Benchmarking in Public Services: The cases of health and education in Portugal

Maria Conceição A. Silva Portela
CEGE-FEG- UCP

Abstract

This paper addresses the measurement of efficiency and benchmarking of public services in Portugal, with particular emphasis on education and health – two services with great benchmarking tradition and with several publicly available tools. The paper intends to present the benchmarking tools, BESP and HOBE, developed by the author, and to show how such tools can be used for improved efficiency. Aggregate efficiency indicators are computed based on Data Envelopment Analysis, and therefore this methodology is also presented by means of an illustrative example.

1. Introduction

In a period of economic turbulence, fierce competition, and demanding customers, the pressures to optimize processes, to improve efficiency and effectiveness, and to reduce costs are constant in every company that wishes to survive and prosper. This has been typically true in for-profit companies, but in the last years non-profit and public sectors have been equally exposed to stringent economic constraints, and only process improvement and efficiency gains can assure cost cuts without quality deterioration. One way to obtain such gains is through benchmarking. Benchmarking can be defined as a continuous process of comparison of performance between different organizations with the purpose of sharing best practices and improve performance (Beckford 1998).

Since the mid 90's public spending in Portugal has been increasing until 2011 (where it suffered a reverse). The health sector was responsible for 23% of total spending in 2013, the education sector was responsible for 19% of total spending in 2013, and the security and public order sector was responsible for 8% of total spending. As can be seen in the graph below (from Pordata) the largest share of total public spending is social security with a share of 37% of total spending. As a result, analysing the efficiency through which education, health and justice are provided, assumes particular relevance not only because these three sectors together account for more than 50% of total spending, but also because the provision of these services is of vital importance for the functioning of any developed economy.

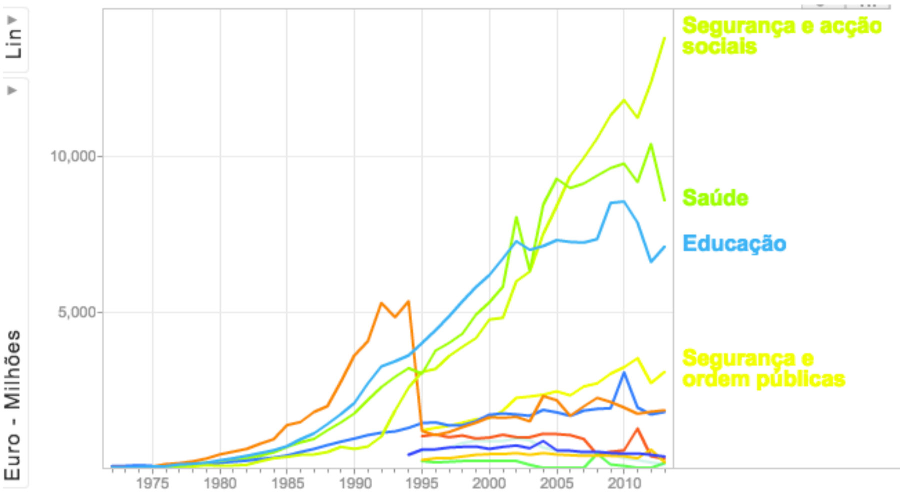


Figure 1: Evolution of public spending by function in Portugal

In this paper we address some efforts that have been made in Portugal for benchmarking public services - particularly in the areas of education and health - two public services with a great tradition of benchmarking and comparative analyses in Portugal and the rest of the world (e.g. in Portugal the Health Regulatory Authority provides a public web platform - SINAS - which allows the comparison of hospitals on certain quality indicators, and in education school rankings, published since 2001 by the media, allow the benchmarking of schools which, despite several criticism, has had great impact on school organization and management).

Benchmarking is typically performed on the basis of Key Performance Indicators (KPIs). The use of individual indicators presents some problems, since in general such analyses assume the independence between KPIs, which may not be a valid assumption. To overcome this problem, it is common the use of aggregate performance indicators or Composite Indicators as defined by OECD (OECD, 2008a). There are several techniques that can be used to construct aggregate indicators, but we will explore in this paper Data Envelopment Analysis (DEA). This technique was developed in 1978 by Charnes, Cooper and Rhodes based on a work of Farrell (1957), and has been applied in various contexts such as schools, banks, retail outlets, post offices, water companies, health centres, hospitals, etc.
The basic principles of DEA will be presented in this paper through graphical analysis and an illustrative example. Real cases will be analysed, with an illustration of the results and their potential impact on management. Focus will be given to the cases of education and health, where the aggregate performance indicators are available within public benchmarking platforms like BESP (Portela et al. 2011) and HOBE (Portela et al. 2014), both developed by the author.

2. Web based benchmarking tools

a. Health benchmarking tools

Benchmarking through the web is not a new concept, and has been put into practice by several entities. In the health sector web platforms have been mainly developed by private entities that provide consultancy in the area of benchmarking, amongst other. There is a growing number of consultancy firms providing benchmarking services to hospitals and other health institutions. Some of these companies also provide a public service by publishing hospital rankings (e.g. U.S. News Best Hospitals, or the Truven Health Analytics 100 Top Hospitals in the US), by providing hospital guides (e.g. Dr. Foster Hospital Guide), or by attributing prizes to performance (e.g. *Caspe Healthcare Knowledge System* awards). In spite of providing some outcomes of the benchmarking exercises to the public, the actual access to the benchmarking platforms of the above consultancy firms is only opened to the health institutions that purchased their services.

Some examples of publicly available benchmarking health tools are the Acute Trust Quality dashboard in the UK, or the Care quality commission in the UK, the former is used to benchmark Trusts, and the later is used for benchmarking hospitals. In Portugal benchmarking efforts have been done by public entities, in spite of the fact that some hospitals use consultancy services for benchmarking purposes. The Health Regulatory Entity (ERS), has developed, with a Joint Commission – Siemens consortium, a web tool named SINAS (ERS, 2011). This tool is intended to provide a rating system (based on 'stars') of hospitals. SINAS is based on hospital indicators on five dimensions (clinical excellence, safety of the patient, comfort of facilities, focus on the patient, and patient satisfaction). Clinical excellence indicators are provided for several specialties (cardiology, general surgery, genecology, etc.) and constructed at the detailed level of the disease, taking into account the procedures required for a 'good quality treatment'. A hospital is considered of clinical excellence III if it is above the 99% upper limit of the confidence interval for the entire sample of hospitals. This tool is interesting in the sense that both public and private hospitals report information to the SINAS. However, in practice for many hospitals and indicators, no evaluation results are indeed available, meaning that the usefulness of the tool is very limited.
Following a management perspective, as opposed to the clinical perspective of SINAS, ACSS (Administração Central do Sistema de Saúde) launched in 2014 a benchmarking web site for hospitals (http://benchmarking.acss.min-saude.pt/), where indicators on 4 dimensions (access, quality, productivity and financial-economics) are monitored on a monthly basis. Hospitals are grouped into 5 clusters and comparisons are made within each cluster. Note that the level of analysis in this web site is the hospital, whereas in SINAS it is the disease, or the patient.

The above shows that the evaluation paradigm is changing in Portugal, and several recent and valid attempts towards benchmarking hospitals have been undertaken. However, there is still much to be done on this subject as most hospitals do not seem to understand the advantages of benchmarking. A culture of accountability, at all levels in the Portuguese society, needs to be developed such that entities being evaluated are not reluctant in providing their data, are careful with the data collection process, are happy with the increased transparency that benchmarking exercises imply, and are not concerned with the increased empowerment of patients that benchmarking exercises bring.

HOBE is a benchmarking platform available at http://feg.porto.ucp.pt/hobe that has been developed within a project financed by the Portuguese foundation for science and technology (FCT). It is at the moment an experimental website and is not being used by hospitals. It has been designed to perform benchmarking at the service level on the basis of indicators relating to the volume of the activity of hospitals and to its costs. It is therefore based on a managerial perspective in the same spirit of the ACSS benchmarking website. Its main advantages are the level of detail (service level rather than hospital level), and the possibility of customization – hospitals can choose their universe of comparison rather than being restricted to benchmarking within a pre-defined cluster. In addition, this website allows the construction of aggregate indicators at the service and at the hospital level. For each service a set of indicators are produced, but these may entail contradictory information and some implicit trade-offs. Therefore it is important to aggregate all the indicators produced for a service into a single indicator that gives information on the overall performance of the service. This is done in the website through a DEA based procedure, that produces an aggregate indicator of performance for each service and allows the service to be compared with its peers. In addition, the HOBE platform also produces an aggregate performance measure for hospitals, which is an aggregate of the individual performance of each of its services. Such an indicator is also based on DEA.

In addition to the web platforms and consulting companies performing hospital benchmarking, there is a vast set of academic studies that perform this type of analysis. On an overview of such studies one can distinguish several levels of analysis. In some cases countries are compared (Afonso and St. Aubyn, 2005), in others the comparison is between hospitals (Moreira, 2008), or health centres (Thanassoulis et al. 2012). Hollingsworth (2008) advocates that hospitals are in general too complex and too heterogeneous to be compared amongst them. He advocates benchmarking at the service or disease level as a better and more informative way of performing benchmarking in the health context. Therefore several studies can be found in the literature using sub-levels of analysis, like intensive care units (Puig-Junoy, 1998 and Derveaux et al. 2009), or haemodialysis units (Kontodimopoulos and Niakas, 2005)). Departments or specialties have also been compared in (Laudicella et al., 2010), and doctors have also been benchmarked in Chilingerian (1995) and in Castro et al. (2013). Groups of diseases based on the DRG have been analysed in Dismuke and Sena (2001).

b. Education benchmarking tools

The education field is also rich in benchmarking tools. The idea of comparing school's performance dates back to the 60's when in the US the Coleman report (in 1966) concludes that schools do not matter, and results of students are mainly determined by their socio-economic background. This has led to a movement of 'school effectiveness research' in the 90's (where it was advocated that schools can make a difference), and the subsequent production of comparative studies between schools. In 1992 league tables were introduced in the UK and in 1993 the Tenessee Value-Added Assessment System (TVAAS) was introduced in Tennessee, US. These tools are still available today, though with several improvements most of which relating to the computation of contextualised value added of schools (seen as the right way to benchmark them). In the UK comparisons between schools can be made through the department for education web site where a set of school indicators are public and comparisons between schools' performance can be undertaken: http://www.education.gov.uk/schools/performance/. In the US the country dimension prevents comparisons of schools across the country, being these comparisons done within states. The TVAAS system is the prevailing system for evaluating schools and teachers in the US (note that this system in now called EVAAS - Education Value Added Assessment System, and has been operationalized by private companies, the most recent of which is SAS, which developed the SAS EVAAS Web reporting software). For details see (OECD, 2008b).

More importantly than national benchmarking of schools is the possibility of international school's benchmarking, made possible through the development of the International PISA program since 2000. Nowadays schools can be benchmarked, not only against other schools in the country, but also against other schools, and school systems in the world. PISA data have give rise to a high number of academic studies that attempt to use these data to guide national education policies. Most notably OECD has produced very comprehensive reports with student results and benchmarking exercises across schools internationally. All results are publically available at http://www.oecd.org/pisa/, where there is also the possibility of exploring the data, and see a set of indicators for a number of chosen countries using the *Education GPS,* available within the PISA website.

In Portugal there is a free web tool that allows the benchmarking of Portuguese secondary schools named BESP (details can be found in Portela et al., 2011). BESP (available at http://feg.porto.ucp.pt/besp) is a visual tool showing for each school its performance on individual indicators (based on the percentile notion,) when the school is compared with the set of all schools in the country, or any other comparative set (like all schools of the same type, all schools in the same district, etc.). The platform has a public and a private area. The public area shows indicators based on data from national exam results databases (publicly available). The private area allows schools to provide additional information to the platform and allows benchmarking on indicators constructed based on these data. In the private area there is a customized tool that allows schools to choose the set of indicators that they want to aggregate to construct a composite performance indicator. This tool is based on DEA, and therefore schools can not only know what their aggregate performance score is, but also who are the benchmark schools based on which this score was obtained.

In addition to web platforms, there is a vast set of academic studies that perform educational benchmarking. There are mainly two type of analysis that can be distinguished in the literature: Analysis on ***School effectiveness,*** where pupil-level data is used to assess the value that a school adds to its students, given their abilities and intrinsic capabilities at the entry of a certain educational stage. This value added perspective implies matching the results for the same student in two or more

different periods, leading to richer outcomes, since within school variation (differential school effectiveness), teacher effects, and between school variation (school effects) can be assessed.
The second level is that of ***School efficiency*** studies, where aggregate school data are used. In this type of studies frontier models (stochastic frontiers and Data envelopment analysis) have been traditionally employed. Such methods see schools as a production process consuming a set of inputs to transform into a set of outputs (outcomes), and identify best performance based on the notion of production frontier.
When pupil level data is used multilevel regression models have been advocated (see OECD 2006 and 2008b). Multilevel models can be seen as the construction of a regression model per school, where differences between the intercepts are seen as the school effect. Some examples of multilevel studies applying pupil-level data can be seen in Ladd and Walsh (2002); Hanushek and Taylor (1990); Gray and Jesson (1996); O'Donoghue et al. (1997). DEA has also been applied to pupil-level data by Portela and Thanassoulis (2001), DeWitte et al (2010), and Portela and Camanho (2010). Nevertheless, the use of DEA has been more prevalent in school efficiency studies where the school is the level of analysis rather than the pupil (see e.g. Mancebón and Bandrés (1999), Kirjavainen and Loikkanen (1998) or Grosskopf et al. (1999)).

c. Other benchmarking tools

In other fields we observe that benchmarking platforms are designed with tools that allow benchmarking on individual indicators, but also on aggregate indicators. DEA has been an extensively used tool to aggregate KPIs into a single performance indicator. Bogetoft and Nielsen (2005) were amongst the first to present some solutions that use DEA and other frontier techniques for benchmarking. These authors advocate that, in order to make better use of benchmarking techniques, their results must be available in web tools. As a result they developed a web tool (with a private scope) to analyze Danish commercial banks, using a DEA model. Recent work by the same authors has resulted in an improved (private) web tool and the foundation of Ibensoft ApS – a company wich develops tailored web platforms for interactive benchmarking for individual companies (Ibensoft ApS, 2013). This tool commercially provides customers with state of the art frontier and econometric models for every industry or sector. The tool dynamically enables travelling along the best practice frontier, choosing the benchmarking direction and also choosing the percentiles with which the client wants to compare his/her organization.
Another example of a web benchmarking tool is the platform iDEAs, from the School of Industrial and Systems Engineering at the Georgia Institute of Technology, which explicitly uses DEA as a tool to compare warehouses. Companies that want to compare with their peers only need to supply their data online, to obtain a measure of relative efficiency, computed by a predefined model (For further information see School of Industrial and Systems Engineering at the Georgia Institute of Technology, 2011).

3. Key performance Indicators (KPIs) and Aggregate indicators based on Data Envelopment Analysis

Benchmarking implies in general the comparison between production units on a set of indicators, usually termed key performance indicators (KPIs). This analysis provides interesting information, as a production unit can understand how it compares with its peers on e.g. operational costs per bed (in the case of a hospital), on average grades of students (in the case of a school), or on sales per employee (in the case of a retail outlet). However, it is important to be aware of the limitations implicit in such comparisons (see Bogetoft and Lars, 2011): (i) KPIs assume constant returns to scale. That is, one assumes that, e.g., the cost per bed is comparable whatever the dimension of the hospital, but this may not be true since bigger hospitals, performing a wider range of services, may naturally present higher costs per bed, and this is not a result of inefficiency, but a result of dimension. (ii) KPIs imply partial evaluations and cannot give an overall picture of performance. That is, a school, e.g., can understand how it performs on a set of, say 10, KPIs but it is not clear how bad performance in some KPIs may, or may not, compensate good performance in others. (iii) In addition, some KPIs may entail trade-offs, meaning that a performance improvement in some KPI, may deteriorate the performance in another (for example improving average grades in schools, may mean neglecting bad students and decreasing completion rates - an equally important KPI). (iv) Lastly KPIs suffer from the Fox Paradox, meaning that a production unit with best performance on individual indicators may present an aggregate performance that is worst than that of a unit with worst performance on the individual indicators. The example in Table 1 shows how this can happen (where CE relates to outpatient appointments and these are divided into first and subsequent appointments, as typically the first time a doctor sees the patient he/she takes more time than in subsequent visits):

Table 1: Fox Paradox illustrated

Hospital	Cost 1st CE/ # 1st CE	Cost CE subs/ # CE subs	Cost CE/ # CE
A	2000/1000 = 2	4000/1000 = 4	6000/2000 = 3
B	30/20 =1.5	800/210= 3.8	830/230 = 3.6

Due to these limitations it is important to complement the KPI analysis with aggregate performance indicators that take several dimensions of the production process into account. This can be done by several means, one of which is Data Envelopment Analysis.

3.1. Data Envelopment Analysis

DEA has been developed as a result of the work of Farrell (1957) by Charnes et al. (1978), for the purpose of measuring the efficiency of non-profit organizations. DEA is a technique that encompasses two definitions of efficiency: (i) Efficiency defined as the distance of production units to the best practice frontier, and (ii) efficiency interpreted as a relative aggregate productivity measure defined as the sum of weighted outputs to weighted inputs. In this latter spirit (that of Charnes et al. (1978)), and considering u_r the weights assigned to outputs y_r (r=1,…,s), and v_i the weights assigned to inputs x_i (i=1,…,m), we can compute for each production unit o its maximum efficiency score through the programming problem (1), where the weights are the decision variables. Problem (1) can be easily transformed into a linear programming problem through the normalization of the denominator of the objective function.

$$Max\left\{h_o = \frac{\sum_{r=1}^{s} u_r y_{ro}}{\sum_{i=1}^{m} v_i x_{io}} \;\middle|\; \frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \le 1,\ j=1,\dots,n,\ \ u_r, v_i \ge 0\right\} \qquad (1)$$

The dual of model (1) allows one to define efficiency as a distance measure from a given point to the frontier (in the spirit of Farrell (1957)). The dual is shown in (2) where decision variables are the Lambdas associated to each production unit j (j=1,…, n) that represent the intensity by which benchmark j has been used in the reference set of unit *o* being assessed.

$$Min\left\{\theta_o \;\middle|\; \sum_{j=1}^{n} \lambda_j y_{rj} \ge y_{ro},\ r=1,\dots,s,\ \sum_{j=1}^{n} \lambda_j x_{ij} \le \theta_o x_{io},\ i=1,\dots,m,\ \lambda_j \ge 0\right\} \qquad (2)$$

To see graphically how a DEA model works consider the following illustrative example, with 7 production units, that use two inputs in the production of a single output. Assume e.g. that inputs are doctors and nurses and that the output is the volume of patients treated in a hospital.

Table 2: Data for illustrative example

DMU	*x1*	*x2*	*y*
A	1	8	2
B	4	7	2
C	15	6	6
D	14	8	8
E	12	16	10
F	12	24	12
H	3	6	2

Data in Table 2 can be represented graphically on a two-dimension graph when inputs are normalized by the output (and therefore the points should read as the amount of each input required to produce one unit of output). The graphical representation is shown in Figure 2.

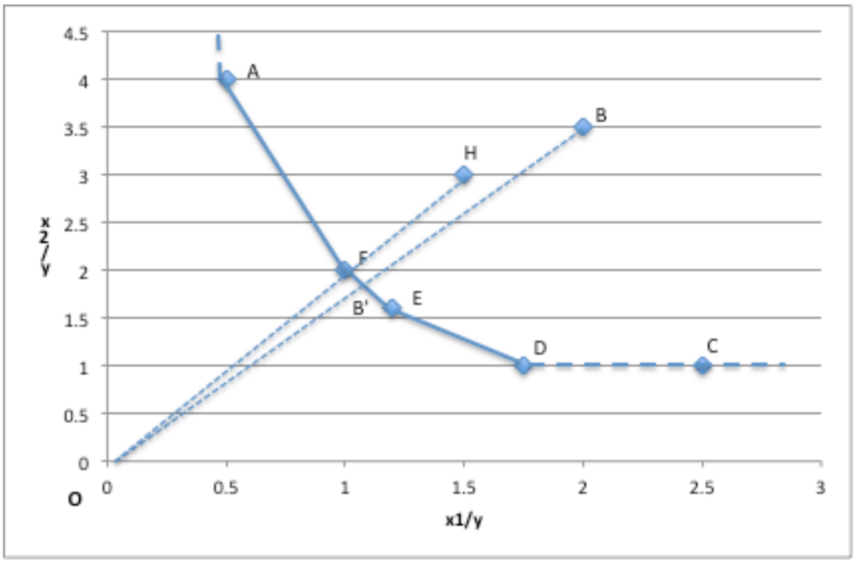


Figure 2: Graphical representation of illustrative example

The efficient frontier is composed by the production units that consume the lower amount of resources for producing the same amount of output. In this case these are units A, F, E and D. Unit C lies on the isoquant but is not considered efficient as it has potential to reduce input 2 and still produce the same amount of output. Unit B lies on the interior of the frontier and can be considered inefficient. Its measure of inefficiency represents the distance between B and the frontier point B', which in model (2) represents θ. Therefore θ for unit B equals OB'/OB. The benchmarks of unit B are units F and E in about the same proportion. The intensities by which the inputs at units F and E are combined convexly to give rise to point B' are the lambda values in model (2). For details see Thanassoulis et al. (2008).

Note that units A and D in Figure 2 use very different combination of inputs to produce one unit of output. In spite of that they are both considered efficient. This is so because we are referring to technical efficiency in Figure 2, and therefore we assume that a unit that is intensive in the use of doctors can be as efficient as one that is intensive in the use of nurses. Putting aside quality issues to simplify the analysis, one can say that from a cost perspective the two units are not equally efficient. The most cost efficient unit will be the one that uses the least expensive mix of inputs.

As a result, if input prices are known an isocost can be added to Figure 2. For prices of input 1 of 2$ and input 2 of 5$, the isocost is represented in Figure 3, where it can be seen that for these prices the cost minimising unit is unit D. The remaining units at the frontier are said to be technically efficient but allocatively inefficient, as they fail to employ the mix of inputs that minimise their cost.

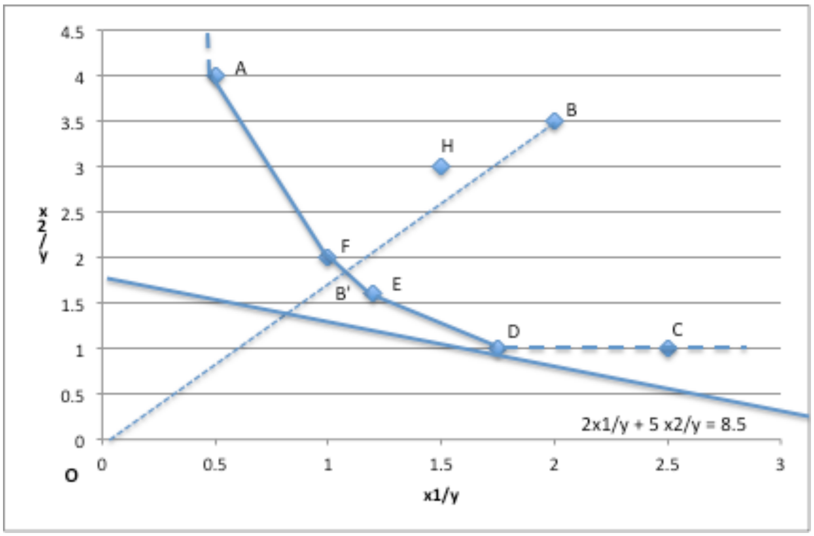


Figure 3: Graphical representation of isocost

The cost minimising problem for unit *o*, when prices of inputs (p_i) are known is given by model (3), where the lambdas and x_i are the decision variables.

$$Min\left\{\sum_{i=1}^{m} p_{io}x_i \middle| \sum_{j=1}^{n}\lambda_j y_{rj} \geq y_{ro},\ r=1,\ldots,s,\ \sum_{j=1}^{n}\lambda_j x_{ij} \leq x_i,\ i=1,\ldots,m,\ \lambda_j \geq 0\right\} \tag{3}$$

When prices are not known the above model can be replaced by model (4), which is only equivalent to model (3) when all units face the same market prices.

$$Min\left\{\sum_{i=1}^{m} C_i \middle| \sum_{j=1}^{n}\lambda_j y_{rj} \geq y_{ro},\ r=1,\ldots,s,\ \sum_{j=1}^{n}\lambda_j C_{ij} \leq C_i,\ i=1,\ldots,m,\ \lambda_j \geq 0\right\} \tag{4}$$

Portela (2014) showed that solving model (4) is equivalent to solving model (5), where all constraints regarding input costs have been aggregated into a single constraint with overall cost.

$$Min\left\{C \middle| \sum_{j=1}^{n}\lambda_j y_{rj} \geq y_{ro},\ r=1,\ldots,s,\ \sum_{j=1}^{n}\lambda_j C_j \leq C,\ \ \lambda_j \geq 0\right\} \tag{5}$$

3.2. An application to the evaluation of hospital services

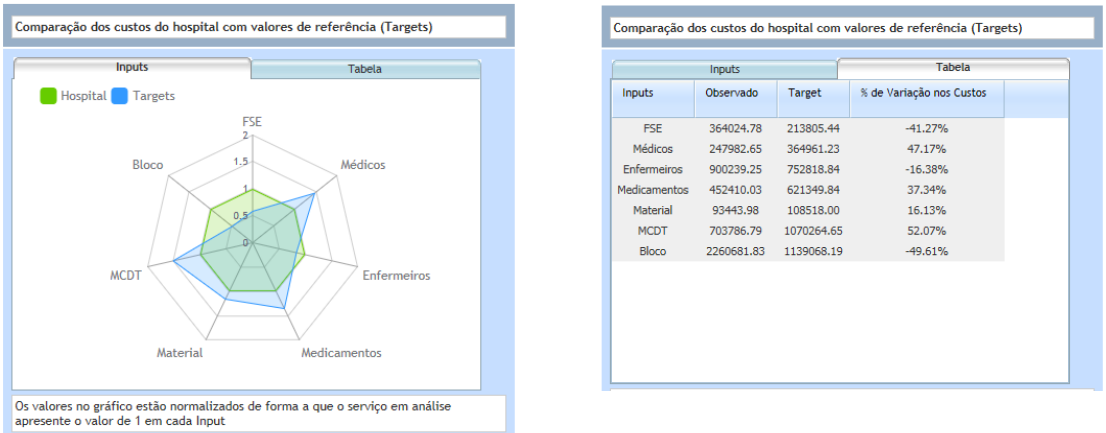
Hollinsworth (2008) advocates that the best level to perform comparisons between hospitals is the service level, as one can say that e.g. Paediatrics departments are more homogeneous amongst hospitals than the hospital itself. In spite of this, there are more studies in the literature at the hospital level than at the service level (some examples of studies at the service level are: Castro et al. (forthcoming), Dervaux et al. (2009)).

We illustrate the application of DEA to the case of the service of general surgery of Portuguese Hospitals following Almeida (2013). This study was also the basis for the DEA models implemented within HOBE. We shall make use of the HOBE platform to present the results of the DEA models here. In the evaluation of hospital services we are interested in computing cost efficiency, and for that purpose we used model (5). The input and output set considered in this evaluation varies per service, but the largest set of inputs and outputs considered can be seen in Table 3.

Table 3: Complete set of inputs and outputs considered in the evaluation of services

Inputs	Outputs
Costs with doctors	N. inpatient days
Costs with nurses	N first outpatient appointments
Costs with medication	N subsequent outpatient appointments
Costs with clinical material	N day hospital sessions
Costs with complementary diagnostic means	N patients in emergency
Costs with surgery units	N days in ICU
Costs with services	N ambulatory surgeries
	N conventional surgeries
	N emergency surgeries

For evaluation purposes the costs have been aggregated into a single cost figure, but only hospitals showing values for the full set of the costs in Table 3 have been considered in the analysis. Taking the example of a hospital (Centro Hospitalar do Nordeste) and using HOBE to compute the efficiency of its general surgery service we see that the efficiency of this service is 85%. Figure 4 shows the outputs in HOBE for this hospital:



Inputs	Observado	Target	% de Variação nos Custos
FSE	364024.78	213805.44	-41.27%
Médicos	247982.65	364961.23	47.17%
Enfermeiros	900239.25	752818.84	-16.38%
Medicamentos	452410.03	621349.84	37.34%
Material	93443.98	108518.00	16.13%
MCDT	703786.79	1070264.65	52.07%
Bloco	2260681.83	1139068.19	-49.61%

Figure 4: HOBE output for the general surgery service of Centro Hospitalar do Nordeste

The figure shows that the general surgery service of this hospital (when compared to its targets, constructed on the basis of the performances observed for benchmarking services) has about the double of service costs (FSE), and of the costs with the surgery unit. Costs with nurses are also higher than target costs, but not as much as the other two. As for the remaining costs they are not higher, but lower than those of its targets, however aggregate costs of this service are higher for this hospital than for its benchmarks, for a similar level of outputs produced. In Figure 5 we can see how this service compares with one of its benchmarks (Centro Hospitalar de Vila Nova de Gaia Espinho)

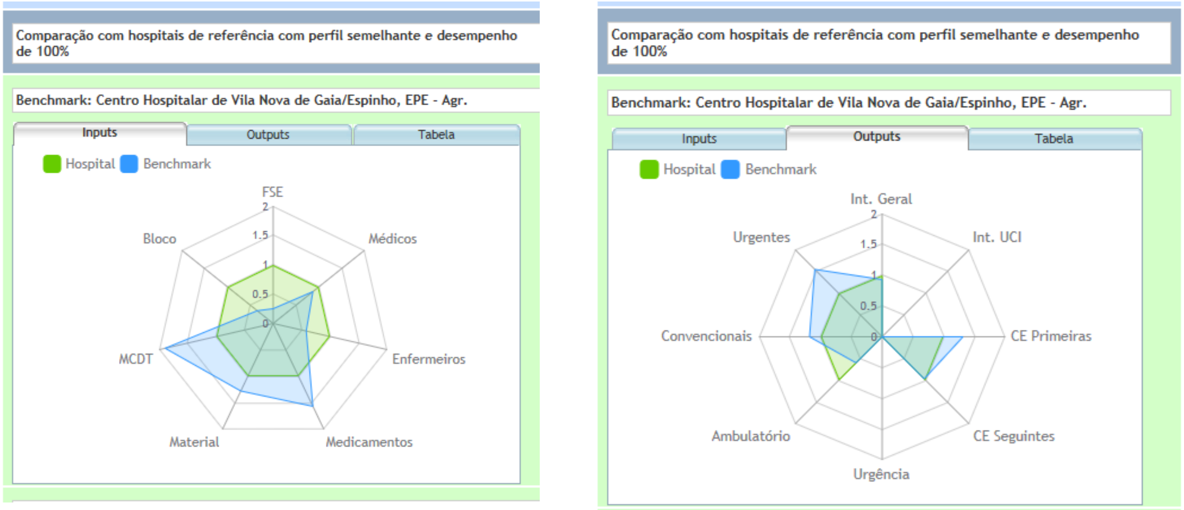


Figure 5: comparison of the service with the service of one of its benchmarks (Centro Hospitalar de Vila Nova de Gaia Espinho)

The Centro Hospitalar do Nordeste shows clearly Higher FSE, higher costs with the surgery unit, and higher nurses and doctors' costs. It shows lower costs with clinical material, medication and complementary diagnosis means, but these do not compensate for the higher costs in the remaining items (note that the values in the radar are normalised such that the observed value of the hospital being evaluated is 1, and therefore one does not have an idea of the magnitude of the differences in monetary value. The actual costs are shown in tables – also available in HOBE). Note that in terms of output production (in the graph on the right of Figure 5), the general surgery service of Centro Hospitalar do Nordeste, shows about the same number of inpatient days and subsequent outpatient appointments as its peer, but a lower number of emergency surgeries, and a lower number of conventional surgeries and first outpatient appointments. The only output where Centro Hospitalar do Nordeste outperforms its peer is ambulatory surgeries. Therefore, this hospital has scope to reduce its overall cost in 15%, and such a reduction should not have an impact on the outputs produced, as it has been shown that similar services are able to produce more outputs with lower costs. Note that implicitly we are making an assumption that the quality between the two services is similar and that reducing costs does not have an impact on service quality. That is why the managerial analysis of hospital performance, although important, should be complemented with a clinical analysis of hospitals too.

Within HOBE each hospital can assess a set of 18 services. The performance of each one can then be aggregated through a DEA model such as model (1) where efficiencies are taken as the outputs of the model, and an unitary fictitious input is used. Such a model imposes additional restrictions such that the weight given to each service is in accordance with the volume of patients treated. The results from this model show an overall picture of the performance of the hospital in each of its services. For Centro Hospitalar do Nordeste the picture is that shown in Figure 6, where it is clear that this hospital has very low performance in the services of pneumatology, urology and anaesthesiology. The reasons for this low performance should be looked at, and actions taken accordingly.

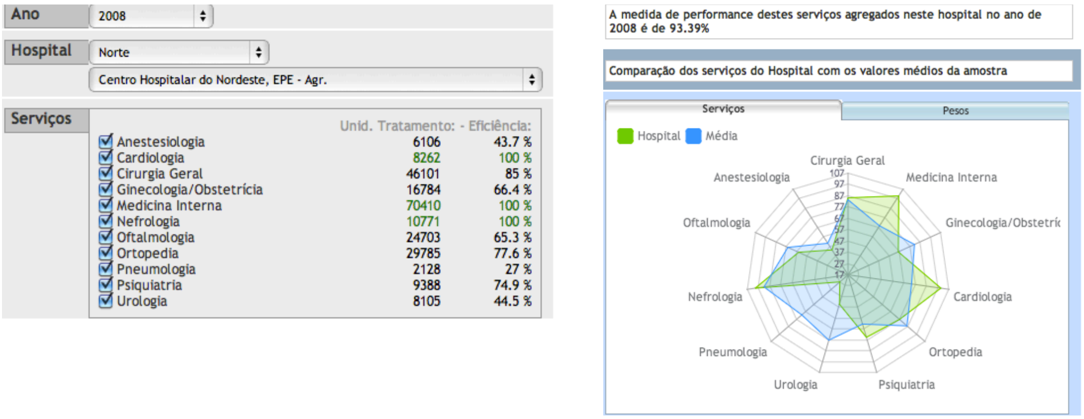


Figure 6: Overall performance of Centro Hospitalar do Nordeste in all its services

HOBE can be used to perform a more detailed analysis on the low performance services by looking carefully at its KPIs. For example, looking at indicators relating to the costs of resources for this hospital's service of pneumatology, one can see that the cost of doctors per hour is amongst the highest (percentile 97%) in a sample of 25 other services. The annual cost of nurses per treatment unit (patient) is also in percentile 97%, indicating values high above the remaining hospitals (see Figure 7). It can be seen in HOBE (but not in the figure below) that the cost of technicians of diagnostic and therapeutic is in the percentile 100% for the service of pneumatology of this hospital. In some other indicators the performance is not so bad, but these high costs, on items that represent a big percentage of the costs of a service, may clearly contribute for the marked inefficiency of this service.

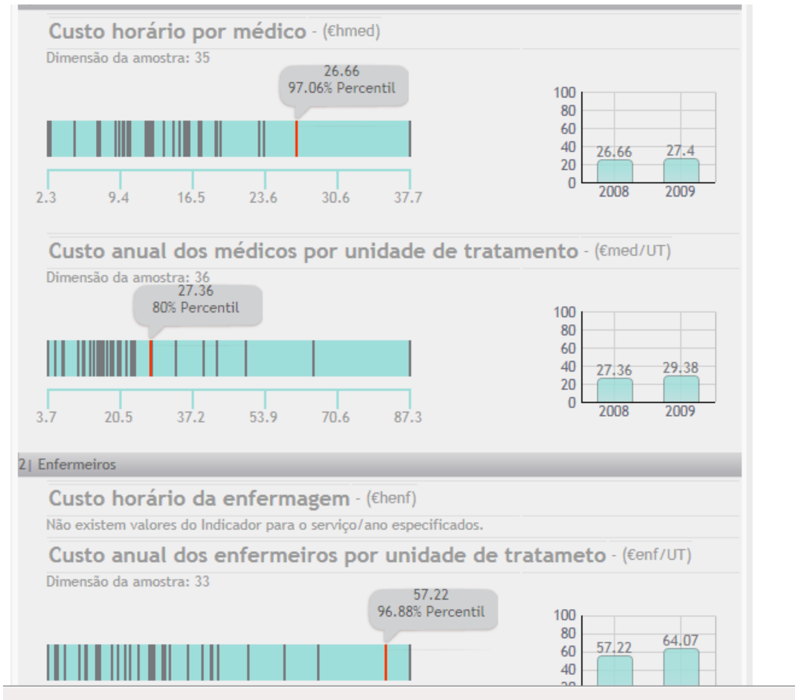


Figure 7: Benchmarking in Hobe for certain KPIs

3.3. An application to the evaluation of schools

Evaluating schools is usually difficult because of the lack of the data regarding socio-economic variables of students, which are known to impact their attainment. There are no academic studies in Portugal regarding the value added of schools taking into account the overall population of Portuguese schools. In BESP, however, an attempt is made to evaluate schools in a contextualized manner. Students' average results on a set of national exams at the 11^{th} and 12^{th} years of the secondary education are taken as outputs of secondary schools. Inputs should ideally relate to the attainment of the same students on entry of the secondary education together with some socio-economic data, like parents education, number of students with economic support from the State, etc. Given the lack of these variables, we used, within BESP, the scores of the school on the 9^{th} grade taken two and three years earlier than the year relating to outputs. The input and output set considered in the assessment of Portuguese secondary schools is shown in Table 4.

Table 4 - Input/Output set for evaluating schools

Inputs	Outputs
Average scores on entry for Mathematics and Portuguese exams (9th grade)	Average scores at the end of Secondary education for each of the following exams: - Portuguese national exam (Pt) - Mathematics exam (Mat) - Biology & Geology national exam (BioGeo) - Physics & Chemistry national exam (FQ) - History national exam (Hist) - Mathematics Applied to Social Sciences national exam (MatAp) - Economy national exam (Econ) - Geography national exam (Geog)

The input specified intends to be a surrogate for the socio-cultural conditions of the cohort of students being assessed on exit, since we would expect that good grades on entry would be associated with good grades on exit. The cohort of students on entry and on exit may, however, be quite different as several new students enter and exit schools in the 10^{th} grade and others abandon the school.

In BESP we implemented a DEA assessment in two ways: an assessment that is available to the general public based on the input/output set shown in Table 4 and an assessment that is available only to schools, where they have some freedom to choose the list of inputs and outputs that can be used in their performance assessment. In both cases the model solved is model (1) with some additional constraints on the weights to force schools to weight more the exams that more students took (see Portela et al., 2012).

For the general public there is available a table with the schools overall performance (in percentage) and a classification of the school in three performance bands: high, medium, low. These bands are a simple representation of the sample higher, medium and lower thirds. Figure 8 shows an extract of the alphabetically ordered list of schools for the year 2008/2009, where, for this year, the low performance band interval range from 0 to 91.6%, the medium performance band from 91.61 to 97.36% and the high performance band from 97.37% to 100%

Legend: High Performance   Medium Performance   Low Performance

School	Performance	Range
Academia de Música de Santa Cecília (Lisboa)	93.96%	
Centro de Educação Integral (Aveiro)	92.16%	
Centro de Estudos Básico e Secundário - CEBES (Porto)	91.22%	
Centro de Estudos de Fátima - CEF (Santarém)	100%	

Figure 8 - List of schools, with the indication of the overall efficiency score

Apart from this table we also show to the general public a radar graph with the average of each input and output indicator for schools in each band as shown in Figure 9 (year 2008/2009). This radar graph gives an idea of how the high, medium of low performing schools stand against each other.

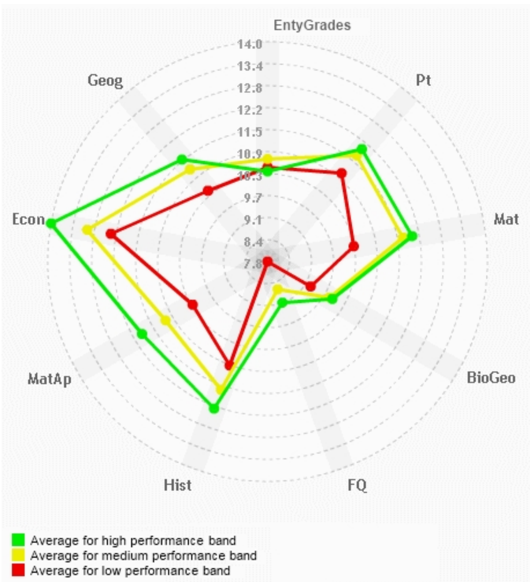


Figure 9 - Radar graph with the average input and outputs in each performance band

For registered schools, the DEA assessment involves a first step where the school can choose the set of inputs and outputs from a pre-defined list. This choice recognises that schools can have different priorities in promoting some outputs in detriment of others. The efficiency results are calculated in real-time, and the information is presented in a radar graph as shown in Figure 10. For this example, the inputs and outputs selected are those shown in Table 4.

In Figure 10 it is possible to see that school Escola Secundária de Coruche has an efficiency value of 79.2%. The graph shows that this school (orange) has similar input than its targets, but output values are considerably lower than the targets, with Physics and Chemistry (FQ) being the case with the biggest scope for improvement. To be considered efficient, this school should have reached the targets while maintaining the current input values.

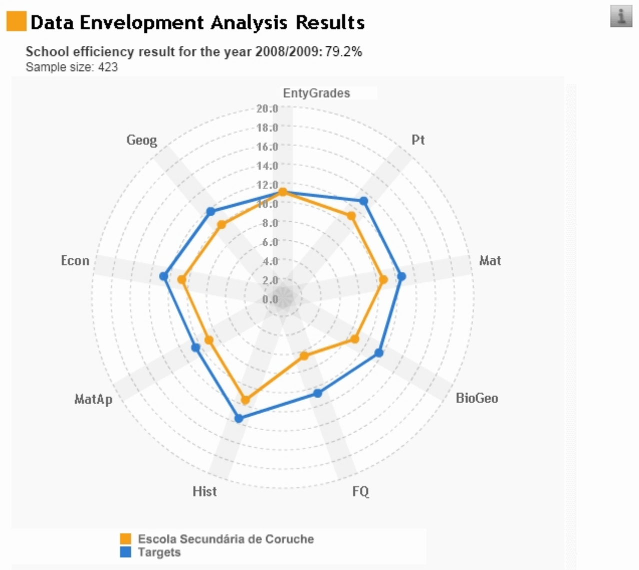


Figure 10 - Radar graph with the school's targets

Figure 11 shows the radar graphs comparing the school values with the most significant benchmarks. The school values are orange and the targets and benchmarks values are blue, following the standard adopted in the BESP platform. In the graph on the left of Figure 10 the benchmark school has a higher input value (EntryGrades) than the school under assessment but this difference is compensated by even higher values for the output variables. As for the graph on the right, the benchmark school has a lower input value, but is able to achieve higher values in every output variable than the school under assessment. For both cases the biggest offset is seen in Physics and Chemistry in accordance with the targets shown in Figure 10.

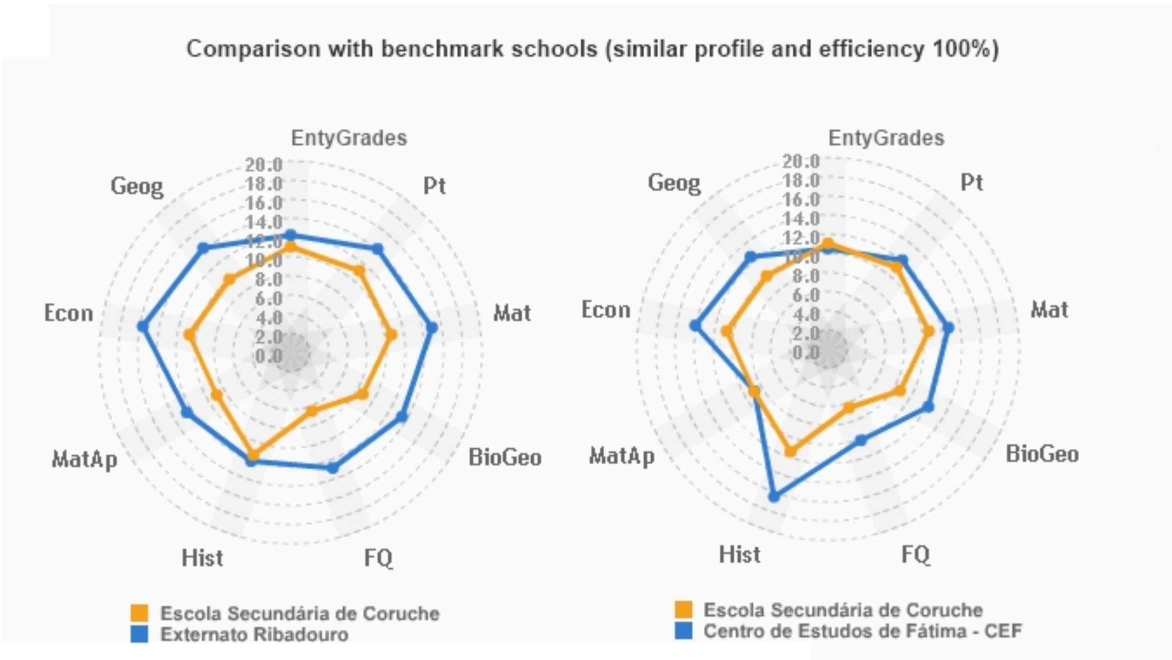


Figure 11 - Radar graph with the school's benchmarks

4. Conclusion

The aim of this paper was to illustrate the concept of benchmarking applied to two public services in Portugal. We reviewed some literature in terms of previous academic studies done in the education and health areas, and also described existing benchmarking web platforms, with some incidence towards the ones existing in Portugal (two of which were developed by the author). The necessity of defining KPIs is addressed and also the main limitations regarding their use. The DEA methodology is presented as a way to circumvent some of the problems of KPIs, and it is explained and illustrated through an example. Finally two empirical applications are presented, with the aid of the developed web-platforms to illustrate the use of the DEA methodology and its potential. No political or practical implications can be taken from the assessments presented because data are old and inaccurate (in the case of health), or because existing data are not enough for taking serious political and managerial conclusions (as in the case of schools). We hope, however, that the empirical applications illustrated that, in the hand of good quality data, the tools proposed can indeed be used as an aid for management actions and political guidance.

References

Afonso, A. and St. Aubyn, M. (2005). Non-parametric Approaches to Education and Health Efficiency in OECD Countries. *Journal of Applied Economics, 8(2), p. 227-246.*

Almeida, D. (2013). Hospital Efficiency as an Aggregate of Services' Efficiency: A DEA Approach in Universidade Católica Portuguesa, Master's dissertation for completion of the MSc in Industrial Engineering and management, FEUP.

Beckford, John. (1998). Quality: A Critical Introduction: Routledge.

Behin-Cara-Pajoh Research Center of Operations Research. (2013). Data Envelopment Analysis Online Software. https://www.deaos.com/.

Bogetoft, P. and Nielsen. (2005). Internet Based Benchmarking. *Group Decision and Negotiation* 14 (3) (May): 195–215.

Bogetoft, P. and Lars, O. (2011). Benchmarking with DEA, SFA, and R, Springer

Campbell, R., and Buetow. (2000). Defining Quality of Care. *Social Science & Medicine* 51 (11) (December): 1611–25.

Castro, R. Portela, MCAS., Camanho, A. (forthcoming). Benchmarking dos Serviços dos Hospitais Portugueses: Uma Aplicação de Data Envelopment Analysis, In Casos de Aplicação da IO.

Charnes, A., Cooper, W. W., and Rhodes, E. (1978). Measuring efficiency of decision making units. European Journal of Operational Research, 2:429–444.

Chilingerian, J. A. (1995). Evaluating physician efficiency in hospitals: A multivariate analysis of best practices. *European Journal of Operational Research, 80, p.548– 574.*

Chilingerian, and Sherman. (2004). Health Care Applications: From Hospitals to Physicians, From Productive Efficiency to Quality Frontiers. In *Handbook on Data Envelopment Analysis*, edited by William W Cooper, Lawrence M Seiford, and Joe Zhu. Boston / Dordrecht / London: Kluwer Academic Publisher.

Dervaux, B., H. Leleu, E. Minvielle, V. Valdmanis, P. Aegerter , B. Guidet (2009). Performance of French intensive care units: A directional distance function approach at the patient level, Int. J. Production Economics 120, 585–594.

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

Dismuke, C. E. and Sena, V. (2001). Is there a Trade-Off between Quality and Productivity? The Case of Diagnostic Technologies in Portugal, Annals of Operations Research 107, 101–116,

ERS. (2011). "SINAS - Sistema Nacional de Avaliação Em Saúde. http://www.websinas.com/sinas/index.html.

Farrell, M. J. (1957). The measurement of productive efficiency. Journal of the Royal Statistical Society, Series A, general 120(Part 3):253–281.

Gray, J. Goldstein, H. and D. Jesson (1996). Changes and improvements in school's effectiveness: trends over five years. Research Papers in Educa- tion 11(1), 35–51.

Grosskopf, S., Hayes, K. J., Taylor, L. L., and Weber, W. L. (1999). Anticipating the consequences of school reform: A new use of DEA. Management Science, 45(4):608–620.

Hanushek, E. and L. Taylor (1990). Alternative assessments of performance of schools: Measurement of state variation in achievement. The Journal of Human Resources 25(2), 179–201.

Hollingsworth, B. (2008). The measurement of efficiency and productivity of health care delivery. Health Economics no. 17 (10):1107-1128

Ibensoft ApS. (2013). Interactive Benchmarking - State-of-the-art in Performance Evaluation. http://www.ibensoft.com/.

Kirjavainen, T. and Loikkanen, H. A. (1998). Efficiency differences of Finnish senior secondary schools: An application of DEA and Tobit analysis. Economics of Education Review, 17(4):377–394.

Kontodimopoulos, N., Niakas, D. (2005). Efficiency measurement of hemodialysis units in Greece with data envelopment analysis. *Health Policy 71, p.195–204.*

Ladd, H. F. and R. P. Walsh (2002). Implementing value-added measures of school effectiveness: getting the incentives right. Economics of Education Review 21(1), 1–17.

Laudicella, M. Olsen, K. R. and Street, A. (2010). Examining cost variation across hospital departments: a two-stage multi-level approach using patient-level data Social Science & Medicine 71, 1872-1881

Mancebón, M. and Bandrés, E. (1999). Efficiency evaluation in secondary schools: the key role of model specification and of ex post analysis of results. Education Economics, 7(2):131–152.

Moreira, S. 2008. Análise dos Hospitais-Empresa: Uma Aplicação da Data Envelopment Analysis. Boletim Económico, Banco de Portugal.

OECD. (2008a) Handbook on Constructing Composite Indicators: METHODOLOGY AND USER GUIDE, OECD (http://composite-indicators.jrc.ec.europa.eu/).

OECD. (2008b) Measuring Improvements in Learning Outcomes: best practices to assess the value added of schools, OECD publications.

Olsen, K.R. and Street, A. (2008) The analysis of efficiency among a small number of organizations: how inferences can be improved by exploiting patient-level data, health Economics, 17. 671-681.

O'Donoghue, C., S. Thomas, H. Goldstein, and T. Knight (1997). 1996 DfEE study of value added for 16-18 year olds in England. DfEE Research Series March.

Portela, MCAS., Camanho, A.S. and Borges, D. (2011), *BESP - Benchmarking of Portuguese Secondary Schools*, Benchmarking: and International Journal 18/2, 240-260.

Portela, MCAS., Camanho A.S. and Borges, D. (2012), Performance assessment of secondary schools: the snapshot of a country taken by DEA, Journal of the Operational Research Society 63, 1098-1115.

Portela, MCAS and A. Camanho (2010). Analysis of complementary methodologies for the estimation of school value-added. Journal of the Operational Research Society 61, 1122–1132.

Portela, MCAS. and E. Thanassoulis (2001). Decomposing school and school type efficiency. European Journal of Operational Research 132(2), 114–130.

Portela, MCAS. (2014), Value and quantity data in economic and technical efficiency measurement, Economics Letters 124, 108-112.

Portela, MCAS., Camanho, A.S., Almeida, D., Nogueira Silva, S., Lopes, L. and Castro, R. (2014), *HOBE - Benchmarking Hospitals through a web based platform, Submitted to Benchmarking: An International Journal.*

Puig-Junoy, J. (1998). Technical efficiency in the clinical management of critically ill patients. Health Economics 7, 263–277.

School of Industrial and Systems Engineering do Georgia Institute of Technology. (2011). iDEAs. http://www2.isye.gatech.edu/ideas/.

Thanassoulis, E, M.C.A.S. Portela, and O. Despic (2008), Data Envelopment Analysis - The Mathematical Programming Approach to Efficiency Analysis, In Fried, HO., Lovell, CAK., and Schmidt, SS., *The Measurement o Productive Efficiency and Productivity Growth*, Oxford University Press, 251-420.

Thanassoulis, E., Portela, MCAS., and Graveney, M. (2012), Estimating the scope for savings in referrals and drug prescription costs in the General Practice units of a UK Primary Care Trust, European Journal of Operational Research, 221, 432-444.