consideration, a multi-criteria problem is mathematically ill-defined. The consequence is that a complete axiomatization of a multi-criteria aggregation convention i.e. a multi-criteria method is quite difficult (Arrow and Raynaud, 1986). To deal with the problem correctly pointed out by Arrow and Raynaud, two main approaches can be distinguished.

- a) The attempt to check under which specific circumstances each method could be more useful than others, i.e. the search of the right method for the right problem (e.g., see Guitouni and Martel, 1998).
- b) The attempt of looking for a complete set of formal axioms that can be attributed to a specific method (e.g., Arrow and Raynaud, 1986).

Here, we will try to isolate some properties that can be considered desirable for a discrete multi-criteria method in the framework of efficiency assessment of education systems. In synthesis, the information contained in the impact matrix useful for solving the so-called multi-criterion problem is:

- Intensity of preference (when quantitative criterion scores are present).
- Number of criteria in favour of a given alternative.
- Weight attached to each single criterion.
- Relationship of each single alternative with all the other alternatives.

Combinations of this information generate different aggregation conventions, i.e. manipulation rules of the available information to arrive at a preference structure. The aggregation of several criteria implies taking a position on the fundamental issue of

compensability. As we have already observed in the Introduction, compensability refers to the existence of trade-offs, i.e. the possibility of offsetting a disadvantage on some criteria by a sufficiently large advantage on another criterion, whereas smaller advantages would not do the same. Thus a preference relation is non-compensatory if no trade-off occurs and is compensatory otherwise. It is important to understand that compensability means that in an education efficiency assessment exercise, an improvement in one of the spending side criteria can easily compensate a worsening in an output criterion such as e.g. PISA scores in science! We can safely state that complete compensability is then not desirable for efficiency assessment of education systems.

Let's now try to find other desirable properties. It has been argued that the presence of qualitative information in evaluation problems concerning socio-economic issues is a rule, rather than an exception (Nijkamp et al., 1990). Thus there is a clear need for methods that are able to take into account information of a "mixed" type (both qualitative and quantitative criterion scores). For simplicity, we refer to qualitative information as information measured on a nominal or ordinal scale, and to quantitative information as information measured on an interval or ratio scale. Moreover, ideally, this information should be precise, certain, exhaustive and unequivocal. But in reality, it is often necessary to use information which does not have those characteristics so that one has to face the uncertainty of a stochastic and/or fuzzy nature present in the data (Munda, 1995; 2008). In conclusion, multi-criteria methods able to tackle consistently the widest types of mixed information should be considered as desirable ones. Examples of this kind of methods are REGIME (Hinlopen and Nijkamp, 1990), EVAMIX (Voogd, 1983), NAIADE (Munda, 1995) and Martel and Zaras (1995) method.

From the mathematical point of view, three main methodologies have been proposed to solve the discrete multi-criterion problem: (i) *multi-attribute value and utility theory*<sup>14</sup>, (ii) *outranking methods* and (iii) the *decision rule approach*.

From a theoretical point of view, MAVT is a very elegant and attractive solution to the discrete multi-criterion problem. From the operational point of view, it assumes a decision-maker who always "believes that in a specified decision context there is a particular preference structure that is appropriate for him" (Keeney and Raiffa, 1976, p. 80). MAVT is based on the following hypothesis: in any decision problem there exists a real value function  $V$  defined on the set  $A$  of feasible actions, which a decision-maker wishes, consciously or not, to examine. This function aggregates the different criteria (generally referred to as attributes) to be taken into consideration, so that the problem can be formulated as:

---

<sup>14</sup> In this context, the terminology "value function" is used when preferences are assumed to be under certainty and "utility functions" when probability distributions are present. Here we consider value functions only.

$\max V(\mathbf{g}(a_n))$  such that  $a_n$  belongs to  $A$

where  $\mathbf{g}(a_n)=[g_1(a_n), \dots, g_M(a_n)]$  and  $V(\mathbf{g}(a_n))$  is a value function aggregating the  $M$  criteria.

The role of the analyst is to determine this function. One of the most important policy consequences of using MAVT functions is that complete compensability is always assumed. As stated clearly by Keeney and Raiffa (1976, p. 66) "our problem is one of value trade-offs". Since there exists a function  $V$  by which criteria (attributes)  $g_1, g_2, \dots, g_M$ , can be aggregated, there must also exist functions  $w_{ij}$  (called trade-offs between the  $i$ -th and the  $j$ -th criteria) measuring the amount that a decision-maker is willing to accept on the  $j$ -th criterion to compensate the loss of a unit on the  $i$ -th criterion (an amount which may vary according to the point considered in the criteria space). In practice determining such trade-offs in precise terms is difficult. Often trade-offs are the basis for discussions between the analyst and the decision-maker towards the construction of the function  $V$ . The simplest and most commonly used, analytical form is the linear aggregation rule.

An important point to consider is that each criterion can have its own value function because of the existence of preference independence. This property is a necessary condition for the existence of a linear aggregation rule. From an operational point of view this means that an additive aggregation function permits the assessment of the marginal contribution of each criterion separately (as a consequence of the preference independence condition). The marginal contribution of each criterion can then be added together to yield a total value. This implies that, for example, among the different aspects of the output variables there are no phenomena of synergy or conflict, i.e. preference independence considers each single score being fully unrelated with all the others, but indeed can courage be evaluated as a positive characteristic of a person, without knowing if he/she is a dedicated criminal or an enthusiastic medical doctor? From an education policy point of view, this implies that, for example, interaction among PISA scores in reading and mathematics are not possible. This is rather unrealistic from a scientific point of view. Thus, an interesting research topic is the study of interactions between criteria, when the preference independence condition is inapplicable. Such research is related to the use of non-additive integrals, e.g., the Choquet integral (see Grabisch 1996).

A particular application of MAVT, that is also completely compensatory, is the analytic hierarchy process (AHP) as developed by Saaty (1980). AHP structures the decision problem into levels which correspond to the decision-maker's understanding of the situation: objectives, criteria, sub-criteria, and alternatives. The decision-maker can focus on smaller sets of decisions by breaking the problem up into levels. The AHP is a very widespread approach with many applications and is one of the few methods that explicitly deals with the issue of hierarchy in decision problems. However, we think that AHP is an adequate decision tool only when the decision-maker is clearly identifiable and expresses her/his

preferences and takes responsibility for the decision outcome. This may be the case in entrepreneurial decisions but is hardly ever so in social decisions.

The concept of partial comparability is the basis for the "outranking methods". The most representative ones are ELECTRE (Roy, 1996) and PROMETHEE (Brans *et al.*, 1986). These methods entail aggregating the criteria into a partial binary relation  $aSb$  (an outranking relation) based on concordance and discordance indexes, and then "exploiting" this relationship. Each of these two steps may be treated in a number of ways according to the problem formulation and the particular case under consideration.

To illustrate the method consider Parliamentary voting. The concordant coalition can be considered as the sum of the votes of the members in favour of a given option; according to a majority voting rule this option will be approved if it obtains more than 50% of the votes. According to the normative tradition in political philosophy, all coalitions, however small, should be given some fraction of the decision power. One measure of this power is the ability to veto certain subsets of outcomes. This explains the use of the condition of non-discordance. In practice, the effect of the discordance test is that even if  $M-1$  criteria support the recommendation of choosing  $a$  over  $b$ , this recommendation must not be accepted if only one criterion is against it with a strength bigger than the veto threshold. This implies that in a situation where all input criteria would support a policy option, this option cannot be accepted if one output criterion, e.g. early school leavers, is very strongly against this option. Of course, this depends on the way in which "very strongly" is defined, i.e. the definition of the veto threshold.

In the 1990s some outranking methods were especially designed to address public policy analysis, one of the most widespread being NAIADE (Munda, 1995). It is a discrete multi-criteria method whose impact matrix may include crisp, stochastic or fuzzy measurements of the performance of an alternative with respect to an evaluation criterion. Thus it is very flexible for actual applications. NAIADE can give the following information:

- ranking of the alternatives according to the set of evaluation criteria (i.e. technical compromise solution/s);
- indications of the distance of the positions of the various interest groups (i.e. possibilities of convergence of interests or coalition formations);
- ranking of the alternatives according to actors' impacts or preferences (i.e. social compromise solution/s).

A common property of all outranking methods is that they are partially non-compensatory, thus weights can be used as "coefficients of importance", rather than trade-offs. Let's clarify this important point. The use of weights with intensity of preference originates compensatory multi-criteria methods and gives the meaning of trade-offs to the weights; on the contrary, the use of weights with ordinal criterion scores originates non-

compensatory aggregation procedures and gives the weights the meaning of importance coefficients. (Bouyssou and Vansnick, 1986; Keeney and Raiffa, 1976; Roberts, 1979). In the decision theory literature, the concept of weights usually referred as symmetrical importance, is the following "... if we have two non-equal numbers to construct a vector in  $R^2$ , then it is preferable to place the greatest number in the position corresponding to the most important criterion." (Podinovskii, 1994, p. 241). This concept of weights as importance coefficients is very intuitive and it is how often weights are derived and used. However often there is a theoretical inconsistency in the way weights are actually used and their real theoretical meaning. In fact when one uses a compensatory approach in practice, such as the linear aggregation rule, one has to determine for each evaluation criterion, a mapping  $\phi_i: x_i \rightarrow R$  which provides at least an interval scale of measurement and to assess scaling constants (i.e. weights) in order to specify how the compensability should be accomplished, given the scales  $\phi_i$  between the different criteria (Roberts, 1979). Note that the scaling constants which appear in the compensatory approach depend on the scales  $\phi_i$ , thus they do not characterise the intrinsic relative importance of individual indicators. From this point of view, ELECTRE II (Roy and Bertier, 1973) and ELECTRE III (Roy, 1978) methods are probably the most interesting methods (although the existence of a veto considered as intensity of preference may create some perplexities on the meaning of weights). A method that for sure needs weights as importance coefficients is the REGIME method when only ordinal criterion scores are used (Hinloopen et al., 1983). However, when mixed information is considered (Hinloopen and Nijkamp, 1990), weights are more likely to be considered as trade-offs and not importance coefficients anymore.

Two issues are connected with all the outranking methods, as well as with other approaches based on pair-wise comparisons. First, the axiom of independence of irrelevant alternatives is not respected. Thus the phenomenon of rank reversal may appear (i.e. the preference between  $a$  and  $b$  can change in function of the fact that a third option  $c$  is considered or not). Second, the Condorcet paradox may appear, i.e. alternative  $a$  may be ranked better than  $b$ ,  $b$  better than  $c$  and  $c$  better than  $a$ . In addition, there is a problem specifically connected with the outranking approach. That is the necessity to establish a large number of "preference parameters", i.e. indifference and preference thresholds, concordance and discordance thresholds and weights. This may cause a loss of transparency and consistency in the model. In the framework of policy analysis, outranking approaches look an interesting assessment framework, but to guarantee consistency with the social process behind the problem structuring, the mathematical aggregation rules need to be kept as simple as possible (see Munda, 2008 for a deeper technical discussion on this issue).

Another methodology that has gained interest recently is the decision rule approach. This aims to elicit preference information and supplying a decision model in the simplest possible way. The preference information required is:

- assignments of alternatives to some predefined classes of merits, such as alternatives  $a$ ,  $b$  and  $c$  are “bad”, alternatives  $e$ ,  $f$  and  $g$  are “medium”, alternatives  $h$ ,  $j$ ,  $k$  are “good”, or
- pairwise comparisons such as alternative  $a$  is weakly preferred to alternative  $b$ , alternative  $c$  is fairly preferred to alternative  $d$ , alternative  $e$  is strongly preferred to alternative  $f$ .

The decision model supplied to the decision-maker is a set of easily understandable “if-then” decision rules induced from the above preference information, such as:

- if on attribute  $g_1$  the evaluation is at most medium and on attribute  $g_2$  the evaluation is at most good, then the overall evaluation is at least medium, or
- if on attribute  $g_1$  alternative  $a$  is at least fairly preferred to alternative  $b$ , and on attribute  $g_2$  it is strongly preferred, then alternative  $a$  is comprehensively at least fairly preferred to alternative  $b$ .

The decision rules, induced from the preference information by using the Dominance-based Rough Set Approach (Greco, Matarazzo and Slowinski 2001), give simple explanations to the preference information and can be critically discussed. A decision-maker can accept some of them and reject others, or modify the corresponding preference information. The final result of this process is a set of decision rules accepted as representative of preferences so that these rules can be applied to the whole set of alternatives. Decision rules not only permit the assignment of comprehensive evaluations to the alternatives, or drawing the conclusion that an alternative is preferred over another one, but also supply an explanation for this, thus increasing transparency.

An interesting advantage of the decision rule approach is that it does not entail the use of any preferential parameter such as weights, thresholds or value functions. As we have already observed, there are many difficulties linked to the use of these parameters, and the possibility of getting rid of them may consequently constitute an interesting advance from a policy point of view. Another advantage of this approach is that it does not require fulfilment of demanding conditions such as preference independence.

### **A NAIADE based example of education efficiency assessment**

The NAIADE method can be considered particularly useful for efficiency analyses in the field of education for four main reasons:

1. It has been explicitly designed for public policy applications;

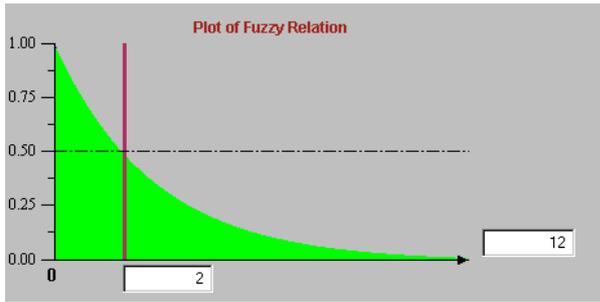
2. it is flexible, since it can deal with different source of information on the criterion scores;
3. compensability can be controlled fully;
4. it can also be used for benchmarking exercises.

The whole NAIADÉ mathematical aggregation procedure can be divided into three main steps:

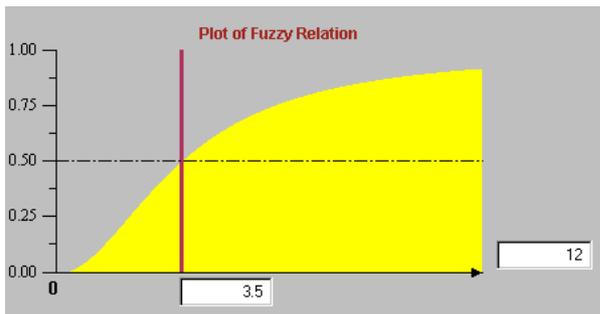
1. pair wise comparison of alternatives according to each criterion,
2. aggregation of all criteria,
3. ranking of alternatives.

Another characteristic of the NAIADÉ approach is the possibility of considering credibility degrees of the preference relations. For example, let's consider the situation described in Figure A1. The value 2, in the example shown, indicates that below this value the difference between the two alternatives is not sufficient to state that one is better than the other. In Figure A2, the value 3.5 indicates that starting from this value, the difference between the options is indeed sufficient to declare that one is better than the other. A problem inherent in the use of precise indifference and preference thresholds is that they can create the strange situation that e.g. in our case up to the value 1.9999 one would conclude that the two options are indifferent and starting from 2.0001 one would definitely state that the preference relation seems plausible. For this reason, credibility degrees of the preference and indifference relations are introduced in NAIADÉ.

Credibility degrees are measured on the *y-axis* (while in the *x-axis* the difference intensities for two options and one single criterion are represented), in the case of indifference they indicate that zero difference intensity makes the credibility equal to 1, and then the greater the difference intensity the smaller the credibility of an indifference relation. This credibility is greater than 0.5 up to the value of the indifference thresholds and smaller than 0.5 starting from the indifference thresholds. The credibility of an indifference relation then necessarily must be a monotonically decreasing function like the one shown in Figure A1. In the case of preference the reverse holds. At zero difference the credibility of preference is zero, then the greater the intensity the more credible the preference relation. This credibility is greater than 0.5 when the preference threshold is overtaken. As a consequence, the credibility degree of a preference relation can only be a monotonically increasing function like the one shown in Figure A2. As one can see thanks to the preference modelling based on the use of credibility degrees, the issue of significance of difference intensities is dealt with properly, and no abrupt transition from indifference to preference is allowed.



**Figure A1 Example of Credibility Degrees of a Fuzzy Indifference Relation**



**Figure A2 Example of Credibility Degrees of a Fuzzy Preference Relation**

Given the information on the pair-wise performance of the alternatives according to each single criterion, it is necessary to aggregate these evaluations in order to take all criteria into account simultaneously. This is done by using a kind of concordance index aggregating the various credibility degrees obtained according to the criteria used. This is done by introducing parameters that allow the user *to establish the degree of compensability* desired in the problem at hand. The final ranking of the alternatives in a complete or partial pre-order ( $\gamma$  problem formulation), is obtained by means of the basic idea of *positive* (i.e. credibility that a number of options is worse than the one considered) and *negative* (i.e. credibility that a number of options is better than the one considered) *flows*.

Any attempt of measuring efficiency should deal with the following two questions: (i) are results robust to the selection of specific variables (inputs and outputs selection)? and (ii) are results robust to the selection of a specific method for efficiency analysis? Discrete multi-criteria evaluation can help in answering both questions since it allows to tackle many variables simultaneously and it is complementary in nature with frontier based methods, such as DEA.

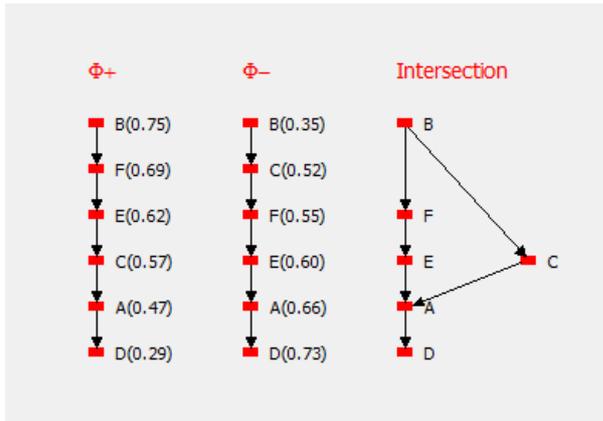
To give an illustrative example, here NAIADe is applied to the impact matrix presented in Table A1. Six imaginary countries have been designed considering a plurality of available data sources for both input and output variables. The result can be considered a

multidimensional measure of efficiency where the various input and output items are integrated all together.

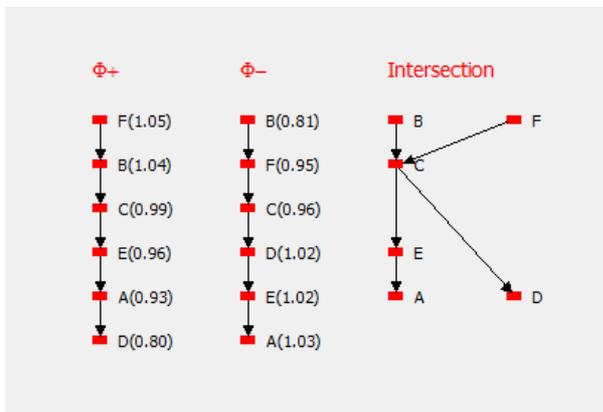
Criteria \ Alternatives	A	B	C	D	E	F
PISA READING	492.000000	494.000000	500.000000	490.000000	515.000000	440.000000
PISA SCIEN	506.000000	510.000000	496.000000	494.000000	520.000000	440.000000
EARSCHLEAV	91.000000	95.000000	74.000000	82.000000	91.000000	82.000000
NEET	91.000000	85.000000	78.000000	77.000000	93.000000	81.000000
EXPGDP	3.300000	2.800000	2.900000	2.900000	3.800000	1.600000
EXPSHARE	6.000000	6.100000	6.000000	5.600000	7.500000	4.200000
EXPGDP-2	3.400000	2.600000	2.500000	2.300000	3.500000	1.700000
EDUATTAIN	83.000000	84.000000	70.000000	87.000000	72.000000	87.000000
EXP-STUD-2	12080.000000	6100.000000	8000.000000	8300.000000	10000.000000	2300.000000

**Table A1. Example of a Possible Impact Matrix for Six Countries**

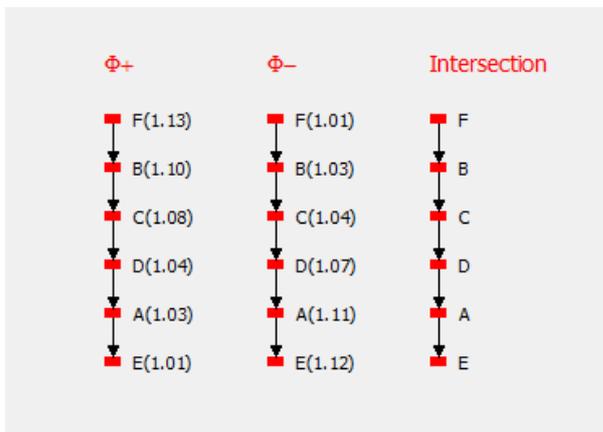
The multi-criteria ranking obtained is presented in Figure A3, where the final ranking is the intersection between rankings obtained through positive ( $\Phi+$ ) and negative flows ( $\Phi-$ ). As one can see, the ranking is very straightforward with the exception of **C** which surely performs worse than **B** and better than **A** and **D**, but it is difficult to compare with **E** and **F**. This difficulty in real-world studies, in general is due to very different characteristics of the objects compared. This ranking has been derived by limiting the compensability among criteria as much as possible. As we have already discussed a low degree of compensability can be considered a desirable property in a multidimensional efficiency assessment of education systems. However, in the search of testing result robustness, it is good practice to check how the final ranking varies if one allows higher degrees of compensability. Figures A4 and A5 show how rankings vary if higher and higher degrees of compensability are allowed in the mathematical aggregation procedure.



**Figure A3 Multi-criteria ranking according to input and output items (Compensability= low)**

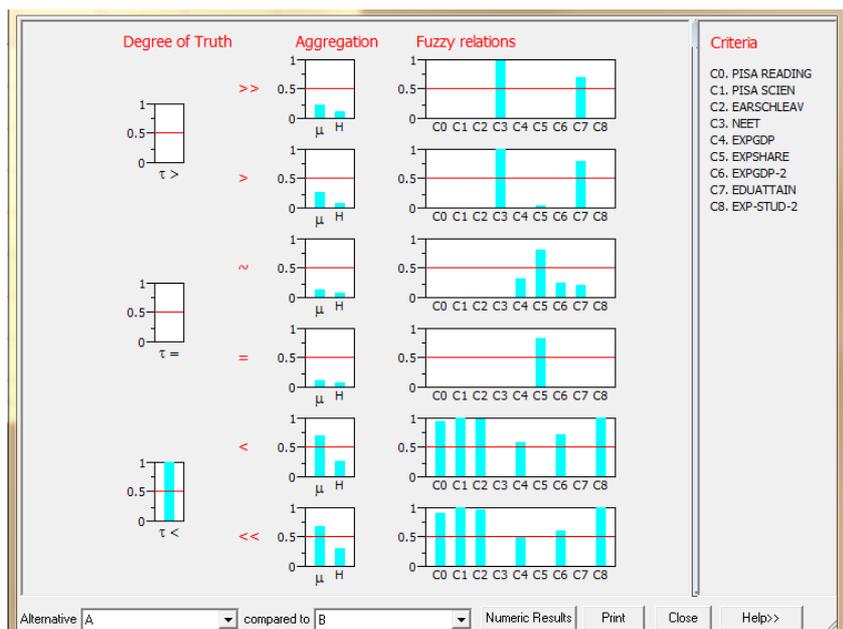


**Figure A4 Multi-criteria ranking according to input and output items (Compensability= medium)**



**Figure A5 Multi-criteria ranking according to input and output items (Compensability= high)**

Finally, one should observe that NAIADe can also be used for benchmarking exercises. In fact it allows the pairwise comparisons between all the options considered; thus the various countries can be compared to e.g. the top performer. These comparisons may have a policy value since one can be fully aware of the mutual weaknesses and strengths on each single evaluation criterion and some policy priorities can be derived. In our example, by considering **B** as the top performer country, the pairwise comparisons supply the results contained in Figures from A6 to A10. The first two columns report results on the credibility of the statement that the overall evaluation considers an option better, equal or worse than another one. The third column provides relevant information for the benchmark exercise; in fact the credibility of the evaluation is referred to each single evaluation criterion. In this way, e.g. by looking at Figure A6, it is possible to deduce that **A** is performing better than **B** (the benchmark) on criteria 3 and 7, it is almost equal to **B** on criterion 5, thus we can conclude that the policy aspects represented by these criteria are dealt with in a satisfactory way. On the contrary, **A** is definitely worse than **B** on 4 criteria, which as a consequence should be considered as top policy priorities to improve the overall performance of **A** in the future.



**Figure A6** Pairwise Comparison Between A and B

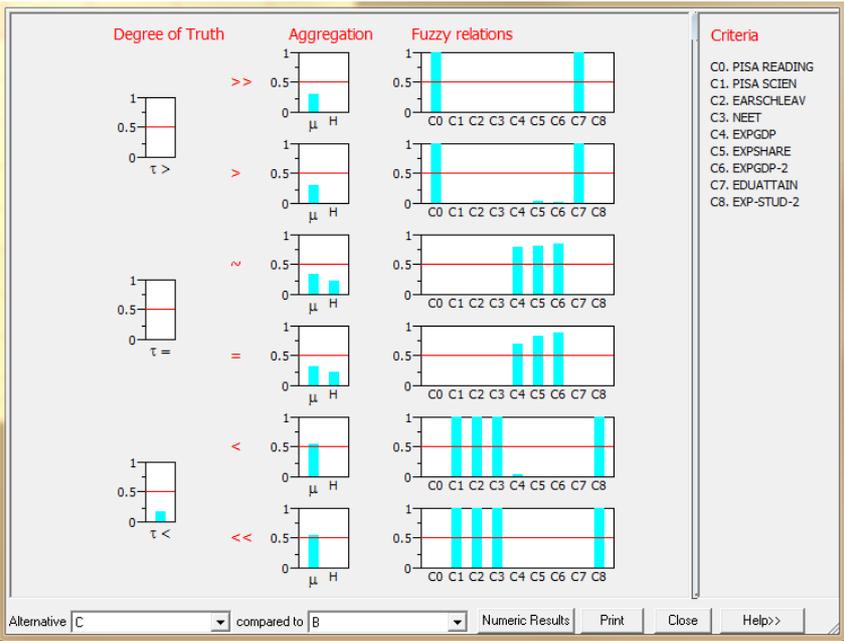


Figure A7 Pairwise Comparison Between C and B

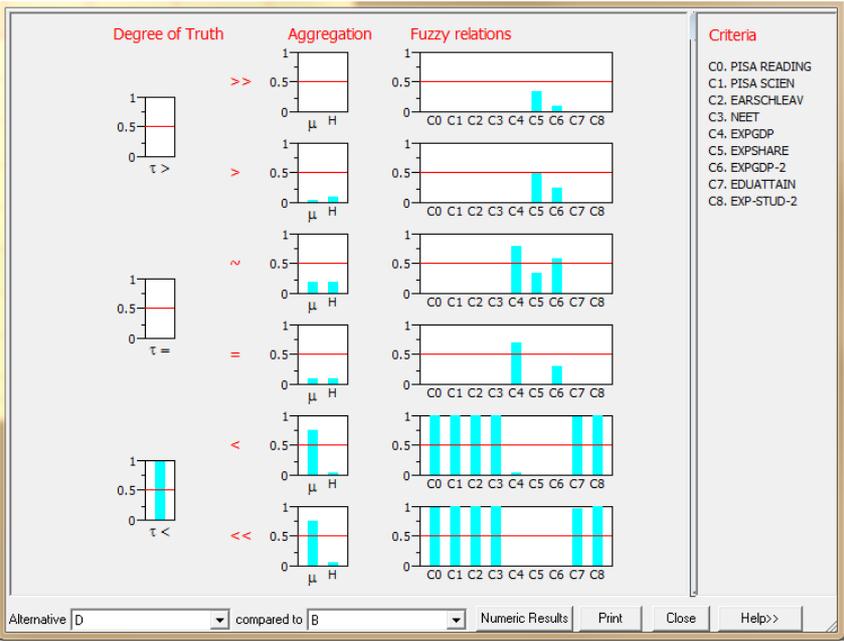


Figure A8 Pairwise Comparison Between D and B

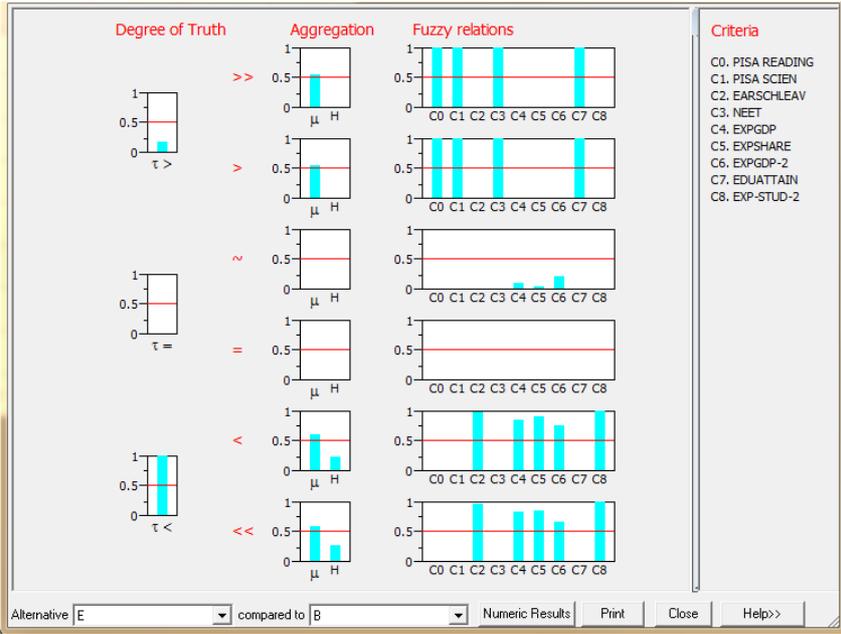


Figure A9 Pairwise Comparison Between E and B

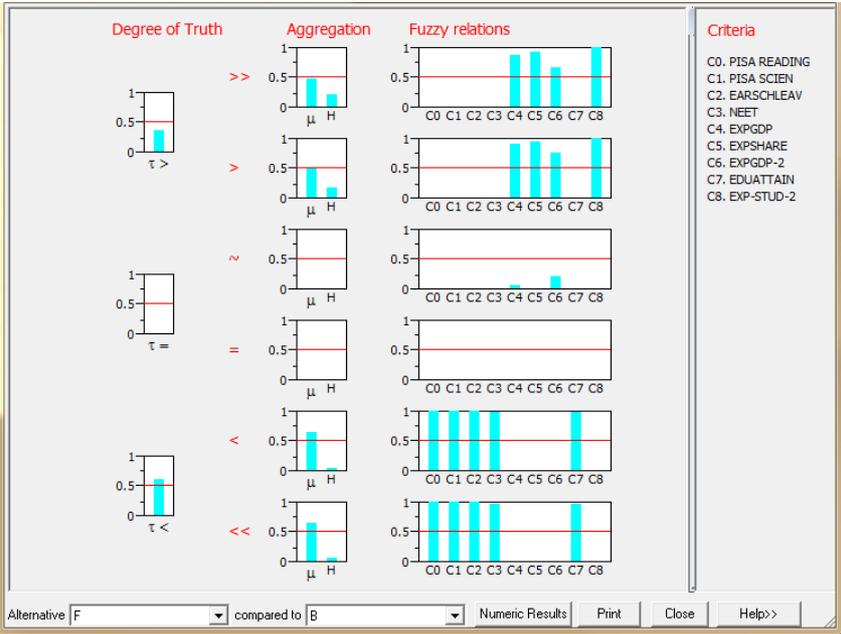


Figure A10 Pairwise Comparison Between F and B

## References

- Afonso, A., & Aubyn, M. S. (2006). Cross-country efficiency of secondary education provision: A semi-parametric analysis with non-discretionary inputs. *Economic Modelling*, 23(3), 476-491.
- Agasisti, T., & Salerno, C. (2007). Assessing the cost efficiency of Italian universities. *Education Economics*, 15(4), 455-471.
- Agasisti, T. (2009). Market forces and competition in university systems: theoretical reflections and empirical evidence from Italy. *International Review of Applied Economics*, 23(4), 463-483.
- Agasisti, T., & Johnes, G. (2009). Beyond frontiers: comparing the efficiency of higher education decision - making units across more than one country. *Education Economics*, 17(1), 59-79.
- Agasisti, T., & Pérez-Esparrells, C. (2010). Comparing efficiency in a cross-country perspective: the case of Italian and Spanish state universities. *Higher Education*, 59(1), 85-103.
- Agasisti, T., and G. Johnes (2010). Heterogeneity and the evaluation of efficiency: the case of Italian universities. *Applied Economics*, 42: 1365–75.
- 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., & Pohl, C. (2012). Comparing German and Italian public universities: Convergence or divergence in the higher education landscape?. *Managerial and Decision Economics*, 33(2), 71-85.
- Agasisti, T., Bonomi, F., & Sibiano, P. (2014). Measuring the “managerial” efficiency of public schools: a case study in Italy. *International Journal of Educational Management*, 28(2), 120-140.
- Agasisti, T., & Belfield, C. (2014). Efficiency in the community college sector: Stochastic frontier analysis. CCRC Working Papers, Teachers College Columbia University.
- Agasisti, T. (2014). The efficiency of public spending on education: An empirical comparison of EU countries. *European Journal of Education*, 49(4), 543-557.
- Agasisti, T., & Bonomi, F. (2014). Benchmarking universities' efficiency indicators in the presence of internal heterogeneity. *Studies in Higher Education*, 39(7), 1237-1255.
- Agasisti, T., & Haelermans, C. (2015). Comparing efficiency of public universities among European countries: Different incentives lead to different performances. *Higher Education Quarterly*.
- Agasisti, T., & Wolszczak-Derlacz, J. (2015). Exploring efficiency differentials between Italian and Polish universities, 2001–11. *Science and Public Policy*, scv026.
- Aghion, P., Dewatripont, M., Hoxby, C., Mas-Colell, A., & Sapir, A. (2010). The governance and performance of universities: evidence from Europe and the US. *Economic Policy*, 25(61), 7-59.

Aigner, D., Lovell, C. K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, 6(1), 21-37.

Alberta Oliveira, M., & Santos, C. (2005). Assessing school efficiency in Portugal using FDH and bootstrapping. *Applied Economics*, 37(8), 957-968.

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

Angilella, S., Corrente, S., Greco, S., Słowiński, R. (2015). Robust Ordinal Regression and Stochastic Multiobjective Acceptability Analysis in multiple criteria hierarchy process for the Choquet integral preference model. *Omega*. doi:10.1016/j.omega.2015.10.010

Antunes, C.H., Alves, M.J., Clímaco, J., *Multi-objective Linear and Integer Programming*, SPRINGER 2016.

Arrow, K.J., Raynaud, H. (1986) *Social choice and multicriterion decision making*. M.I.T. Press, Cambridge.

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

Barr, N. (2004). Higher education funding. *Oxford Review of Economic Policy*, 20(2), 264-283.

Barro, R. J. (2001). Human capital and growth. *American Economic Review*, 12-17.

Battese, G. E., & Coelli, T. J. (1995). A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics*, 20(2), 325-332.

Baumol, W.J., J.C. Panzar and R.D. Willig (1982). *Contestable Markets and the Theory of Industry Structure*, New York: Harcourt Brace Jovanovich.

Beasley, J. E. (1990). Comparing university departments. *Omega*, 18(2), 171-183.

Benhabib, J., & Spiegel, M. M. (1994). The role of human capital in economic development evidence from aggregate cross-country data. *Journal of Monetary Economics*, 34(2), 143-173.

Berlinski, S., Galiani, S., & Gertler, P. (2009). The effect of pre-primary education on primary school performance. *Journal of Public Economics*, 93(1), 219-234.

Bland, C. J., Center, B. A., Finstad, D. A., Risbey, K. R., & Staples, J. (2006). The impact of appointment type on the productivity and commitment of full-time faculty in research and doctoral institutions. *Journal of Higher Education*, 89-123.

Bloom, N., Lemos, R., Sadun, R., & Van Reenen, J. (2015). Does management matter in schools?. *The Economic Journal*, 125(584), 647-674.

Bolli, T., Olivares, M., Bonaccorsi, A., Daraio, C., Aracil, A. G., & Lepori, B. (2016). The Differential Effects of Competitive Funding on the Production Frontier and the Efficiency of Universities. *Economics of Education Review*.

Bonaccorsi, A., & Daraio, C. (Eds.). (2007a). *Universities and strategic knowledge creation: Specialization and performance in Europe*. Edward Elgar Publishing.

Bonaccorsi, A., & Daraio, C. (2007b). The differentiation of the strategic profile of higher education institutions. New positioning indicators based on microdata. *Scientometrics*, 74(1), 15-37.

Bonaccorsi, A., Daraio, C. & Simar L. (2007). Efficiency and productivity in European universities: exploring trade-offs in the strategic profile. Chapter in Bonaccorsi & Daraio (2007a).

Bonaccorsi, A. (Ed.). (2014). *Knowledge, Diversity and Performance in European Higher Education: A Changing Landscape*. Edward Elgar Publishing.

Bordons, M., Fernández, M., & Gómez, I. (2002). Advantages and limitations in the use of impact factor measures for the assessment of research performance. *Scientometrics*, 53(2), 195-206.

Bouyssou, D., Vansnick, J.C. (1986) Non-compensatory and generalized non-compensatory preference structures. *Theory and Decision*, 21, pp. 251-266.

Brans, J.P., Mareschal, B. and Vincke, Ph. (1986) How to select and how to rank projects. The PROMETHEE method. *European Journal of Operational Research*, vol.24, pp. 228-238.

Bratti, M., Checchi, D., & De Blasio, G. (2008). Does the expansion of higher education increase the equality of educational opportunities? Evidence from Italy. *Labour*, 22(s1), 53-88.

Campbell, J. P., & Oblinger, D. G. (2007). Academic analytics. *EDUCAUSE Review*, 42(4), 40-57.

Carvalho Chaves, M.C., Soares de Mello, J.C., Angulo-Meza, L. (2016), Studies of some duality properties in the Li and Reeves model, *Journal of the Operational Research Society*, vol.67, 474-482.

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 Operational Research*, 2, 429-444.

Charnes, A., Cooper, W. W., Lewin, A. Y., & Seiford, L. M. (Eds.). (2013). *Data envelopment analysis: Theory, methodology, and applications*. Springer Science & Business Media.

Chowdry, H., Crawford, C., Dearden, L., Goodman, A., & Vignoles, A. (2013). Widening participation in higher education: analysis using linked administrative data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(2), 431-457.

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

Clímaco, J.C.N., & Antunes, C. H. (1987). TRIMAP - an interactive tricriteria linear programming package. *Foundations of Control Engineering*, vol.12, 101-119.

Clímaco, J.C.N., & Antunes, C. H. (1989). Implementation of an user friendly software package - a guided tour of TRIMAP. *Mathematical and Computer Modelling*, vol.12, 1299-1309.

Clímaco, J. C. N., Soares de Mello, J. C. C. B. e Angulo- Meza, L. (2008), Performance Measurement – From DEA to MOLP, in Adam, F. and Humphreys, P. (Eds), Encyclopedia of Decision Making and Decision Support Technologies, Information Science Reference, Hershey, 709-715.

Clímaco, J., Soares de Mello, J.C., Angulo Meza, L.(2010). A study of highways performance with a MOLPDEA model and an interactive tri-criteria linear programming package (TRIMAP). Brazilian Journal of Operations and Production Management, vol. 7, n.1, pp. 163-179.

Coelli, T. J., Rao, D. S. P., O'Donnell, C. J., & Battese, G. E. (2005). An introduction to efficiency and productivity analysis. Springer Science & Business Media.

Cohn, E., Rhine, S. L., & Santos, M. C. (1989). Institutions of higher education as multi-product firms: Economies of scale and scope. The Review of Economics and Statistics, 284-290.

Cohn, E., & Cooper, S. T. (2004). 15 Multi-product cost functions for universities: economies of scale and scope. International handbook on the economics of education, 579.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. (1966). Equality of educational opportunity. Washington, DC, 1066-5684.

Cooper, W. W., Seiford, L. M., & Zhu, J. (Eds.). (2011). Handbook on data envelopment analysis (Vol. 164). Springer Science & Business Media.

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

Cordero Ferrera, J. M., Cebada, E. C., Chaparro, F. P., & González, D. S. (2011). Exploring educational efficiency divergences across Spanish regions in PISA 2006. Revista de Economía Aplicada, 19(57), 117-146.

Cordero, J. M., Santín, D., & Sicilia, G. (2015). Testing the accuracy of DEA estimates under endogeneity through a Monte Carlo simulation. European Journal of Operational Research, 244(2), 511-518.

Corrente, S., Greco, S., Kadziński, M. and Słowiński, R. (2013). Robust ordinal regression in preference learning and ranking. Machine Learning, 93(2-3), 381-422.

Cunha, M., & Rocha, V. (2012). On the efficiency of Public higher education institutions in Portugal: An exploratory study. FEP Economics and Management, Working Paper.

Daouia, A., & Simar, L. (2005). Robust nonparametric estimators of monotone boundaries. Journal of Multivariate Analysis, 96(2), 311-331.

Daraio, C., & Simar, L. (2005). Introducing environmental variables in nonparametric frontier models: a probabilistic approach. Journal of Productivity Analysis, 24(1), 93-121.

Daraio, C., & Simar, L. (2007). Advanced Robust and nonparametric methods in efficiency analysis: methodology and applications. Springer Science & Business Media.

Daraio, C., Bonaccorsi, A., Geuna, A., Lepori, B., Bach, L., Bogetoft, P., ... & Fried, H. (2011). The European university landscape: A micro characterization based on evidence from the Aquameth project. Research Policy, 40(1), 148-164.

Darling-Hammond, L. (2000). Teacher quality and student achievement. *Education Policy Analysis Archives*, 8, 1.

De la Torre, E., Agasisti T. & Perez-Esparrells C. (2016). How including third mission indicators do change universities' efficiency scores: empirical evidence on Spanish universities, Working Paper – to be presented at the 2016 Annual Conference of the Association for Education Finance and Policy (AEFP).

De Witte, K., & Moesen, W. (2010). Sizing the government. *Public Choice*, 145(1-2), 39-55.

De Witte, K., & López-Torres, L. (2015). Efficiency in education: a review of literature and a way forward. *Journal of the Operational Research Society*.

Di Liberto, A., Schivardi, F., & Sulis, G. (2015). Managerial Practices and Students' Performance. *Economic Policy*, 30(84), 683-728.

Dill, D. D. (1997). Higher education markets and public policy. *Higher Education Policy*, 10(3-4), 167-185.

Dill, D.D. & Soo, M. (2005). Academic quality, league tables, and public policy: A cross-national analysis of university ranking systems. *Higher Education*, Volume 49, Issue 4, pp 495–533.

Dronkers, J., & Robert, P. (2008). Differences in scholastic achievement of public, private government-dependent, and private independent schools: a cross-national analysis. *Educational Policy*, 22(4), 541-577.

Emrouznejad, A., Parker, B. R., & Tavares, G. (2008). Evaluation of research in efficiency and productivity: A survey and analysis of the first 30 years of scholarly literature in DEA. *Socio-Economic Planning Sciences*, 42(3), 151-157.

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

Estelle, S. M., Johnson, A. L., & Ruggiero, J. (2010). Three-stage DEA models for incorporating exogenous inputs. *Computers & Operations Research*, 37(6), 1087-1090.

Etzkowitz, H., & Leydesdorff, L. (2000). The dynamics of innovation: from National Systems and "Mode 2" to a Triple Helix of university–industry–government relations. *Research Policy*, 29(2), 109-123.

Eugène, B. (2008). The efficiency frontier as a method for gauging the performance of public expenditure: a Belgian case study. *National Bank of Belgium Working Paper*, (138).

European Commission (2015). *Education and training monitor 2015*. Publications Office of the European Union.

Fandel, G. (2007). On the performance of universities in North Rhine-Westphalia, Germany: Government's redistribution of funds judged using DEA efficiency measures. *European Journal of Operational Research*, 176(1), 521-533.

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General)*, 120(3), 253-290.

Figueira, J., Greco, S. and Ehrgott, M. (eds.) (2016) Multiple-criteria decision analysis. State of the art surveys. Springer International Series in Operations Research and Management Science, New York.

Figlio, D. N. (1999). Functional form and the estimated effects of school resources. *Economics of Education Review*, 18(2), 241-252.

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

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

Glass, J. C., McKillop, D. G., & Hyndman, N. (1995). Efficiency in the provision of university teaching and research: an empirical analysis of UK universities. *Journal of Applied Econometrics*, 10(1), 61-72.

Golany, B., & Tamir, E. (1995). Evaluating efficiency-effectiveness-equality trade-offs: a data envelopment analysis approach. *Management Science*, 41(7), 1172-1184.

Grabisch, M. (1996). The application of fuzzy integrals in multicriteria decision making. *European Journal of Operational Research*, 89(3), 445-456.

Greco, S., Matarazzo, B., and Slowinski, R. (2001). Rough sets theory for multicriteria decision analysis. *European Journal of Operational Research*, 129(1), 1-47.

Greene, W. (2005). Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *Journal of Econometrics*, 126(2), 269-303.

Greene, W. H. (2008). The econometric approach to efficiency analysis. In *The measurement of productive efficiency and productivity growth*, 92-250.

Greenwald, R., Hedges, L. V., & Laine, R. D. (1996). The effect of school resources on student achievement. *Review of Educational Research*, 66(3), 361-396.

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

Grosskopf, S., & Moutray, C. (2001). Evaluating performance in Chicago public high schools in the wake of decentralization. *Economics of Education Review*, 20(1), 1-14.

Guitouni, A., Martel, J.M. (1998) Tentative guidelines to help choosing an appropriate MCDA method. *European Journal of Operational Research*, 109, pp. 501-521.

Gulbrandsen, M., & Slipersaeter, S. (2007). The third mission and the entrepreneurial university model. In *Universities and strategic knowledge creation*, 112-143.

Gundlach, E., Wossmann, L., & Gmelin, J. (2001). The decline of schooling productivity in OECD countries. *The Economic Journal*, 111(471), 135-147.

Guskey T.R. (2007). Multiple Sources of Evidence: An Analysis of Stakeholders' Perceptions of Various Indicators of Student Learning. *Educational Measurement: Issues and Practice*, Volume 26, Issue 1, Pages 19–27.

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

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

Hanushek, E. A. (1979). Conceptual and empirical issues in the estimation of educational production functions. *Journal of Human Resources*, 351-388.

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

Hanushek, E. A. (1997). Assessing the effects of school resources on student performance: An update. *Educational evaluation and policy analysis*, 19(2), 141-164.

Hanushek, E. A., & Luque, J. A. (2003). Efficiency and equity in schools around the world. *Economics of Education Review*, 22(5), 481-502.

Hanushek, E. A. (2006). School resources. *Handbook of the Economics of Education*, 2, 865-908.

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

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

Hanushek, E. A., & Woessmann, L. (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *Journal of Economic Growth*, 17(4), 267-321.

Hashimoto, K., & Cohn, E. (1997). Economies of scale and scope in Japanese private universities. *Education Economics*, 5(2), 107-115.

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

Hicks, D. (2012). Performance-based university research funding systems. *Research Policy*, 41(2), 251-261.

Hinloopen, E., Nijkamp, P., Rietveld, P. (1983) Qualitative discrete multiple criteria choice models in regional planning. *Regional Science and Urban Economics*, pp. 77-102.

Hinloopen, E., Nijkamp, P. (1990) Qualitative multiple criteria choice analysis, the dominant regime method. *Quality and Quantity* 24, pp. 37-56.

Ho William, Prasanta K. Dey, Helen E. Higson, (2006). "Multiple criteria decision-making techniques in higher education", *International Journal of Educational Management*, Vol. 20 Iss: 5, pp.319 - 337

Izadi, H., Johnes, G., Oskrochi, R., & Crouchley, R. (2002). Stochastic frontier estimation of a CES cost function: the case of higher education in Britain. *Economics of Education Review*, 21(1), 63-71.

James, E. (1978). Product mix and cost disaggregation: A reinterpretation of the economics of higher education. *Journal of Human Resources*, 157-186.

Johnes, G., & Johnes, J. (1993). Measuring the research performance of UK economics departments: an application of data envelopment analysis. *Oxford economic papers*, 332-347.

Johnes, J., & Johnes, G. (1995). Research funding and performance in UK university departments of economics: a frontier analysis. *Economics of Education Review*, 14(3), 301-314.

Johnes, G. (1997). Costs and industrial structure in contemporary British higher education. *The Economic Journal*, 107(442), 727-737.

Johnes, J. (2004). Efficiency measurement. *International Handbook on the Economics of Education*, chapter 16.

Johnes, J., Thanassoulis, E., Lenton, P., & Emrouznejad, A. (2005). An exploratory analysis of the cost structure of higher education in England. DfES publications.

Johnes, G., Johnes, J., & Thanassoulis, E. (2008). An analysis of costs in institutions of higher education in England. *Studies in Higher Education*, 33(5), 527-549.

Johnes, G., and J. Johnes (2009). Higher education institutions' costs and efficiency: taking the decomposition a further step. *Economics of Education Review*, 28: 107–13.

Johnes, G., and A. Schwarzenberger (2011). Differences in cost structure and the evaluation of efficiency: the case of German universities. *Education Economics*, 19: 487–500.

Johnes, G. (2015). Evaluating the efficiency of public services. IZA World of Labor.

Johnson, A. L., & Ruggiero, J. (2014). Nonparametric measurement of productivity and efficiency in education. *Annals of Operations Research*, 221(1), 197-210.

Johnstone, D. B. (2004). The economics and politics of cost sharing in higher education: comparative perspectives. *Economics of Education Review*, 23(4), 403-410.

Jongbloed, B., & Vossensteyn, H. (2001). Keeping up performances: An international survey of performance-based funding in higher education. *Journal of Higher Education Policy and Management*, 23(2), 127-145.

Joumady, O., & Ris, C. (2005). Performance in European higher education: A non-parametric production frontier approach. *Education Economics*, 13(2), 189-205.

Kao, C., & Hung, H. T. (2008). Efficiency analysis of university departments: An empirical study. *Omega*, 36(4), 653-664.

Keeney, R., Raiffa, H. (1976) *Decisions with multiple objectives: preferences and value trade-offs*. Wiley, New York.

Kempkes, G., & Pohl, C. (2010). The efficiency of German universities—some evidence from nonparametric and parametric methods. *Applied Economics*, 42(16), 2063-2079.

Köksal, G., & Nalçacı, B. (2006). The relative efficiency of departments at a Turkish engineering college: A data envelopment analysis. *Higher Education*, 51(2), 173-189.

Koshal, R. K., & Koshal, M. (1995). Quality and economies of scale in higher education. *Applied Economics*, 27(8), 773-778.

Koshal, R. K., & Koshal, M. (1999). Economies of scale and scope in higher education: a case of comprehensive universities. *Economics of Education Review*, 18(2), 269-277.

Koshal, R. K., & Koshal, M. (2000). State Appropriation and Higher Education Tuition: what is the relationship?. *Education Economics*, 8(1), 81-89.

Krueger, A. B. (2003). Economic considerations and class size. *The Economic Journal*, 113(485), F34-F63.

Kuo, J. S., & Ho, Y. C. (2008). The cost efficiency impact of the university operation fund on public universities in Taiwan. *Economics of Education Review*, 27(5), 603-612.

Laband, D. N., & Lentz, B. F. (2003). New estimates of economies of scale and scope in higher education. *Southern Economic Journal*, 172-183.

Ladd, H. F. (Ed.). (1996). *Holding schools accountable: Performance-based reform in education*. Brookings Institution Press.

Laredo, P. (2007). Revisiting the third mission of universities: toward a renewed categorization of university activities?. *Higher Education Policy*, 20(4), 441-456.

Levin, H. M., & McEwan, P. J. (2001). *Cost-effectiveness analysis: Methods and applications (Vol. 4)*. Sage.

Li, X.-B., & Reeves, G. R., (1999). A multiple criteria approach to data envelopment analysis. *European Journal of Operational Research*, vol.115(3), 507-517.

Luedemann, E., Schuetz, G., Woessmann, L., & West, M. R. (2009). *School accountability, autonomy and choice around the world*. Edward Elgar.

Malen B. & Michael Knapp (1997). Rethinking the multiple perspectives approach to education policy analysis: implications for policy-practice connections, *Journal of Education Policy*, 12:5, 419-445.

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

Martel, J.M., Zaras, K. (1995) Stochastic dominance in multicriteria analysis under risk. *Theory and Decision*, 35 (1), pp. 31-49.

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

Mensah, Y. M., & Werner, R. (2003). Cost efficiency and financial flexibility in institutions of higher education. *Journal of Accounting and Public Policy*, 22(4), 293-323.

Moed, H. F., Burger, W. J. M., Frankfort, J. G., & Van Raan, A. F. (1985). The use of bibliometric data for the measurement of university research performance. *Research Policy*, 14(3), 131-149.

Molas-Gallart, J., & Castro-Martínez, E. (2007). Ambiguity and conflict in the development of 'Third Mission' indicators. *Research Evaluation*, 16(4), 321-330.

Moore, M. H. (1995). *Creating public value: Strategic management in government*. Harvard University Press.

Munda, G. (1995) *Multicriteria evaluation in a fuzzy environment*. Physica-Verlag, Contributions to Economics Series, Heidelberg.

Munda G. (2008) *Social multi-criteria evaluation for a sustainable economy*, Springer, Heidelberg, New York.

Muthén, B. O. (1991). Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement*, 28(4), 338-354.

Nikel J., Lowe J. (2010). Talking of fabric: a multi-dimensional model of quality in education, *Compare: A Journal of Comparative and International Education*, Volume 40, Issue 5.

Nigsch, S., & Schenker-Wicki, A. (2015). Frontier efficiency analysis in higher education. In *Incentives and Performance* (pp. 155-170). Springer International Publishing.

Nijkamp P., Voogd M. (1985) - An informal introduction to multicriteria evaluation, in Fandel G., Matarazzo B., Spronk J. (eds.) - *Multiple criteria decision methods and applications*, Springer-Verlag, Berlin-Heidelberg, pp. 27-40.

Nijkamp, P., Rietveld, P. and Voogd, H. (1990) *Multicriteria Evaluation in Physical Planning*. Amsterdam: North-Holland.

OECD (2002). *Education at a Glance 2002 - Glossary*, OECD, Paris.

OECD (2014). *PISA 2012 Results in focus. What 15-year-olds know and what they can do with what they know*, OECD Paris.

Parteka, A., & Wolszczak-Derlacz, J. (2013). Dynamics of productivity in higher education: cross-European evidence based on bootstrapped Malmquist indices. *Journal of Productivity Analysis*, 40(1), 67-82.

Perelman, S., & Santin, D. (2011). Measuring educational efficiency at student level with parametric stochastic distance functions: an application to Spanish PISA results. *Education Economics*, 19(1), 29-49.

Perry, L., & McConney, A. (2010). Does the SES of the school matter? An examination of socioeconomic status and student achievement using PISA 2003. *The Teachers College Record*, 112(4), 7-8.

Pinheiro, R., Langa, P. V., & Pausits, A. (2015). The institutionalization of universities' third mission: introduction to the special issue. *European Journal of Higher Education*, 5(3), 227-232.

Powell, B. A., Gilleland, D. S., & Pearson, L. C. (2012). Expenditures, efficiency, and effectiveness in US undergraduate higher education: A national benchmark model. *The Journal of Higher Education*, 83(1), 102-127.

Raudenbush, S. W., & Willms, J. D. (Eds.). (1991). *Schools, classrooms, and pupils: International studies of schooling from a multilevel perspective*. Academic Press.

Ray, S. C. (1991). Resource-use efficiency in public schools: a study of Connecticut data. *Management Science*, 37(12), 1620-1628.

Roberts, F. S. (1979) *Measurement theory with applications to decision making, utility and the social sciences*. Addison-Wesley, London.

Robst, J. (2001). Cost efficiency in public higher education institutions. *Journal of Higher Education*, 730-750.

Romer, P. M. (1990). Human capital and growth: theory and evidence. In *Carnegie-Rochester Conference Series on Public Policy* (Vol. 32, pp. 251-286). North-Holland.

Rossell C.H. (1993). Using Multiple Criteria To Evaluate Public Policies The Case of School Desegregation, *American Politics Research*, Vol 21, Issue 2.

Roy, B., Bertier, P. (1973) La methode ELECTRE II. Une application ou media planning, in M.Ross (ed.). *Operational research '72*, North Holland, Amsterdam, pp. 291-302.

Roy, B. (1978) ELECTRE III: Un algorithme de classement fondé sur une représentation floue des préférences en présence de critères multiples. *Cahiers du Centre d' Etudes de Recherche Opérationnelle Paris*, 20 (1), pp. 3-24.

Roy B. (1996) *Multicriteria methodology for decision analysis*, Kluwer, Dordrecht.

Ruggiero, J. (1998). Non-discretionary inputs in data envelopment analysis. *European Journal of Operational Research*, 111(3), 461-469.

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

Ruggiero, J. (2004). Performance evaluation in education. In *Handbook on data envelopment analysis* (pp. 323-346). Springer US.

Ruggiero, J. (2004b). Performance evaluation when non-discretionary factors correlate with technical efficiency. *European Journal of Operational Research*, 159(1), 250-257.

Ryan, J. F. (2004). The relationship between institutional expenditures and degree attainment at baccalaureate colleges. *Research in Higher Education*, 45(2), 97-114.

Saaty, T.L. (1980) *The analytic hierarchy process*. McGraw Hill, New York.

Sacerdote, B. (2011). Peer effects in education: How might they work, how big are they and how much do we know thus far?. *Handbook of the Economics of Education*, 3, 249-277.

Sammons, P. (1995). *Key characteristics of effective schools: A review of school effectiveness research*. B & MBC Distribution Services.

Sarrico, C. S., & Dyson, R. G. (2000). Using DEA for planning in UK universities—an institutional perspective. *Journal of the operational Research Society*, 51(7), 789-800.

Sarrico, C. S., Teixeira, P. N., Rosa, M. J., & Cardoso, M. F. (2009). Subject mix and productivity in Portuguese universities. *European Journal of Operational Research*, 197(1), 287-295.

Sarrico, C. S., Rosa, M. J., & Coelho, I. P. (2010). The performance of Portuguese secondary schools: An exploratory study. *Quality Assurance in Education*, 18(4), 286-303.

SCRGSP (Steering Committee for the Review of Government Service Provision) 2006, *Report on Government Services 2006*, Productivity Commission, Canberra.

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

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

Smith, P. C., & Street, A. (2005). Measuring the efficiency of public services: the limits of analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168(2), 401-417.

Soares de Mello, J.C.C.B., Clímaco, J.C.N., & Angulo-Meza, L. (2009). Efficiency evaluation of a small number of DMUs: an approach based on Li and Reeves's model. *Pesquisa Operacional*, vol. 29, 97-110.

St. Aubyn, M. S., Garcia, F., & Pais, J. (2009). Study on the efficiency and effectiveness of public spending on tertiary education (No. 390). Directorate General Economic and Monetary Affairs (DG ECFIN), European Commission.

Steuer, R. (1986). *Multiple criteria optimization; theory, computation, and application*, Wiley Series in Probability and Mathematical Statistics - Applied, Wiley.

Stevens, P. A. (2005). A stochastic frontier analysis of English and Welsh universities. *Education Economics*, 13(4), 355-374.

Stufflebeam D. (2001). Evaluation models. *New Directions for Evaluation*, Issue 89, pp. 7-98.

Sutherland, D., Price, R., & Gonand, F. (2010). Improving public spending efficiency in primary and secondary education. *OECD Journal: Economic Studies*, 2009(1), 1-30.

Tajnikar, M., & Debevec, J. (2008). Funding system of full - time higher education and technical efficiency: case of the University of Ljubljana. *Education Economics*, 16(3), 289-303.

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

Thanassoulis, E., Kortelainen, M., Johnes, G., & Johnes, J. (2011). Costs and efficiency of higher education institutions in England: a DEA analysis. *Journal of the Operational Research Society*, 62(7), 1282-1297.

Todd, P. E., & Wolpin, K. I. (2007). The production of cognitive achievement in children: Home, school, and racial test score gaps. *Journal of Human capital*, 1(1), 91-136.

Tone, K. (2004). Malmquist productivity index. In *Handbook on Data Envelopment Analysis* (pp. 203-227). Springer US.

Tsionas, E. G. (2002). Stochastic frontier models with random coefficients. *Journal of Applied Econometrics*, 17(2), 127-147.

Tsionas, E. G., & Kumbhakar, S. C. (2014). Firm heterogeneity, persistent and transient technical inefficiency: a generalized true random - effects model. *Journal of Applied Econometrics*, 29(1), 110-132.

Tzeng G. H., Cheng-Hsin Chiang, Chung-Wei Li (2007). Evaluating intertwined effects in e-learning programs: A novel hybrid MCDM model based on factor analysis and DEMATEL. *Expert Systems with Applications* 32, 1028–1044.

Van Raan, A. F. (2005). Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics*, 62(1), 133-143.

Veiderpass, A., & McKelvey, M. (2015). Evaluating the performance of higher education institutions in Europe: a nonparametric efficiency analysis of 944 institutions. *Applied Economics*, 1-11.

Voogd, H. (1983) *Multicriteria evaluation for urban and regional planning*, London, Pion.

Webber, D. A., & Ehrenberg, R. G. (2010). Do expenditures other than instructional expenditures affect graduation and persistence rates in American higher education?. *Economics of Education Review*, 29(6), 947-958.

Wößmann, L., & Schütz, G. (2006). Efficiency and equity in European education and training systems. Analytical report for the European commission prepared by the European expert network on economics of education (EENEE) to accompany the communication and staff working paper by the European commission under the same title, 26.4, 1, 2007-2013.

Woessmann, L. (2007). International evidence on school competition, autonomy, and accountability: A review. *Peabody Journal of Education*, 82(2-3), 473-497.

Woessmann, L. (2008). Efficiency and equity of European education and training policies. *International Tax and Public Finance*, 15(2), 199-230.

Wolszczak-Derlacz, J., & Parteka, A. (2011). Efficiency of European public higher education institutions: a two-stage multicountry approach. *Scientometrics*, 89(3), 887-917.

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

Worthington, A. C., & Dollery, B. E. (2002). Incorporating contextual information in public sector efficiency analyses: a comparative study of NSW local government. *Applied Economics*, 34(4), 453-464.

Worthington, A. C., & Higgs, H. (2011). Economies of scale and scope in Australian higher education. *Higher Education*, 61(4), 387-414.

Zhu, J. (Ed.). (2015). *Data Envelopment Analysis: A Handbook of Models and Methods* (Vol. 221). Springer.

***Europe Direct is a service to help you find answers  
to your questions about the European Union.***

**Freephone number (\*):**

**00 800 6 7 8 9 10 11**

(\*). The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

More information on the European Union is available on the internet (<http://europa.eu>).

## **HOW TO OBTAIN EU PUBLICATIONS**

### **Free publications:**

- one copy:  
via EU Bookshop (<http://bookshop.europa.eu>);
- more than one copy or posters/maps:  
from the European Union's representations ([http://ec.europa.eu/represent\\_en.htm](http://ec.europa.eu/represent_en.htm));  
from the delegations in non-EU countries ([http://eeas.europa.eu/delegations/index\\_en.htm](http://eeas.europa.eu/delegations/index_en.htm));  
by contacting the Europe Direct service ([http://europa.eu/europedirect/index\\_en.htm](http://europa.eu/europedirect/index_en.htm)) or  
calling 00 800 6 7 8 9 10 11 (freephone number from anywhere in the EU) (\*).

(\*). The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

### **Priced publications:**

- via EU Bookshop (<http://bookshop.europa.eu>).

## JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.



**EU Science Hub**  
[ec.europa.eu/jrc](https://ec.europa.eu/jrc)



@EU\_ScienceHub



EU Science Hub - Joint Research Centre



Joint Research Centre



EU Science Hub

