Abstract

Purpose – The purpose of this paper is to illustrate the implementation of a web-based platform integrating benchmarking and data envelopment analysis (DEA) for the Portuguese secondary schools.

Design/methodology/approach – The benchmarking platform is designed around a set of key performance indicators that are displayed using benchmarking graphs. These indicators are also aggregated through the methodology of DEA to provide a summary measure of performance.

Findings – The benchmarking platform developed enables schools to perform internal and external evaluation through a standard model that is based on indicators of school activities. It encourages schools' efforts of continuous improvement and increases society awareness regarding schools' context and results obtained.

Practical implications – The benchmarking platform can be useful for schools and general public. For the general public, there is a tool that allows the construction of user-defined rankings online and benchmarking tools that allow the comparison of performance of a specific school with others. In addition to these features, schools have in the BESP platform a repository of historical data, and the possibility to see a set of graphs that show for some indicators their evolution over time.

Originality/value – This paper describes a breakthrough in the Portuguese education context. The BESP platform is the first in this context to combine DEA and benchmarking tools in a web-based environment, designed to enable real-time performance assessments.

Keywords Benchmarking, Secondary schools, Performance measures, Data analysis, Portugal

Paper type Technical paper

BESP is a benchmarking platform for secondary schools whose aim is to help schools in their internal and external evaluation. For that purpose, a set of key performance indicators was developed allowing schools to analyse, in real time, their performance relative to other schools, and the evolution of their performance over time. The platform allows the comparison of a school with others in a given indicator, shows an aggregate performance indicator, constructed through data envelopment analysis (DEA) and allows the construction of school league tables based on user-defined criteria. The indicators are constructed from data that comes from two sources: public databases and schools’ private information. Public data comprise exam results on various subjects at the beginning and at the end of secondary education, and private data are collected...
through online questionnaires. The users of the platform (schools and general public) have different access profiles, with schools having full access and the general public having restricted access, only to indicators constructed from public data.

1. Introduction

In the last couple of years, social and political awareness towards the quality of education and school evaluation has been steadily increasing worldwide. This concern led to the creation of new working committees (e.g. the European Union effective school self-evaluation (ESSE)) to define schools’ improvement factors.

In Portugal, school evaluation is a recent reality. Several studies, programs or projects on education evaluation were put into action since the 1990s but the number of these initiatives is still very small. This may be due to the fact that in Portugal there has not been a tradition of school evaluation, and the perception of its vital role to promote quality improvements is not yet shared by all schools’ stakeholders.

An increasing awareness for the importance of school evaluation prompted Portugal to follow the European trend, and led to new legislation (e.g. “Dispatch 484/2006” about integrated evaluation systems at the university level and “Law 31/2002, 20th December” about evaluation in the basic and secondary education).

In Portugal, the current procedure for schools’ external evaluation is carried out by the Portuguese General Inspectorate of Education (IGE). Visits to schools are conducted by an IGE team, which produces a report evaluating the school on several domains: results, school management and organization, leadership, auto-regulation and improvement capacity. Besides, external evaluation, schools are obliged (by legislation) to have internal evaluation programs. However, in practice, each school follows its own internal evaluation program (since only general guidelines on the evaluation dimensions are provided by legislation), and in a large number of schools this evaluation is not done on a systematic basis.

While schools struggle with mandatory evaluation procedures, the Portuguese newspapers publish school rankings every year based on student results in national exams. These rankings are controversial due to the lack of contextualization (i.e. they do not consider the socio-cultural context or student achievements before entering the school). While criticized by many (Matos et al., 2006) and despite not being a particularly fair ranking, they are a starting point for school and society awareness.

The development of the internet platform BESP – Benchmarking of Portuguese Secondary Schools (http://besp.mercatura.pt) was motivated by the desire to help schools in their evaluation procedures, specifically those with secondary education. The BESP platform is a supporting base for internal and external evaluation of schools, as well as a tool for fostering the accountability of schools.

In BESP, a set of key performance indicators plays a central role. These indicators cover not only student results, but also the school context, resources and processes, a structure that follows the context, input, product and process (CIPP) model from Stufflebeam (2003). Benchmarking is performed for each indicator through the identification of the school’s percentile, allowing the school to know the percentage of schools lying above and below its own level in a given indicator (Costa et al., 2007). Adopting this benchmarking exercise aids school administrators to pursue best practices and to set targets based on comparisons with the best organizations in the industry (Camp, 1989). For other examples of benchmarking methodologies, see Ribeiro and Cabral (2006).
Besides the benchmarking purpose, BESP has a useful and innovative feature that concerns the integration of DEA models into the benchmarking exercise. DEA is a methodology that allows the measurement of efficiency of a decision-making unit (a school in our case) in the processes of using a set of resources (inputs) to produce a set of results (outputs). DEA enables a relative efficiency assessment, meaning that schools are compared with other schools in a set of variables previously chosen as inputs and outputs. As efficiency is a relative measure based on direct comparisons with other schools, the sample used influences the quality of the results obtained. The DEA model implemented in the BESP platform aims to produce a summary measure of performance (or efficiency), based on which school outputs are compared in a contextualised way, i.e. taking into account some contextual indicators, such as student grades on entry or their socio-economic context. In BESP, the assessment of schools through DEA adopts a value-added perspective. This perspective can be seen as the progress schools help students to make relative to their different starting points. This perspective of assessment was previously applied to the Portuguese education context in Portela and Camanho (2009), where student results on entry and student results on exit were considered as inputs and outputs of the assessment, respectively (Portela and Thanassoulis, 2001).

The objective of this paper is to present the BESP platform, describing the tools available and their potential, emphasising their utility for schools and the general public. Behind the creation of BESP are wider objectives related to an effort to understand the factors contributing to school performance and quality, the critical performance areas in each school and an identification of true benchmark schools that can be used by others as role models. BESP is a first step towards these wider objectives. In order to understand performance differences between schools, other type of studies, mainly qualitative, may also be conducted focusing particularly on the worst and best schools. BESP collects and treats data on schools performance over time, and therefore it is a valuable instrument to support this type of studies. A deeper analysis of the performance of Portuguese schools is, however, outside the scope of this paper, whose main goal is to describe the key features of the BESP platform.

The contributions of the paper are twofold. On the one hand, the paper contributes to the literature by putting forward a new platform for benchmarking and evaluating schools. On the other hand, it builds on existing benchmarking tools by presenting for the first time in education (as far as the authors are aware) a DEA composite indicator for evaluating schools on a set of key performance indicators. The DEA functionality allows the computation of efficiency scores for schools in real time and further allows schools to choose the dimensions on which they wish to be compared with other schools.

This paper unfolds as follows. Section 2 presents a brief review of previous school performance assessment studies and benchmarking exercises supported by web platforms. Section 3 describes some methodological aspects behind the creation of BESP. Section 4 presents the BESP platform, describing each of the tools available and the type of results that are displayed within the platform. Section 5 concludes the paper.

2. State of the art
Benchmarking has become a common practice in education, as well as in other sectors. Benchmarking can be defined as “the continuous process of measuring products, services and practices against the toughest competitors or those companies recognised as industry leaders” (Camp, 1989, p. 10). According to Camp (1989), to be successful,
benchmarking must be continuous and systematic. It cannot be performed once and disregarded thereafter on the belief that the task is done. Costa et al. (2006) also urge the importance of designing a set of indicators that will effectively support companies in creating improvement measures.

In this brief review, we describe some previous school performance studies and attempt to present benchmarking results interactively through web platforms. While studies abound on the former, there are few examples of internet benchmarking platforms in the educational context.

School benchmarking studies are varied and can range from simple rankings of schools to more complex studies, where regression-based or frontier methods (from which we can differentiate between stochastic frontier methods, which are regression-based, or non-parametric DEA models, which are linear programming models) are used to compare schools. Examples of regression studies include Levitt and Joyce (1987), Sammons et al. (1996) or Tavares et al. (2002). Examples of stochastic frontier models include Pereira and Moreira (2007), who applied stochastic frontiers to the case of Portuguese schools.

Frontier methods based on DEA (Charnes et al., 1978) have become very popular as a way to evaluate schools, rank them and suggest routes for improvement. DEA works by comparing each unit (schools in the context of this paper) with a frontier constituted by units that show maximum performance in a set of output indicators, for a given level of resources or inputs used. Therefore, the schools located at the frontier (or a subset of these) can be regarded as benchmarks for the schools located below the frontier. A detailed analysis of the practices of benchmark units can provide guidelines for improvement.

There are many DEA studies on schools in the literature. School performance studies differ widely in aim and type of results produced. Most studies assessing schools' performance use national exam scores of students as the main outcome of schools, given the ease of access to such data. Most differences between studies happen on the factors used for contextualising these exam scores. This is generally done through the consideration of student grades on entry and other socio-economic variables as inputs of the model (e.g. parents’ average education, ratio of economic deprivation, or percentage of free school meals). When output variables are related to attainment on exit and the aim is to try to contextualise these results with pupil-related factors that contribute for explaining such attainment (including grades on entry and/or some socio-economic characteristics of pupils), then the perspective of the evaluation is school’s value added. This evaluation considers the contribution that the school provides to a given cohort of students. For details on the concept of value added, see for example Meyer (1997). The literature on measuring school’s value added is quite vast and most models, particularly those with student level data, are based on multilevel modelling (see, e.g. Goldstein (1999) and Ferrão and Goldstein (2009) with an application to the Portuguese case). We however restrict our focus to DEA models, since these are directly related to the models implemented in the BESP platform.

An example of a DEA study adopting a value-added perspective can be found in Kirjavainen and Loikkanen (1998), who studied the efficiency of Finnish secondary schools. The outputs reflect student achievement and the inputs reflect the admission level of students, their parental context, as well as other variables related to the school (teaching and non-teaching hours) and variables related to teachers (experience and education). Following a similar approach, but using data at the student level, Portela and Thanassoulis (2001) used DEA to decompose student under-attainment that could be attributed to the school and to student effects. This study also adopted a value-added
perspective since scores on entry were treated as inputs and scores on exit were treated as outputs. See also Portela and Camanho (2009), who used a similar approach to compute the value added of Portuguese students and schools. Cherchye et al. (2010) also used student-level data to assess efficiency of Flemish schools using DEA. The approach used was slightly different from the studies described above. Efficiency was first computed at the student level, considering as inputs the total number of instruction units assigned to a student and the total number of “equal education opportunity” instruction units assigned to that student. The outputs defined represented two test scores. These resulting efficiency scores were then adjusted with environmental variables reflecting the socio-economic status (SES) of students and entry levels. The results suggested that SES variables and student achievement on entry have a positive impact on exam results. However, for lower levels of SES or lower levels of exam levels on entry, the impact on efficiency is higher than that verified at higher levels of these variables.

Mancebo and Bandrés (1999) used DEA to assess Spanish schools with a different perspective of schools’ efficiency. The authors also considered achievements of students in national test scores as outputs (as indeed most of the studies analysing schools’ efficiency do, given the ease of access to this type of data), but included as inputs the operating expenses of the school per pupil, the number of teachers per pupil and a set of contextual variables measuring the academic quality of students. The inclusion of cost variables on the input side of a school’s assessment takes into account not only value added but also value for money (Mayston, 2003). In Grosskopf et al. (1999), cost-related input variables were also used for the efficiency assessment of Texas school districts.

In terms of benchmarking platforms, i.e. web-based interfaces for the evaluation of schools, there are some examples worth mentioning. In the UK, the Department for Children Schools and Families (DCSF) publishes every year school performance tables (www.dcsf.gov.uk/performancetables). Two important objectives of these tables are “informing parents in their choice of school and providing schools with an incentive to raise their standards” (OECD, 2008, p. 46). These tables contain raw data for the student examination results at the end of compulsory education and contextual value-added scores for a restricted set of key indicators (the value-added model comprises prior attainment and contextual variables) to predict attainment (Ray, 2006).

The UK’s web-based application “report and analysis for improvement through school self-evaluation” (RAISE) (www.raiseonline.org) enables schools to analyse performance data in greater depth as part of their self-evaluation process. This application is only available to schools, providing them a wider range of data than that provided by the performance tables of DCSF. The results of the analysis are shown in percentile line graphs or snake plots (a system replicated in BESP), and box-plots enabling the identification of possible weaknesses and strengths of schools (for details on this tool see Ray et al., 2008).

Also in the UK, there is the School Financial Benchmarking (SFB) web site with restricted access to schools (http://sfb.teachernet.gov.uk). This web site is based on financial data from the consistent financial reporting framework that holds financial data from all maintained schools in England. This web site enables schools to compare their expenditure with that of similar schools. Among the different functionalities, schools can access their standard data, create “what if scenarios” and select other schools against which to compare their income and expenditure.
Norway has been using for the last few years a national school accountability system (http://skoleporten.utdanningsdirektoratet.no/english/Sider/default.aspx). This system started in 2004 and its central element is a web-based application containing data on indicators for results, resource use and learning environment. The indicators are intended to provide standardised and comprehensive information concerning Norwegian schools to different stakeholders, such as school managers, government, parents, students and general public (Hægeland, 2006).

Sweden, through the National Agency for Education (www.skolverket.se/sb/d/190) publishes data for all levels of the education system. Besides, the data on several different indicators, the agency also publishes expected results for each individual school, estimated with linear regression. School performance indicators are then calculated as the difference between the school actual results and its expected results.

In Table I, we present a summary of the functionalities available in the web platforms described above and compare these functionalities with those available within BESP. Note that in some cases we could not check the existence of certain functionalities either because the access is restricted or due to the unavailability of an English version of the platform (in the Swedish and Norwegian cases). Note that BESP has some additional features that no other educational platform comprises. In particular, it makes extensive use of radar graphs showing details on the percentiles of the school in several sub-indicators (e.g. the average grades on each subject included in the indicator “aggregate average grade”, the percentage of success in each secondary school year included in the indicator “aggregate success rate”). BESP also uses DEA to compute an aggregate measure of school performance and allows a customisation of this performance measure, to make it consistent with each school educational project, as will be seen in Section 4.

Outside the education sector, there are other examples of internet-based benchmarking platforms. The construction industry is a particularly good example of a sector with numerous benchmarking web sites.

The BESP platform gained from the knowledge acquired in the development of a benchmarking platform in the Portuguese construction industry: icBench – developed by the same entity that was involved in the development of BESP. The platform icBench (www.icbench.net) uses a set of key performance indicators, based on which external benchmarking comparisons are made. In addition, the platform also acts as a repository of companies’ internal data, which can aid managerial decision making. The conceptual model of this web site was used as a reference for the development of BESP.

| System                | Restricted school area | Access to general public | League tables | Snake plot | Radar graph | DEA | Longitudinal analysis | VA |
|-----------------------|------------------------|--------------------------|---------------|------------|-------------|-----|-----------------------|-----|
| Performance tables (UK)| ×                      | ✓                        | ✓             | ✓          | ×           | ×   | ×                     | ✓   |
| RAISE (UK)            | ✓                      | ×                        | ×             | ✓          | ×           | ×   | ?                     | ✓   |
| SFB (UK)              | ✓                      | ×                        | ×             | ✓          | ×           | ×   | ×                     | ×   |
| Skoleporten (Norway)  | ✓                      | ✓                        | ✓             | ✓          | ×           | ×   | ?                     | ×   |
| Skolverket (Sweden)   | ✓                      | ✓                        | ?             | ?          | ×           | ✓   | ×                     | ×   |
| BESP (Portugal)       | ✓                      | ✓                        | ✓             | ✓          | ✓           | ✓   | ✓                     | ✓   |

Table I. Comparison of school benchmarking platforms
resulting in similarities in the site structure and graphical presentation of data (for details on this tool see Costa et al, 2007).

Two examples of other construction industry benchmarking platforms worth mentioning can be found in the USA (BM&M – www.construction-institute.org) and the UK (KPIzone – www.kpizone.com).

The platforms previously mentioned, aside from BESP, do not include an aggregate measure of performance based on individual indicators, and in particular none of them uses DEA to compute such a measure. There are, however, some examples of platforms that use DEA to produce aggregate performance indicators online. One example is iDEAs (www.isye.gatech.edu/ideas) by the Georgia Institute of Technology in the USA. This platform is targeted to warehouses or other industrial systems and combines benchmarking and DEA to allow managers to benchmark their performance against others. Users that complete online questionnaires for a set of indicators can then evaluate the performance of a particular warehouse by comparison to a large number of other warehouses.

In Denmark, there is an internet-based benchmarking system, applied to Danish commercial and savings banks, that incorporates a DEA model. The user interface allows interactivity and control over the model parameters. This interaction enables individualized analyses, which is a differentiating feature in relation to other web-based applications. This tool enables comparing the bank under assessment with benchmark banks. For details on this tool, see Bogetoft and Nielsen (2005).

3. Methodology
School performance indicators play a central role in BESP. The choice of indicators to use was a complex process that involved essentially two sub-processes: investigating existing models in the literature, and undertaking meetings with school managers to understand the central issues for self-evaluation procedures that should be considered in benchmarking exercises.

There are some performance evaluation models in education that inspired the basic structure of the key performance indicators available in BESP. ESSE puts forward three main dimensions for school evaluation: inputs, processes and outputs. The “European report on the quality of school education” sets four areas of evaluation: attainment, success and transition, monitoring of school education, resources and structures (Directorate-General for Education and Culture, 2000). Common Assessment Framework is a self-evaluation model specifically developed to support European public organizations. It was adapted from European Foundation for Quality Management to focus primarily on public sector schools in the areas of leadership, strategy, planning, people, partnerships, resources and processes. The CIPP model of Stufflebeam (2003) is one of the most broadly accepted models in the education sector. It distinguishes four dimensions of analysis: context, input, process and product. Context evaluations assess needs, problems, assets, priorities and outcomes. Input evaluations assess alternative approaches, action plans and budgets. Process evaluations assess the implementation of plans to help staff carry out their activities. Product evaluations identify and assess outcomes, intended and unintended, in the short and the long term.

The indicator set constructed within BESP adopted the CIPP model, classifying indicators into four major areas: context, resources, results and processes. In total, the BESP platform considers 123 indicators spread over these four dimensions of evaluation. Examples of indicators considered within each of category include the following.
Context indicators. These indicators concern the socio-economic context of the school, such as parents’ education (measured by the average number of years in school), economic deprivation level of the school, rate of selectivity of the school and average instruction level in the school area.

Resources indicators. These indicators concern the resources used in daily school activities, such as school state budget per student, external financial contributions, number of teachers per student, teachers’ average age, teachers’ average number of years in career, number of teaching staff per non-teaching staff, number of computers per student, schools’ own revenues per student and teaching staff salary per student.

Results indicators. These indicators concern the main outcomes from the school activities, such as average results at the end of secondary education achieved by the school students in national exams. These indicators are computed for a set of eight national exams with the highest number of students enrolled, and aggregated into a single indicator using an weighted average. Examples of other indicators include average results on entry (i.e. average results obtained by school’s students at the end of basic education on Portuguese and mathematics exams), percentage of students that finished secondary education in three years, percentage of students that entered public universities and percentage of failure in each subject.

Process indicators. These indicators concern the practices and procedures of the school. These indicators are the only ones that are constructed based on ordinal data (on a Likert scale) collected from the online questionnaire. Some examples of indicators available in BESP under this heading are: extra-curricular projects, parents’ participation, training courses attended by teachers, teachers’ motivation, group work, excellence board and discipline system.

The selection of the indicators that should be considered within each of these broad dimensions, and the specification of the formulas underlying their calculation, was supported by visits of the research team to secondary schools, which provided the opportunity to meet with school managers. Two workshops with school directors were also organised to discuss in group, with about 20 participants, the type of indicators that could be more useful for school management. The issues discussed concerned data availability, the effort required for data collection, relevance of certain indicators and compromises between quantity of indicators and quality of the information provided. When the BESP platform was at a pilot stage, some more meetings with school managers were conducted with the aim of introducing the platform, and at the same time getting feedback on the utility of the set of indicators available and finding ways to improve the platform.

4. BESP platform
The BESP platform organizes information regarding Portuguese secondary schools’ performance. The Portuguese education system comprises two levels: basic education and secondary education. These educational levels follow a common curriculum in all schools determined by the central government. Basic education in Portugal is compulsory and comprises three cycles. To conclude the third cycle, which normally happens at the age of 15, students are submitted to national exams on the subjects of Portuguese and mathematics. Secondary education comprises three years (tenth, 11th and 12th). To conclude the secondary education students that are enrolled in scientific-humanistic studies, which are intended to prepare for tertiary education, are submitted
to national exams. Apart from scientific-humanistic studies students may enrol in other studies within a school, such as professional studies, technological studies or specialized arts. The BESP platform has been designed to include indicators both for the scientific-humanistic studies and the professional studies (which are the ones with the highest percentage of students in the country). However, the indicators developed for professional studies depend entirely on the private data entered by schools. Therefore, our main focus in this paper is on the students that follow scientific-humanistic studies.

Part of the BESP homepage is shown in Figure 1.

This page has a diagram that allows users to navigate in the web site and shows that the performance indicators play a central role in the benchmarking platform. It can also be seen from this figure that the data used in the computation of indicators come from two sources: the public databases of the Ministry of Education, with students’ results on national exams at the end of basic education (ENEB database) and at the end of secondary education (ENES database), and schools private data, collected through questionnaires available within BESP. The questionnaires are created on a yearly basis.

The BESP platform has been designed for two types of public: schools managers and general population. School users have a reserved area where they can fill in school details, access the questionnaire to be answered online and have full access to the functionalities concerning benchmarking and internal evaluation. These features are only partially available to the general public since only the indicators based on publicly available data are shown. The following sections explain the various functionalities within the BESP platform and indicate whether the functionalities are available only to schools or both to schools and the general public.

Figure 1.
BESP homepage
4.1 Benchmarking
The benchmarking functionality within BESP aims at comparing, for each indicator, the school under assessment with a selected sample of schools. This functionality is available both for the general public and the schools, but full access is only available to schools. To perform benchmarking for a given school, the user must choose the sample for comparison and the year of assessment (five years of data are currently available in the platform). Schools can be compared with those in the same district, with schools of the same type of ownership (public or private), with schools with the same or lower values of average instruction level in the school area (meaning schools located in the same or less advantaged socio-cultural conditions), with schools in the same type of location (central urban, peripheral urban, or rural), or with the full sample of Portuguese secondary schools. The objective of restricting the comparator set of schools is to allow for more contextualised comparisons, which are not usually a concern in traditional league tables. That is, a school can compare itself with the full set of Portuguese schools if this is considered to be a fair assessment, or otherwise the school can choose to be compared only with schools in a similar context.

The benchmarking results shown in the platform for each indicator appear in the form of a snake plot, as shown in Figure 2.

This is an example of a benchmarking graph for a given school in the indicator “Average results for national exams at the end of secondary education”. In this type of graph, the vertical axis represents the value of the indicator and the horizontal axis indicates the percentile. A horizontal line is drawn intercepting the accumulated curve on the value obtained by the school in the indicator considered and a vertical line intercepts the horizontal axis in the percentile value. For this particular case, the school has an average result of 11.24 (in a scale from 0 to 20) and this value corresponds to the percentile 73.39 per cent. This means that the results of the school for this indicator are
better than those observed in 73.39 per cent of the national schools, which compose the sample selected for the evaluation.

The indicator above is an aggregate of several indicators representing the average result on national exams for a number of subjects. For aggregate indicators such as this, BESP also presents a radar graph that allows a detailed visualisation of the percentiles of the individual indicators that constitute the aggregate indicator (Figure 3).

For this particular school, it is clear that the worst performance happens for the average result in mathematics applied to social sciences (where the percentile of the school is the lowest), and the best performance happens for the average result in Portuguese (where the percentile of the school is the highest).

4.2 Internal evaluation
BESP allows the school to explore changes on its own performance over time, in addition to the benchmarking analysis consisting of direct comparisons with other schools. This can be done within the functionality of “internal evaluation”, which is only available to registered users (i.e. schools). The outcomes that are shown within this functionality are column graphs that display the evolution of the school over time in a given indicator. This is a useful instrument for a school self-evaluation since it can be used for diagnosing areas in need of improvement. The graphs presented in the “internal evaluation” section are shown in Figure 4.

Figure 4 shows the evolution of a school over a period of four years in the indicator “Average results on national exams of secondary education” (graph on the left). In absolute terms, the result of this indicator varies between 11.07 and 12.41, with fluctuations over the years. It is worth mentioning that the value of this indicator in a given year is related to the level of difficulty of the national exam in that year. Therefore, to complement the analysis and remove the effect related to variations in the level of difficulty of the exams, the indicator is also presented in relative terms (i.e. the average result for each school is divided by the national average). Thus, the indicator computed relative to the national average shows the evolution of school performance, taking into account the context of the universe of Portuguese secondary schools. For the school
shown in Figure 4, it is possible to see that the results were above the national average (above 1.0) every year, but steadily decreasing over the years, with a minimum value of 1.05 in the last year analysed. Selecting the details button shown in Figure 4 provides a graphic illustration of the results of the individual exams (i.e. bar graphs for each exam) that underlie the calculation of the aggregate indicator.

For the indicators constructed with private data, the visualisation of graphs for internal evaluation or benchmarking depends on the completion of the online questionnaires by registered school users. To increase the representativeness and quality of the performance comparisons, schools collaboration by completing every year the questionnaires available within the BESP platform is essential.

4.3 Overall, efficiency measure
Analyzing the BESP outcomes, it is possible to have an idea of how a school is performing in several dimensions. However, the multiplicity of indicators makes it difficult to gain an overall view of school performance, since a school may perform very well on some indicators, but poorly on others. Therefore, we created within BESP a performance indicator that gives an overall idea of school performance when several indicators are considered together. Typical composite indicators are only applied to outcomes, but DEA provide the means to account for both resources and outcomes.

The set of indicators that were used for constructing a DEA composite indicator are shown in Table II. These indicators were divided into input indicators and output indicators, since it was important to consider input variables to contextualise school outcomes.

The outputs considered are related to attainment on exit of secondary education, disaggregated on various subjects. In order not to bias the results by taking into account subjects with few students evaluated, we considered for each school only the average scores for exams with more than ten students. If a school does not have a score on a given subject, or if the number of students doing the exam on that subject is lower than ten, the score considered was null (this implies that in the performance assessment the school cannot take into account this subject).

In relation to the inputs, we considered the average results on entry of secondary education. Given the flow of students over time within a school, it is not possible to obtain an exact match of the cohort of students on entry and on exit. To obtain an input set that could be considered a good approximation of the cohort of students that followed
secondary education within each school, we specified the inputs and outputs with a time lag of three years, i.e. we considered the average scores of students on exams of basic education (input) three years before the year considered for the exam results of secondary education (outputs).

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. As seen before, this input/output set is coherent with a value-added perspective of the schools’ performance (note however, that in general value-added measures of schools are computed at the student level and not at the aggregate level, as in this case). For the computation of a composite indicator, the selection of the contextual variables (inputs) and outcomes (outputs) was restricted to data publicly available. Some schools already started using the platform and entering data, but the universe of schools for which private data are available is still small.

In BESP, we implemented the DEA assessment in two ways: an assessment that is available to the general public based on the input/output set shown in Table II 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.

For the general public, BESP shows a table with the schools overall performance (in percentage) and a classification of the school in three performance bands: high, medium and low. These bands are simple representations of the sample higher, medium and lower thirds.

Figure 5 shows an extract of the ordered list for the year 2008/2009. For this year, the low-performance band interval range is from 62.0 to 91.6 per cent, the medium performance band from 91.6 to 97.4 per cent and the high-performance band from 97.4 to 100 per cent.

Table II. Input/output set

| Inputs | Outputs |
|--------|---------|
| Average scores on entry for mathematics and Portuguese exams (ninth grade) | Average scores at the end of secondary education for each of the following exams: Portuguese national exam (Pt) Mathematics exam (Mat) Biology and geology national exam (BioGeo) Physics and chemistry national exam (FQ) History national exam (Hist) Mathematics applied to social sciences national exam (MatAp) Economy national exam (Econ) Geography national exam (Geog) |

Figure 5. 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 input and output indicators for schools in each band, as shown in Figure 6 (year 2008/2009). This radar graph gives an idea of how the high, medium of low-performing schools stand against each other. For example, we can see in Figure 6 that schools with high-performance have lower input than schools in other bands, but are able to achieve considerably higher output levels.

The general public can also see a histogram of the distribution of the overall performance measure, where each dot represents a school and its colour represents the school’s performance band (high, medium and low) (Figure 7).

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. The flexibility allowed in the selection of inputs and outputs recognise that schools can have different priorities in promoting some outputs in detriment of others. The indicators that can be selected for inputs are: parents’ level of education, economic deprivation index, average scores on entry at Portuguese and mathematics (ninth grade), average Portuguese score on entry (ninth grade) and average mathematics score on entry (ninth grade). The outputs available include those in Table II plus the percentage of students proposed to secondary level national exams, the percentage of students that finished secondary education in three years, the percentage of students that entered public universities and the percentage of dropouts.

Note that apart from the inputs that the school can choose to “contextualise” its outputs, the school can also restrict the comparators (i.e. the sample of other schools...
considered in the benchmarking analysis). This contributes to enable a contextualised assessment by comparing the school under evaluation only to schools with the same type of ownership (public/private), the same type of location (rural/urban centre/urban periphery), or within the same district or council. It is also allowed to restrict the comparators to schools with a similar or lower average instruction level in the council.

The efficiency results are calculated in real time, and the information is presented in a radar graph as that shown in Figure 8. For this example, the inputs and outputs selected are those in Table II.

Figure 8 shows a radar graph with the school output and input values in orange and the average values for the sample selected in blue. The efficiency score is also indicated next to the radar chart (100 per cent in this example). An efficiency value of 100 per cent indicates that, for the input and output set selected, the school achieved the best possible results when compared to other schools in the sample. It is possible to see in Figure 8 that the school is better than the sample average in every output, with the exception of mathematics. However, there is no evidence that this output could be improved without reducing at least one of the others outputs or increasing the input levels.

Performance values lower than 100 per cent indicate that there is potential for improvement. For schools that do not attain the maximum efficiency score, the BESP platform produces a radar chart as shown in Figure 9, comparing the school current values with its targets. The targets for the variables are obtained as by-products of the DEA model, and result from a linear combination of the input and output levels observed in the benchmark schools, i.e. schools that are an example of best observed performance (with a score of 100 per cent). The input/output values of the benchmark schools shown in Figure 10 are used to produce the targets seen in Figure 9 for the school under assessment.

In Figure 9, the school under assessment has output values considerably lower than the targets. Physics and chemistry (FQ) is the subject with the biggest scope for improvement.
The DEA model used for this assessment is output oriented, meaning that the inputs are assumed to be fixed, and only potential output improvements are explored.

Figure 10 shows the radar graphs comparing the school values with the most significant benchmarks. The school original values are represented in orange and the targets and benchmarks values are in 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. In the graph on the right, the benchmark school has a lower input value, but for every output variable it is able to achieve higher values than the school under assessment. In both cases, the biggest offset corresponds to physics and chemistry, in accordance with the targets shown in Figure 9.

4.4 Other features

Rankings. The BESP platform has other features from which the most relevant is the elaboration of online rankings of schools, similar to those published annually in the Portuguese media. The advantage of using the web platform for this purpose is that it allows the user to choose the parameters based on which the ranking is made. The user may choose, for example, to compute a ranking of schools for specific exams, or for the average result on the eight national exams with more students enrolled. The ranking list
Data envelopment analysis results
School efficiency result for the year 2008/2009: 79.2%
Sample size: 423

Figure 9.
Radar graph with the school’s targets

Comparison with benchmark schools (similar profile and efficiency 100%)

Figure 10.
Radar graph with the school’s benchmarks
may include all secondary schools, only schools in the same geographical location (same district or same council), or only schools with the same type of ownership (public or private). It is also possible to restrict the sample to female or male students only, or to schools that did more than a certain number of exams. In this page, the user can elaborate ordered lists of schools from the national exams database. For the elaboration of these rankings, we consider only internal students that do the national exam for the first time (retakes are excluded because different exams cannot be used for the same comparison).

**Report creation.** For registered schools, it is possible to produce automatically a report with all the information available in the platform for a given year. Alternatively, a school may choose to include only some pre-selected indicators in the report. This functionality produces a pdf document, which can be very useful for distribution to school teachers or stakeholders. Nevertheless, the graphs can be printed individually directly from the platform, if desired.

**5. Conclusions and future research**

In a large number of countries, school evaluation has become a routine, although in Portugal it is still an emergent reality. The BESP platform was created with the aim to promote increased accountability of Portuguese schools, and make available a tool to support self-evaluation and benchmarking processes. BESP allows schools to make performance comparisons (external evaluation) with other schools on several indicators and provides the opportunity to carry the mandatory internal evaluation processes with a standard tool. The BESP platform also acts as a repository of school data for a large number of indicators. This is also innovative in the Portuguese context, as the data of public databases were not integrated, so there was no access to historical information concerning national schools.

This paper described the BESP platform in some detail, showing the main functionalities available: benchmarking snake plots, internal evaluation graphs and the DEA composite indicator of performance. This feature was given particular emphasis, since it is indeed new in relation to the functionalities available in benchmarking platforms of schools available in other countries. This composite indicator is available for the general public, and enables the classification of schools in three performance bands. For registered schools, this functionality allows an automatic computation of a DEA performance measure with input and output variables selected by the user (that is, the functionality is parametrizable by schools and allows flexibility in the choice of the most relevant variables to be considered in the performance evaluation).

Despite the usefulness and utility of BESP (recognised by all the schools that are already using the platform), it also has a few limitations. The main limitation of BESP is the dependency from data introduced directly by schools. That is, some indicators can only be computed if schools provide data by completing annually the BESP questionnaires. As soon as a registered school answers the online questionnaires, it can have access to a repository of data that shows the evolution of performance over time for all indicators available within BESP. However, as far as benchmarking is concerned, for indicators based on private data (such as the percentage of students entering university) the school may not have an adequate number of schools to be compared with, since the universe of comparison depends on the number of schools that previously introduced their data in the platform. Therefore, the full potential of BESP will only be exploitable when a reasonable number of schools introduce data on the platform.
The research team is currently making efforts to obtain additional school data from the Ministry of Education. This would have a great impact in BESP, as it would increase the reliability of the information available in BESP, and reduce significantly the questionnaire size, and consequently diminish the effort required to schools concerning the provision of data. Furthermore, BESP could have a bigger number of active indicators and a meaningful universe of comparison for most indicators.

There are other aspects that will be subject to future developments. In particular, the team involved in BESP expects to expand the platform to other cycles of education (currently BESP only deals with secondary schools). In addition, the DEA models will be further developed to consider restrictions on the weights assigned to the outputs associated to exam scores. Currently, the DEA model gives complete freedom to the assignment of weights to input and output variables (although schools that made less than ten exams on a given subject cannot use that subject as an output). In order to better analyse the performance of schools and better discriminate their efficiency levels, the DEA formulation will be improved by linking the weights assigned to each exam to the number of exams made by the school students. Such restrictions should prevent undesirable weighting schemes, such as placing more weight on the exam of a subject that very few students attended than on other subjects with a larger number of students enrolled.

References
Bogetoft, P. and Nielsen, K. (2005), “Internet based benchmarking”, Group Decision and Negotiation, Vol. 14 No. 3, pp. 195-215.
Camp, R.C. (1989), Benchmarking: The Search for Industry Best Practices that Lead to Superior Performance, ASQC Quality Press, Milwaukee, WI.
Charnes, A., Cooper, W.W. and Rhodes, E. (1978), “Measuring the efficiency of decision making units”, European Journal of Operational Research, Vol. 2 No. 4, pp. 429-44.
Cherchye, L., De Witte, K., Ooghe, E. and Nicaise, I. (2010), “Efficiency and equity in private and public education: a nonparametric comparison”, European Journal of Operational Research, Vol. 202 No. 2, pp. 563-73.
Costa, D., Formoso, C., Kagioglou, M., Alarcón, L. and Caldas, C. (2006), “Benchmarking initiatives in the construction industry: lessons learned and improvement opportunities”, Journal of Management in Engineering, Vol. 22 No. 4, pp. 158-67.
Costa, J.M., Horta, I., Guimaraes, R., Falcão e Cunha, J., Nóvoa, H. and Sousa, R.S. (2007), “icBench – a benchmarking tool for Portuguese construction industry companies”, International Journal for Housing Science and Its Applications, Vol. 31 No. 1, pp. 33-41.
Directorate-General for Education and Culture (2000), European Report on the Quality of School Education – Sixteen Quality Indicators, Official Publications of the European Communities, Luxembourg, available at: http://ec.europa.eu/education/policies/educ/indic/rapinen.pdf (accessed 18 December 2009).
Ferrão, M.E. and Goldstein, H. (2009), “Adjusting for measurement error in the value added model: evidence from Portugal”, Quality & Quantity, Vol. 43, pp. 951-63.
Goldstein, H. (1999), Multilevel Statistical Models, Multilevel Models Project, London Institute of Education, London, available at: www.sozioologie.uni-halle.de/langer/multilevel/books/goldstein.pdf (accessed 2 February 2010).
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, Vol. 45 No. 4, pp. 608-20.
Hægeland, T. (2006), *School Performance Indicators in Norway: A Background Report for the OECD Project on the Development of Value-added Models in Education Systems*, OECD, Paris.

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*, Vol. 17 No. 4, pp. 377-94.

Levitt, M.S. and Joyce, M.A.S. (1987), *The Growth and Efficiency of Public Spending*, Cambridge University Press, Cambridge.

Mancebón, M.J. and Bandrés, E. (1999), “Efficiency evaluation in secondary schools: the key role of model specification and of ex post analysis of results”, *Education Economics*, Vol. 7 No. 2, pp. 131-52.

Matos, M., Lopes, C., Nunes, S. and Venâncio, I. (2006), “Reflexões sobre os rankings do Secundário”, *Investigação Operacional*, Vol. 26, pp. 1-21.

Mayston, D. (2003), “Measuring and managing educational performance”, *Journal of Operational Research Society*, Vol. 54, pp. 679-91.

Meyer, R.H. (1997), “Value-added indicators of school performance: a primer”, *Economics of Education Review*, Vol. 16 No. 3, pp. 283-301.

OECD (2008), *Measuring Improvements in Learning Outcomes – Best Practices to Assess the Value-Added of Schools*, OECD, Paris.

Pereira, M.C. and Moreira, S. (2007), “A stochastic frontier analysis of secondary education output in Portugal”, Estudos e Documentos de Trabalho, Banco de Portugal Working Paper 6|2007.

Portela, M.C.A.S. and Camanho, A.S. (2009), “Analysis of complementary methodologies for the estimation of school value-added”, *Journal of the Operational Research Society*, Vol. 61 No. 7, pp. 1122-32.

Portela, M.C.A.S. and Thanassoulis, E. (2001), “Decomposing school and school-type efficiency”, *European Journal of Operational Research*, Vol. 132 No. 2, pp. 357-73.

Ray, A. (2006), *School Value-added Measures in England – A Paper for the OECD Project on the Development of Value-added Models in Education Systems*, Department for Education and Skills, London, available at: www.dcsf.gov.uk/research/data/uploadfiles/RW88.pdf (accessed 7 February 2010).

Ray, A., Evans, H. and McCormack, T. (2008), “The use of national value added models for school improvement in English schools”, *Revista de Educación*, Vol. 348, pp. 47-66.

Ribeiro, L. and Cabral, J. (2006), “A benchmarking methodology for metalcasting industry”, *Benchmarking: An International Journal*, Vol. 13 Nos 1/2, pp. 23-35.

Sammons, P., Thomas, M.P., Owen, C. and Pennell, H. (1996), *Assessing School Effectiveness: Developing Measures to Put School Performance in Context*, OFSTED Publications Centre, London.

Stufflebeam, D.L. (2003), “The CIPP model for evaluation”, in Kellaghan, T. and Stufflebeam, D.L. (Eds), *International Handbook of Educational Evaluation*, Kluwer Academic Publishers, Boston, MA, pp. 31-62.

Tavares, L.V., Graça, P.M. and Tavares, M.M.V. (2002), *Estudo SEDES: Assimetrias Regionais do Desempenho educativo*, Observatório de Prospectiva da Engenharia e da tecnologia e Centro de Sistemas Urbanos e Regionais do IST, Lisboa.
About the authors

Maria Conceição A. Silva Portela is Assistant Professor in Quantitative Methods at Catholic University of Portugal. She has a PhD in Management Science from Aston University, UK (2004). She has published several refereed papers in the theory and application of DEA as an efficiency and productivity measurement tool (e.g. in European Journal of Operational Research, Journal of Operational Research Society, Journal of Productivity Analysis, Annals of Operations Research, Computers and Operations Research and Journal of Banking and Finance). She is Co-author of two book chapters in Handbook on Data Envelopment Analysis, Kluwer Academic Publishers and The Measurement of Productive Efficiency and Productivity Growth, Oxford University Press. She has research, consultancy and teaching experience in DEA, having applied it to various fields such as bank branches, schools, retail outlets, water companies and general practice units. She has been responsible for two projects financed by the Portuguese Foundation of Science and Technology on the assessment and benchmarking of Portuguese schools. Maria Conceição A. Silva Portela is the corresponding author and can be contacted at: csilva@porto.ucp.pt

Ana Santos Camanho is an Assistant Professor at the School of Engineering of the University of Porto. She has a PhD in Industrial and Business Studies from Warwick Business School, UK (1999). The main research area is operational research/management science, with focus on efficiency and productivity measurement techniques. She is Pro-director of the Master and Doctoral Programs in Industrial Engineering and Management at the School of Engineering of the University of Porto. She has several papers in peer-reviewed international journals in the management science field (e.g. European Journal of Operational Research, Journal of Productivity Analysis, Omega – The International Journal of Management Science, Annals of Operations Research, Computers and Operations Research, Journal of the Operational Research Society, Journal of Construction Engineering and Management and Fisheries Research). She has been involved in research projects in the following areas: banking, education, fisheries, retailing, construction industry and urban quality of life.

Diogo Novoa Borges is a Researcher at Catholic University of Portugal. The author has a Master in Computer Science from University of Porto (2009). The main areas of interest are information technology and operations research. The author has been involved in a project on the assessment and benchmarking of Portuguese schools.

To purchase reprints of this article please e-mail: reprints@emeraldinsight.com
Or visit our web site for further details: www.emeraldinsight.com/reprints