Performance indicators and rankings in higher education

To be published in:

Public and Private Benefits of Higher Education: An Appreciation of the Work of Gareth Williams, Institute of Education Press (2016)

Jill Johnes

The Business School, University of Huddersfield, Queensgate, Huddersfield, HD1 3DH, UK

Email: j.johnes@hud.ac.uk

Telephone: +44 (0)1484 472231

Performance indicators and rankings in higher education[1]

Introduction

Publicly funded sectors are under pressure to use resources efficiently, and awarding funds on the basis of performance is one approach to trying to achieve a more efficient use of taxpayers' money. Resource allocation to the UK higher education sector has long followed such a policy. The 1980s funding cuts to UK higher education prompted rapid development of performance indicators as a means of encouraging accountability: '…higher education policy was dominated by two main concerns: to help reduce public expenditure; and to increase efficiency by encouraging institutions to 'earn' a larger proportion of their income from both government and non-government sources, and to be explicitly accountable for it' (Williams, 1992b, pp3-4). This observation strongly resonates with the current situation in English higher education.[2]

The performance of a traditional firm is relatively straightforward to measure since it is typically assumed to have the objective of profit maximisation. The firm's accounts therefore provide an indication of how well the firm is performing against this benchmark. Any firm where the assumption of profit maximisation is not applicable cannot have its performance assessed in this manner; non-profit institutions universities fall into this category and a conventional approach is inappropriate. In the UK, the need for performance indicators in the higher education sector has long been recognised (Department of Education and Science, 1985). Proposed performance indicators initially focused on a particular output or operation (Jarratt, 1985) and were at best simple ratios such as the proportion of students with 'good' degrees or the cost per student. Media interest in these performance indicators was also stimulated and the performance of universities in certain key areas of interest to both tax

---

[1] Jill Johnes, Professor of Production Economics, The Business School, University of Huddersfield, Queensgate, Huddersfield, HD1 3DH. I am grateful to Gerry Steele (Lancaster University) for comments on an earlier draft.
[2] See HEFCE, http://www.hefce.ac.uk/funding/efficient/.

payers and prospective students was put under the spotlight: for example, labour market destinations (Dixon, 1985); completion rates (Dixon, 1989); and achievement rates (Dixon, 1976).

It is a simple step to go from performance indicators to rankings and league tables. We should highlight at the outset the distinctions between performance indicators and university rankings; the differences revolve around the presentation and usage of the two. The former are usually a set of quantitative data on the performance of HEIs typically used by policy-makers in assessing whether resources are being used efficiently. The latter are lists of HEIs, often produced by commercial publishers, ranked in descending order of performance according to a set of quantitative data and presented in the format of a league table (Usher and Medow, 2009). Rankings draw attention to relative performance, and have largely been aimed at the general public, in particular prospective students and their parents to help them in making an informed choice about their university. The first serious[3] media rankings of universities and colleges (at the institution level) appeared in *US News and World Report* in 1983 (Dill, 2009), and since 1994 the publication of university league tables based on various individual measures of performance has become commonplace in the UK, USA and elsewhere (Yorke, 1997). Distinctions between performance indicators and rankings are blurred by the fact that official performance data often underpin the rankings produced by the media. Moreover, interest and usage of rankings has widened to universities themselves (as an internal auditing and resource allocation tool) and to governments, especially with the recent development and regular publication of global rankings of universities across the world (for example, the 'Shanghai rankings', or the rankings of the *Times Higher Education*).

[3] There have been some rankings of US HEIs produced prior to 1983. But these have been produced *ad hoc* and do not follow the methodology of many of today's media rankings.

Higher education institutions (HEIs) are multi-product firms with complex production processes, so that indicators based on simple ratios (and rankings which utilize these ratios) are unlikely to capture the true picture of performance. My own PhD thesis, for which Gareth Williams was external examiner, represents an early attempt to address these issues and to suggest ways of measuring the value-added of HEIs rather than raw outputs (Johnes and Taylor, 1990b). In the ensuing thirty years, as techniques have developed to capture performance in a multi-output multi-input production framework, it has become possible to develop more sophisticated indicators. There is a trade-off, however, between simplicity and complexity; between providing indicators which are easy to construct and interpret, and computing performance measures which more closely capture the production process but whose construction is poorly understood by the layperson. The continued interest in media rankings suggests that simplicity is the current winner. But the simple approach can provide a misleading picture of performance with adverse consequences for institutional behaviour, as will be explored further in the following sections.

This chapter examines the development and use of performance indicators and rankings in the context of higher education, from the use of individual indicators and their amalgamation into a composite measure, to the construction of efficiency measures from an economist's perspective of universities as multi-product firms. The potential effects of these indicators and rankings on national policy and institutional behaviour are then considered together with proposals for an alternative way forward. The chapter ends with final conclusions. While the ideas and methodological approaches are of general application and interest, illustrative examples are typically from the UK.

Developing measures of performance in higher education: a simple approach

Higher education in the UK and elsewhere often receives funding from government. This can be potentially problematic as the objectives of those funding higher education and those running universities may not be in alignment leading to a classic principal-agent problem (Johnes, 1992). The government which provides the funds (the principal) can only imperfectly observe the actions of those running the HEIs (the agents) and so resorts to performance indicators to ensure that its own goals are being met. This raises the question of how to construct meaningful performance indicators.

One approach to performance measurement involves deriving various indicators and using these (or a subset) for a particular purpose. For example, funding for a particular activity can be given on the basis of performance in that area; or prospective students might examine performance in areas which are of particular relevance to their own interests. In arriving at a set of appropriate performance indicators, two main issues need to be initially addressed:

i) Level of analysis: What are the entities being measured?

ii) Dimensions: What are the dimensions by which performance should be measured?

The separate performance indicators produced by official agencies such as the Higher Education Funding Council for England (HEFCE) and the Higher Education Statistics Agency (HESA) in the UK are used (often by the media) to produce rankings of universities which purport to highlight relative performance of HEIs over a spectrum of dimensions. Thus a third issue which needs to be addressed is:

iii) Weights: How can individual dimensions be aggregated into one composite index?

We will consider each of these in turn.

Level of analysis

Within higher education, we might want to measure the performance of individual academics, departments, academic programmes, institutions or even the whole sector (against higher education sectors in other countries). The level of analysis depends on the targeted audience and the purpose of the evaluation. The opening statement of HESA on their performance indicators which they have published since 2002/03 (prior to which these were published by HEFCE) makes it clear that the entity in which they are interested is the higher education provider i.e. the university.[4] Lower-level performance (eg of departments), however, might be of interest to managers within institutions, and country-level performance to governments interested in the performance of the domestic higher education sector relative to those of global competitors.

Dimensions

The dimensions by which we measure university performance depend, like level of analysis, on the target audience and purpose of evaluation. In the context of government monitoring of university performance, the dimensions should relate closely to the outcomes from higher education most valued by society (Dill, 2009). Research and teaching are two obvious dimensions of interest. HEFCE undertakes a periodic review of UK universities' research on behalf of all the funding councils. The most recent review was the Research Excellence Framework (REF) the results of which were published in 2014.[5] This was preceded by various Research Assessment Exercises (RAEs) undertaken in 1986, 1989, 1992, 1996, 2001 and 2008. The greater part of funds distributed by the funding councils to universities for research purposes have been allocated on the basis of REF/RAE results (Harman, 2011), and there is a clear link between funding and measured performance. The UK is not alone in

---

[4] See HESA, https://www.hesa.ac.uk/pis.
[5] The results of the REF can be found here: http://www.ref.ac.uk/.

funding research in this way (Dill, 2009); Australia, for example, uses similar performance-based schemes to fund research and research training.[6]

The UK has no analogous mechanism for funding teaching, where student numbers drive allocations. Student numbers determine resources in two ways: through the university tuition fee; and through HEFCE resourcing which is currently linked to student numbers by subject.[7] The latter is therefore not performance-related and the former is linked to teaching performance only insofar as students might choose their university and programme on the basis of teaching reputation. In practice, of course, this requires reliable indicators to inform potential students. A Teaching Excellence Framework (TEF) to mirror the REF is therefore currently being mooted in the UK as a means of linking funding to teaching performance with the explicit aim of improving teaching quality and ensuring students and taxpayers of efficient use of resources.[8]

There may be more than two dimensions of interest, however, and a variety of performance indicators can be found. Jarratt (1985) provides an early attempt at suggesting performance indicators which could provide useful information to stakeholders. The development of performance indicators has steadily progressed in the UK since then with numerous indicators now being published annually (Pollard *et al.*, 2013), and the media compile their own rankings from these official sources, as well as any additional data which they have collected.

It is not difficult to find data on all aspects of university activities. HESA provides a huge amount of data on UK higher education including indicators relating to:[9] widening participation rates; non-continuation rates; module completion rates; research output; and

[6] See https://education.gov.au/research-block-grants.
[7] See http://www.hefce.ac.uk/lt/howfund/.
[8] See https://www.gov.uk/government/speeches/teaching-at-the-heart-of-the-system.
[9] Source: https://www.hesa.ac.uk/pis accessed 29th June 2015.

graduate employment. The media address the requirements of prospective students whose interests lie with feedback from current students, the reputation of the institution or simply with the facilities at each institution. *The Complete University Guide*, for example, examines 10 aspects of activity likely to be of interest to prospective students:[10] entry standards; student satisfaction; research assessment; research intensity; graduate prospects; student-staff ratio; academic services spend; facilities spend; good honours; and degree completion.[11]

Numerous dimensions can pose problems of interpretation if performance varies across different measures. Indeed there is no reason to expect that a university that is good in one area will necessarily be good in another. Take the rankings from *The Complete University Guide*. A simple rank correlation of the 10 measures in the most recent ranking (2015-16) illustrates the point (table 1). While the majority of indicators are highly correlated, 17 pairs have a correlation coefficient below 0.4[12]. This means that a university's position in the ranking changes dramatically depending on which indicator is used: one HEI, for example, is ranked top on academic services, second on research intensity, but bottom on facilities, and amongst the bottom 10 universities on student satisfaction. Indeed the student satisfaction indicator generally appears to provide a noticeably different picture compared to the other measures. This result should perhaps come as no surprise since, in contrast to the other variables, it is based on perceptions and opinions.

Table 1 here

The complexities of interpreting performance measures is surely apparent. If, like the media rankings, the objective is to use the quantitative data to produce a composite index of overall

[10] Source: http://www.thecompleteuniversityguide.co.uk/league-tables/methodology/ accessed 29th June 2015. Note that this particular university guide is chosen purely for illustrative purposes; conclusions form any analysis presented here can be generalised across all university guides.
[11] An additional environmental ranking, derived from a 2014 report from the Higher Education Academy (Drayson *et al.*, 2014), is also provided for those interested in the 'green' credentials of universities.
[12] A rank correlation vale close to 1 represents a strong positive correlation while a value close to 0 indicates no relations between the variables.

performance then these disparate measures need to be combined into a meaningful aggregate index. This is not an insignificant problem.

Weights

The purpose of a single index derived from information across multiple indicators is to summarise information across dimensions and to provide an indicator that is easier to interpret than a set of many separate measures (Saltelli *et al.*, 2005). This means that all parts of society can use the indicators, from policy-makers to the general public, and this promotes accountability. Comparisons can be made not just amongst the group of entities being assessed but also over time (assuming calculation is consistent over time). Deriving a single index requires a set of weights in order to aggregate the separate measures.

The easiest approach is to apply an equal weighting across all indicators;[13] alternatively different weightings can be assigned. *The Complete University Guide* assigns weightings of between 0.5 and 1.5 to the 10 individual measures (table 2). Different publications use different weightings (and underlying indicators). Publishers of media rankings generally do not explain why they have chosen these weightings, or the fact that other weightings could be equally legitimate and potentially provide different rankings (Usher and Medow, 2009).

Table 2 here

Ideally weightings should reflect the preferences of the target audience, but deriving preferences for a group from the preferences of the individuals within that group is notoriously difficult. To construct a meaningful overall index is therefore fraught with difficulties. Is any purpose served by computing a composite index? The danger is that the apparent ease of interpretation provided by a composite index conceals the fact that the

[13] Performance indicators are usually standardised to produce a z-score before calculating an overall ranking. This ensures that the composite index is not affected by the units of measurement of the components underlying it.

picture of performance which it represents is misleading, particularly if a) it is unrepresentative of all the dimensions that it purports to cover and b) inappropriate weightings are used. This in turn can lead, for example, to inappropriate policy development or unsuitable choice of university by potential students.

To illustrate these points, consider again *The Complete University Guide* where rankings from the overall indicator are strongly correlated with those from all the separate indicators *with the exception of* those relating to student satisfaction and facilities spend (table 2). Stakeholders for whom these dimensions are of particular interest would appear to be poorly served by the overall ranking.

In the absence of information on the relative importance of each dimension (and hence of an appropriate weighting system), it might still be possible to reduce numerous indicators to a manageable number of dimensions using such techniques as principal components analysis, data envelopment analysis (in particular the 'benefit of the doubt' approach introduced by Cherchye *et al.* (2007)), the analytic hierarchy process, or co-plot (Johnes, 2015). Details and a critique of these techniques can be found elsewhere (Johnes, 2015; Saltelli *et al.*, 2005). We can illustrate the potential advantages of using one of these techniques (principal components analysis) in the context of *The Complete University Guide* data. The objective of a principal components analysis is to explain as much of the variation in the original data (the 10 dimensions in *The Complete University Guide*) with as few variables as possible. More details regarding principal components can be found in Saltelli *et al.* (2005).

The weightings for each of the 10 principal components calculated from *The Complete University Guide* data are displayed in table 3 along with the associated percentage variation accounted for by each principal component. If we use the Kaiser criterion (Saltelli *et al.*, 2005) to select the principal components which are adequate to represent the data, we are left with the first two principal components. The weights (table 3) of these two principal components suggest that the first principal component is mainly a combination of all dimensions *apart from* those reflecting student satisfaction and facilities spend, while the second principal components largely represents the combined dimensions of student satisfaction and facilities spend. This should come as no real surprise given the rank correlations already presented in table 1.

Table 3 here

A plot of the first two principal components is displayed in figure 1. Universities in the top right of the plot score highly on both principal components and therefore have good performance across all 10 dimensions. The converse is true of universities located to the bottom and left of the plot. The leading diagonal quadrants represent mixed performance. Numbers next to the plotted points in figure 1 are the rankings obtained in the composite measure (using the weightings as described above) of *The Complete University Guide*, with 1 representing top performance[14]. Both the top- and bottom-ranked universities are in the top right and bottom left quadrants of the principal components plot, suggesting that the composite ranking and the principal components provide a similar message. There are, however, many instances where the messages from the composite ranking and the principal components plot are mixed. For example, the university ranked 113th in *The Complete University Guide* appears in the bottom right hand quadrant of the scatter plot: while it performs badly on 8 of the indicators, it is amongst the top 5 universities on the basis of the second principal component (reflecting student satisfaction and facility spend). Conversely, the university ranked 13th is in the top left hand quadrant but is amongst the bottom 5 on the basis of the second principal component (reflecting student satisfaction and facility spend).

[14] Note that some rankings appear twice because of tied values.

Figure 1 here

The rank correlations between the first two principal components and the overall ranking (table 4) confirm that performance across all ten dimensions is adequately captured *neither* by the first principal component *nor* the composite ranking. A single indicator is insufficient to capture all the information contained in these 10 measures. This finding is in line with results of a similar analysis of university rankings from *The Guardian* and the *THES* (see HEFCE, 2008, Appendix C). The general message is that, in trying to give a simple overview of performance, composite indicators can be misleading. As they sacrifice information which may be of interest or policy relevance to users of the performance assessment little is gained from their construction.

Table 4 here

Methods for measuring performance: an economist's perspective

Let us now return to the basic idea underpinning performance indicators, namely as tools for government to assess the efficiency with which HEIs use publicly provided resources. The simple approach to efficiency measurement already discussed derives separate indicators which at best reflect reputation and resources but do not adequately capture the *efficiency of resource use*. Efficiency requires a knowledge of the outputs of universities, inputs going into those outputs, and the production relationship between them (Johnes and Taylor, 1990b). This invokes the idea of 'value added' or, from an economist's perspective, 'technical efficiency'.

HEIs are multi-product organisations and produce (in simple terms) teaching, research and third mission (the last reflecting universities' wider social engagement). Initial attempts to derive measures of value added applied Ordinary Least Squares (OLS) regression methods[15] to separate measures of universities' outputs (Johnes and Taylor, 1990b). At a time when proposed performance indicators included degree completions, classes of degrees, destinations of graduates and unit costs (Jarratt, 1985), such analyses proved useful in demonstrating that these suggested measures were affected by characteristics of the HEI. Thus much of the inter-university variation in unit costs was shown to be a consequence of subject mix, student mix and the ratio of students to staff (Johnes, 1990; Johnes and Taylor, 1990b). Similarly much of the inter-university variation in the percentage of graduates gaining employment was explained by subject mix, along with factors such as the academic ability of students on intake and location of the HEI (Johnes and Taylor, 1989a; Johnes and Taylor, 1989b; Johnes and Taylor, 1990b). The percentage of students gaining firsts and upper seconds and non-completion rates were also strongly related to the academic ability of students recruited by universities as well as factors such as library facilities, the percentage of students living in halls, and type of university (Johnes, 1997; Johnes and Taylor, 1989c; Johnes and Taylor, 1990a; Johnes and Taylor, 1990b; Johnes and Taylor, 1990c).

The problem with these analyses is that they separately examine the production of each output which raises the difficulty of interpreting multi-dimensional information (Johnes, 1996). Another problem is that the approach ignores synergies which surely exist in the university production process (Chizmar and McCarney, 1984; Chizmar and Zak, 1984). The reason that universities produce, for example, undergraduate and postgraduate teaching alongside research and interaction with business is that these are all joint products: there are

[15] OLS regression is a common method for estimating a linear relationship between observations of a variable Y (university output in this case) and an explanatory variable X (or explanatory variables $X_1, \dots, X_k$). In the case of one explanatory variable it therefore provides estimates of the coefficients (or parameters) a and b in the linear relationship $Y = a + bX$. OLS regression relies on certain underlying assumptions in order for the parameter estimates to have desirable statistical properties.

savings from producing these within one production unit rather than separately. A simple portrayal of the higher education production relationship is provided in figure 2.

Figure 2 here

The idea of linking resource allocation to performance is that universities which are efficient at transforming inputs into outputs should receive more resources than those which are inefficient. Performance indicators should ideally represent the efficiency with which universities transform inputs into outputs. We are interested, for example, in how much more output universities could produce from given inputs (known as an output-oriented approach), or how many fewer inputs could be used to produce given outputs (known as an input-oriented approach). Consider a visual presentation of the first of these questions. Let us assume that universities produce two outputs (say, graduates and research) from one input (say, staff). For the sector as a whole there will be a production possibility frontier (PPF), which represents the maximum outputs which can be achieved from given input with current technology (figure 3).

Figure 3 here

We can use this frontier as a benchmark against which the production of an individual university can be measured. In figure 3, the observed production point of university F (i.e. the combination of research/staff and teaching/staff) lies inside the PPF and so it is clearly less efficient than it could be. One way of measuring that inefficiency is to take a ray from the origin through point F, and projecting it on to the PPF at point F'. The technical efficiency (TE) of university F is then measured as the ratio $\mathrm{TE} = \mathrm{OF}/\mathrm{OF}'$.

The problem is how to estimate the PPF (figure 3). OLS regression is clearly an unsatisfactory approach to estimating a frontier because it estimates an average vector *through* the data rather than a *frontier*. Resulting efficiency ratios will therefore be calculated against an incorrectly estimated PPF. Furthermore, efficiency ratios derived from an OLS function are based on OLS regression errors.[16] Two frontier estimation techniques have been developed which overcome these problems, and which help in the construction of performance indicators: data envelopment analysis (DEA) and stochastic frontier analysis (SFA). We will briefly consider each of these approaches.

DEA (Banker, Charnes and Cooper, 1984; Charnes, Cooper and Rhodes, 1978; Charnes, Cooper and Rhodes, 1979) is a non-parametric frontier estimation technique[17] which can handle a production situation with both multiple outputs and multiple inputs, and does not require *a priori* specification of a functional form. It estimates using linear programming methods a piecewise linear PPF (see figure 4) which allows the performance of each institution to be measured *relative to institutions with similar missions or objectives*. This makes DEA attractive in the context of higher education where missions and objectives can differ substantially. An additional merit of DEA is that it provides benchmark information to those institutions which are performing inside the frontier. In figure 4, point F' represents a more efficient virtual production point to which university F should aspire. Since point F' is a linear combination of the outputs (relative to input) of universities B and C, target input and output levels can be derived, to which university F should aspire.

Figure 4 here

There are many examples of empirical studies which have applied DEA (and related non-parametric techniques) to measuring the efficiency of universities (Beasley, 1990; Beasley, 1995; Duh *et al.*, 2014; Fandel, 2007; Flegg and Allen, 2007; Flegg *et al.*, 2004; Giménez and Martínez, 2006; Glass *et al.*, 2006; Johnes, 2006; Johnes, 2008; Johnes, 2014b; Thanassoulis

[16] An underlying assumption of OLS regression is that the errors are randomly distributed, hence estimated efficiencies are highly unsatisfactory (Johnes, 1996).
[17] Non-parametric estimation means that the technique does not rely on any underlying assumptions such as the data being distributed in a certain way.

et al., 2011). DEA studies of UK universities tend to find that average efficiency across the sector as a whole is fairly high (typically 80% or above), but the specific findings depend on sample used (the more restricted the sample the higher the average efficiency) and the time period covered. DEA has also been used to assess the efficiency of individual academic departments or programmes within an institution (Casu, Shaw and Thanassoulis, 2005; Colbert, Levary and Shaner, 2000; Kao and Hung, 2008; Kao and Liu, 2000; Moreno and Tadepalli, 2002; Ray and Jeon, 2008), central administration or services across universities (Casu and Thanassoulis, 2006; Simon, Simon and Arias, 2011), and to make efficiency comparisons across different national education systems (Giménez, Prior and Thieme, 2007).

DEA is a deterministic non-parametric approach, with the disadvantages that random fluctuations in the data are not allowed for, there are no conventional tests of significance or methods for drawing inference, and efficiency estimates are particularly affected by sample size. This means that great care should be taken in choosing the variables to represent the inputs and outputs in any DEA model; the model specification should be consistent with the higher education production process. In addition DEA has not been extended to address specific issues of modelling in a panel data context. In its favour, recent developments in DEA include the incorporation of bootstrapping techniques[18] to produce confidence intervals and bias-corrected estimates of efficiency, and the development of hypothesis tests to assess the significance of specific inputs and/or outputs (Banker, 1996; Johnes, 2006; Pastor, Ruiz and Sirvent, 2002).

Stochastic frontier analysis (SFA) is a parametric frontier estimation method, which allows for stochastic errors in the data (figure 5) and provides parameter estimates and associated

[18] In statistics bootstrapping is a method which relies on random sampling (with replacement) of the original data in order to estimate a sampling distribution of a required statistic. In the context of DEA, bootstrapping techniques can, for example, generate a sampling distribution for the efficiency score which then allows estimation of a confidence interval around the score. See Johnes (2004b) for more detail.

significance tests (Aigner, Lovell and Schmidt, 1977; Meeusen and van den Broeck, 1977). Following Jondrow *et al.* (1982) SFA also allows the estimation of technical efficiency for each university. These features of SFA make it an attractive methodological tool which has frequently been used particularly in studies relating to policy development (Abbott and Doucouliagos, 2009; Johnes, Johnes and Thanassoulis, 2008; Johnes and Schwarzenberger, 2011; Johnes, 2014b; McMillan and Chan, 2006; Stevens, 2005).

Figure 5 here

Many of the SFA empirical studies relate to the context of university costs where SFA lends itself to the framework of a single left-hand side variable (costs) and multiple right-hand side variables. Average efficiency levels for the English higher education sector are estimated to be around 70% and there is considerable variation around this mean (Johnes, Johnes and Thanassoulis, 2008; Johnes *et al.*, 2005). The parameters of the SFA cost function can be used to estimate the presence of economies of scale (reduction in cost per unit from increasing output) and economies of scope (reduction in costs from producing two or more outputs jointly); recent evidence reveals that typically economies of scale are exhausted and that there are diseconomies of scope in English higher education (Izadi *et al.*, 2002; Johnes and Johnes, 2009; Johnes, Johnes and Thanassoulis, 2008; Johnes *et al.*, 2005). This suggests that, for the HEI of average size, there are no further opportunities for economies of scale from expansion in size; moreover, the existence of diseconomies of scope suggests that the opportunities for economies from sharing resources across different outputs have been exhausted (indeed HEIs may already be producing too many outputs simultaneously).

An exception to these SFA empirical studies of costs is one by Johnes (2014b) who examines technical efficiency in a multi-input, multi-output framework. Average efficiency levels for the English higher education sector are estimated to be around 70% to 80%, and the estimated parameters provide insights into potential input substitutability: there is least scope for substituting between academic staff and administrative inputs whilst academic staff and capital are the inputs with the greatest potential for substitution. Perhaps of greater interest (particularly from a policy viewpoint) is that universities which ultimately merge typically have greater flexibility in terms of input substitution than those which do not (Johnes, 2014a).

The downside of the SFA approach is that the assumptions which underpin it (regarding the distribution of efficiencies and the stochastic error, as well as the functional form of the function being estimated) are often made for ease of analysis; results can be biased as a consequence. SFA is not a benchmarking tool and provides no precise information on how managers of an institution can alter inputs and/or outputs to improve their efficiency.

In comparisons of DEA and SFA applied to higher education, SFA generally provides lower efficiency estimates than DEA, and rank correlations of efficiencies derived from the two methods are positive but very low (Johnes, 2014b; Kempkes and Pohl, 2010). Policy-makers and users of performance indicators should be aware that the relative position of universities is not consistent across methodological approaches.

Potential effects of performance indicators and rankings

We consider in this section the effects on subsequent efficiency of measuring performance using the simple and frontier approaches.

Measuring performance using individual indicators or a composite index

The problem with the simple approach stems directly from what might seem its main advantage: transparency – of the method and the data used. While this makes it easy for a HEI to see its strengths and weaknesses and alter behaviour accordingly, it also means that rankings are open to manipulation and gaming. According to Goodhart's "law", a variable which is used to measure performance is open to manipulation by those whose performance is being measured (Johnes, 1992; Pollard *et al.*, 2013). Changing behaviour is a desirable consequence of performance measurement *only* if the changed behaviour genuinely improves *performance* rather than simply *rank*.

Rankings are important to individual institutions: national and global rankings can be used by other institutions to identify suitable collaborative partners; they can be used by students to inform their choice of university; by prospective academic employees seeking new posts ; and by employers for recruitment (Hazelkorn, 2015; Saisana, d'Hombres and Saltelli, 2011). This means that a university has an incentive to change its behaviour in response to the rankings; but changed behaviour may not benefit performance.

Many popular measures of performance are under the control of the HEI. Graduation rates, for example, can be improved by more effective teaching delivery – the desired effects of university performance assessment – or by lowering standards (so-called 'grade inflation'), on which there is mixed evidence in the UK and USA (Bachan, 2015; Johnes, 2004a; Johnes and Soo, 2015; Popov and Bernhardt, 2013). More generally, there is confirmation of concern from senior managers of universities that some measures in league tables are vulnerable to 'cheating' behaviour (Rolfe, 2003), and evidence that universities are manipulating, or influencing data in order to raise their rankings (Hazelkorn, 2015). There have been claims, for example, that students have been pressured to provide favourable responses to the National Student Survey in the UK (Newman, 2008). Gaming behaviour by universities is unlikely to achieve the efficiency objective of performance assessment: 'The pernicious effect of this competitive pursuit of academic prestige is that it is a highly costly, zero-sum game, in which most institutions as well as society will be the losers, and which diverts resources as well as administrative and faculty attention away from the collective actions within universities necessary to actually improve student learning …' (Dill, 2009, p6).

Gaming behaviour can mislead those using university rankings. National policy towards higher education, for example, can be based on fallacious information: a policy of merging HEIs is being rolled out in France, Russia and China in the belief that global rankings of domestic HEIs can be favourably affected (Shin and Toutkoushian, 2011). Merger policy has also been promoted in the UK in the belief that greater size leads to greater visibility in the world rankings as well as greater efficiency (Jump, 2014). Given the problems with rankings, however, much more research into the wider likely effects of any policy initiative should be undertaken; in the case of merger policy, there has been little statistical research into the benefits of merging universities and this is an area which should be explored further (Johnes, 2014b).

Even if gaming behaviour is not a serious problem, by focusing on improvement of the components which underpin the media rankings, HEIs are in danger of becoming much more homogeneous. For example, the underlying components of the rankings are often biased towards research activity, particularly to research in the sciences (Dill, 2009), and this could lead to a HEI altering its mission to scientific research activity even though it might formerly have pursued teaching excellence (Shin and Toutkoushian, 2011). In addition, the highly-ranked elite universities become the benchmarks for the lower-ranked HEIs to mimic, thereby ensuring a reduction in diversity between universities (Morphew and Swanson, 2011). Yet diversity in higher education is desirable because it stimulates a dynamic sector giving more choice to students (HEFCE, 2012), and reducing this choice might be socially undesirable because of the negative impact on student access caused by imperfect geographical mobility (De Fraja and Valbonesi, 2012; Kelchtermans and Verboven, 2010).

Finally university rankings suggest a precision which is unlikely to be supported by detailed examination of the data: the methodology is such that differences in rankings, which can appear large, conceal the fact that there are only very small differences in the scores from which the rankings have been derived (Longden, 2011).[19] It is therefore important to know whether or not the differences in rankings between HEIs are 'real' or significant in a statistical sense. Little work has focused on this aspect in the context of individual measures or composite indexes. An exception is work by Smith, McKnight and Naylor (2000) who examine the performance of HEIs in the UK on the basis of the first destinations of graduates, and find that the differences in performance are significantly different only for the top 10 and the bottom 10 institutions. Thus, rankings not only leave the higher education sector open to undesirable behaviour and consequences, but these effects may be based on rankings that have little meaning. This point is pursued further below.

Performance measures based on frontier estimation techniques

Deriving efficiency scores using DEA or SFA involves complex procedures such that the end-user loses direct engagement with the data. The advantage is that it becomes more difficult to alter behaviour merely to affect a *position* in the rankings, so these approaches are less likely to incite gaming behaviour. However, the availability of managerial information on benchmarks which inefficient universities should use to improve their performance means that HEIs still have the opportunity genuinely to improve their *efficiency*.

This might lead one to suppose that rankings based on frontier estimation techniques are more reliable than the simple rankings based on aggregating individual measures. There are two *caveats* to this. First the production relationship in higher education is extremely difficult to model, and incorporating quantity and quality of all aspects of a university's activities can challenge the estimation methods. Second, as with the simple approach, point estimates of each university's relative position are highly suspect. The estimation of confidence intervals

[19] Related to this is the problem of the volatility of rankings over time which might be due to changing methodology or might have alternative explanations. This idea is explored elsewhere (Longden, 2011, pp96-99).

around efficiency estimates derived from both SFA and DEA suggests that there is considerable overlap in performance between many universities (Johnes, 2014b).

Groupings rather than point estimates

Point estimates of rankings are misleading since there is likely no significant difference in efficiency between many universities, and possibly even damaging if they result in undesirable gaming behaviour. How can we gain some idea of the performance of universities whilst avoiding these problems? An idea being mooted is that performance *groupings*, rather than point estimates, would be more appropriate (Bougnol and Dula, 2006).

This raises the question of how to construct the groups. One suggestion by Barr, Durchholz and Seiford (2000) is to use DEA to produce tiers of universities (known as ‘peeling the DEA onion’): the first application of DEA to the data produces a set of efficient universities which are removed to form the top tier. DEA is then applied to the truncated data set, and the efficient universities removed to form the second tier. This process, or ‘peeling’, continues until all universities are assigned to a tier (figure 6).

Figure 6 here

The advantage of DEA is that it can be applied in either context: producing a composite index from a number of performance indicators or the production approach relating inputs to outputs. In the former case, the ‘benefit of the doubt’ version of DEA is used whereby the indicators are classified as ‘outputs’ and the single ‘input’ is equal to one for all universities (Cherchye *et al.*, 2007). This therefore constructs a composite index using objectively derived weightings which differ for each university and are constructed to show each HEI in its best possible light. In practice this means that each university is measured against universities with similar mission or objectives, and hence diversity in the sector is preserved.

We illustrate the peeling method using the data from *The Complete University Guide* 2015-16 and applying the DEA ‘benefit of the doubt’ model to the data. The peeling approach yields 4 groupings of universities (shown in table 5; universities are ordered alphabetically within group). Alongside the HEIs in each tier are the ranking assigned in *The Complete University Guide* 2015-16. The average of the rankings of the universities in each tier are very broadly in line with the tiers produced in that the average for the first tier is the lowest, and so on. But there are some big differences between ranking and tier for some universities. This might arise because of the calculation of the weightings in DEA; but it is possible to restrict the weighting assigned to each measure if this were desired.

Table 5 here

This analysis is offered purely as an example of how a tiered approach to performance assessment might work in practice; alternative approaches should be explored and evaluated. It seems, however, that a move away from specific rankings can only have beneficial effects on the performance of the higher education sector.

Conclusions

This chapter summarises approaches to performance assessments and rankings of universities particularly over the last 30 to 40 years. Various approaches – ranging from individual indicators through composite indexes to the technical efficiency approach permitted by recent developments in frontier estimation techniques – have been presented and compared. The chapter ends with the suggestion that the potentially pernicious and self-seeking effects of the commonly-applied simple approach to performance measurement could be reduced by adopting a tiered performance approach, using frontier estimation, to produce groupings rather than specific rankings. Throughout, the possible approaches have been illustrated using data from recent media rankings in the UK. It is apparent that different approaches deliver different conclusions and the user of performance indicators and rankings should beware: university rankings should come with a serious health warning and be handled with care. Indeed, the words of Gareth Williams from more than 20 years ago continue to be relevant today: ‘Like all quantitative performance indicators these figures raise more questions than they answer’ (Williams, 1992a, p147).

Table 1: Rank correlations of 10 indicators from *The Complete University Guide*

	1	2	3	4	5	6	7	8	9
1. Entry standards									
2. Student satisfaction	0.17								
3. Research assessment	0.84*	0.14							
4. Research intensity	0.73*	0.27*	0.73*						
5. Graduate prospects	0.79*	0.29*	0.67*	0.61*					
6. Staff-student ratio	0.73*	0.22*	0.77*	0.69*	0.59*				
7. Academic services spend	0.57*	0.11	0.60*	0.54*	0.53*	0.62*			
8. Facilities spend	0.21*	0.11	0.27*	0.27*	0.23*	0.21*	0.29*		
9. Good honours	0.86*	0.23*	0.78*	0.76*	0.75*	0.68*	0.54*	0.15	
10. Degree completion	0.76*	0.33*	0.68*	0.62*	0.77*	0.63*	0.48*	0.27*	0.75*

Notes: * = significant at the 5% significance level. Note that *The Complete University Guide* uses student-staff ratio (indictor 6) and this has been reversed for the purposes of the correlation table to ensure that a higher value is consistent with more favourable performance.

Table 2: Weightings used to produce an overall performance indicator in *The Complete University Guide* and rank correlation between the overall ranking and its components

	Weight	Correlation
1. Entry standards	1.0	0.91
2. Student satisfaction	1.5	0.35
3. Research assessment	1.0	0.86
4. Research intensity	0.5	0.78
5. Graduate prospects	1.0	0.83
6. Staff-student ratio	1.0	0.82
7. Academic services spend	0.5	0.64
8. Facilities spend	0.5	0.34
9. Good honours	1.0	0.89
10. Degree completion	1.0	0.85

Source: http://www.thecompleteuniversityguide.co.uk/league-tables/methodology/ accessed 29th June 2015

Table 3: Weightings for the 10 principal components (PC) associated with *The Complete University Guide* data

	Principal components									
***The Complete University Guide* dimensions**	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
1. Entry standards	0.39	-0.05	0.01	-0.07	-0.11	-0.09	-0.18	0.28	0.24	-0.81
2. Student satisfaction	0.13	0.65	-0.59	0.37	0.25	0.02	0.03	0.07	0.07	-0.02
3. Research assessment	0.35	-0.09	-0.00	-0.28	0.40	-0.40	0.66	0.14	-0.07	0.06
4. Research intensity	0.36	0.07	0.04	-0.21	0.35	-0.13	-0.43	-0.68	-0.20	-0.04
5. Graduate prospects	0.36	0.06	-0.06	0.10	-0.47	-0.39	-0.22	0.28	-0.54	0.25
6. Staff-student ratio	0.35	-0.21	0.07	0.07	0.37	0.66	-0.13	0.37	-0.29	0.12
7. Academic services spend	0.27	-0.35	0.19	0.80	0.02	-0.13	0.13	-0.24	0.19	0.05
8. Facilities spend	0.07	0.61	0.77	0.07	0.05	-0.00	0.02	0.11	0.07	0.05
9. Good honours	0.38	-0.02	-0.09	-0.24	-0.13	0.01	-0.22	0.10	0.68	0.50
10. Degree completion	0.34	0.16	-0.05	-0.11	-0.51	0.45	0.46	-0.39	-0.07	-0.08
% variation	58.1	10.8	9.4	6.1	5.2	2.9	2.5	2.0	1.9	1.1

Table 4: Rank correlations between the first two principal components and the university ranking

	1	**2**
1. University ranking		
2. Principal component 1	0.98	1.00
3. Principal component 2	0.15	0.08

Note: Data sourced from *The Complete University Guide* 2015-16

Table 5: Groupings of universities produced by the peeling approach applied to data from *The Complete University Guide* 2015-16

Tier 1 **(Average ranking 42.15)**	**Rank**	***Tier 2*** **(Average ranking 54.85)**	**Rank**	***Tier 3*** **(Average ranking 69.53)**	**Rank**	***Tier 4*** **(Average ranking 98.71)**	**Rank**
Bath	11	Abertay	95	Aberdeen	40	Bedfordshire	110
Birmingham	18	Aberystwyth	86	Anglia Ruskin	115	Bolton	121
Bishop Grosseteste	117	Aston	32	Arts University Bournemouth	57	Brighton	76
Bristol	15	Bangor	58	Bournemouth	54	Canterbury Christ Church	106
Brunel University London	49	Bath Spa	70	Central Lancashire	91	Derby	94
Buckinghamshire New	113	Birmingham City	88	Chester	93	London Metropolitan	126
Cambridge	1	Bradford	63	Chichester	77	London South Bank	119
Cardiff	31	City	41	De Montfort	54	Manchester Metropolitan	73
Cardiff Metropolitan	79	Cumbria	111	Dundee	42	Newman	120
Coventry	48	East Anglia	16	Edge Hill	82	Northumbria	60
Durham	5	Falmouth	70	Edinburgh	20	Plymouth	90
East London	124	Glasgow	30	Edinburgh Napier	92	Portsmouth	59
Essex	34	Glasgow Caledonian	83	Gloucestershire	80	Salford	96
Exeter	10	Glyndwr	123	Greenwich	107	Southampton Solent	122
Heriot-Watt	37	Goldsmiths, University of London	50	Huddersfield	74	Staffordshire	103
Imperial College London	4	Harper Adams	60	Hull	63	Teesside	98
Loughborough	11	Hertfordshire	75	Kingston	104	Worcester	105
Middlesex	89	Keele	46	Leeds Beckett	114		
Northampton	83	Kent	22	Lincoln	51		
Oxford	2	King's College London	23	Liverpool	39		
Queen's, Belfast	36	Lancaster	9	Manchester	28		
Royal Agricultural University	85	Leeds	19	Nottingham	25		
St Andrews	5	Leeds Trinity	101	Nottingham Trent	53		
St George's, University of London	43	Leicester	24	Oxford Brookes	54		
Strathclyde	38	Liverpool Hope	97	Queen Margaret	78		
Surrey	8	Liverpool John Moores	68	Queen Mary, University of London	33		
		London School of Economics	3	Reading	29		
		Newcastle	26	Royal Holloway, University of London	34		
		Robert Gordon	63	Sheffield Hallam	72		
		Roehampton	66	SOAS, University of London	43		
		Sheffield	27	South Wales	102		
		Southampton	14	St Mary's, Twickenham	109		
		Sussex	21	Stirling	47		
		Trinity Saint David	125	Sunderland	116		
		Ulster	66	Swansea	45		
		University College London	13	University for the Creative Arts	52		
		University of the Arts, London	80	West London	108		
		Warwick	7	West of Scotland	118		
		West of England, Bristol	62	Westminster	100		
		York	17	Winchester	86		

Figure 1: Plot of first two principal components

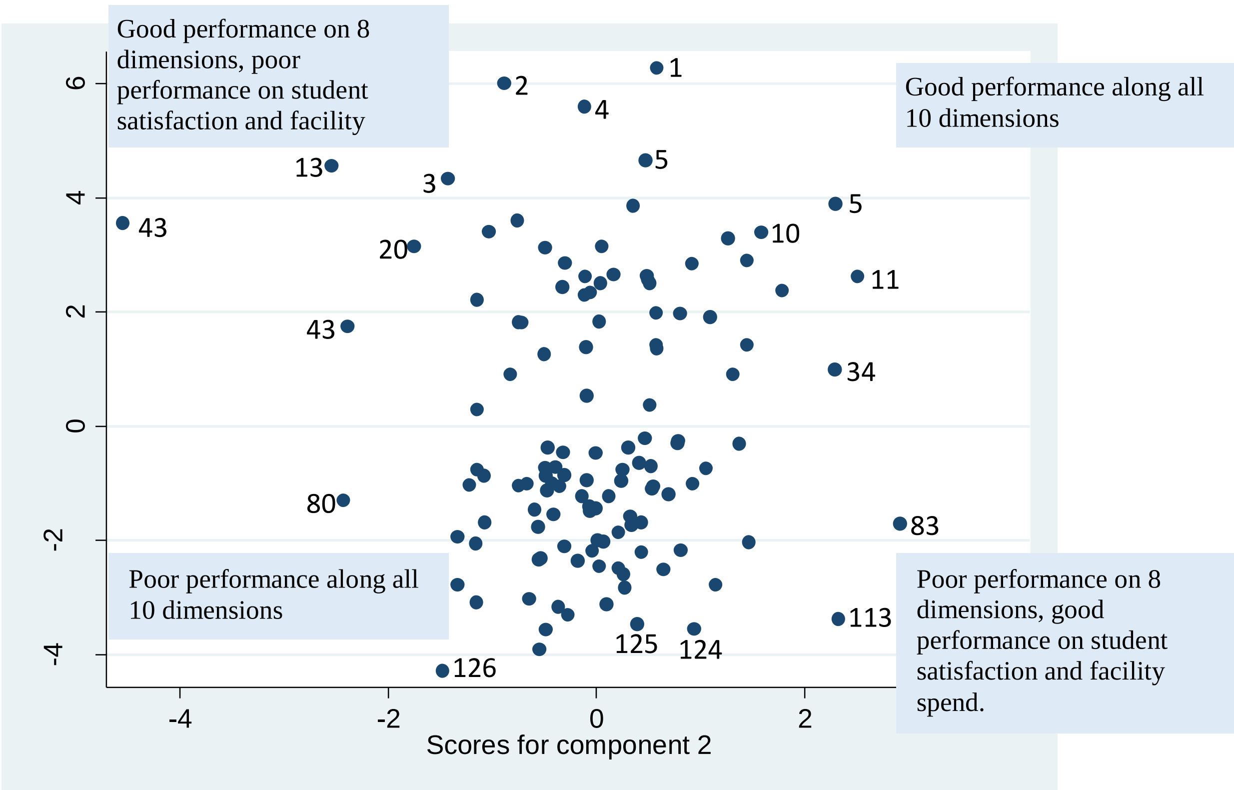
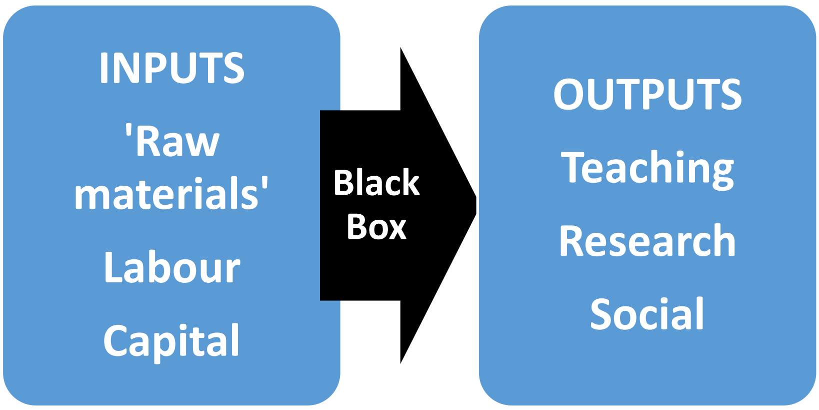


Figure 2: A production framework for universities



4

Figure 3: Measuring technical efficiency of a university

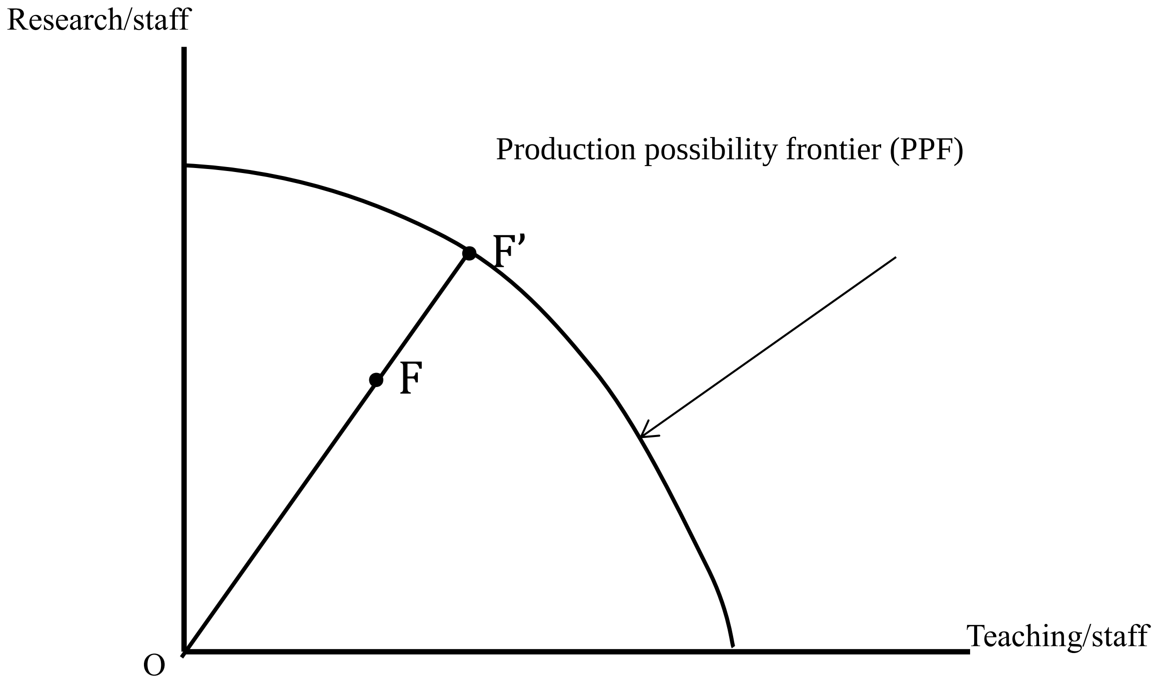


Figure 4: The PPF estimated by DEA

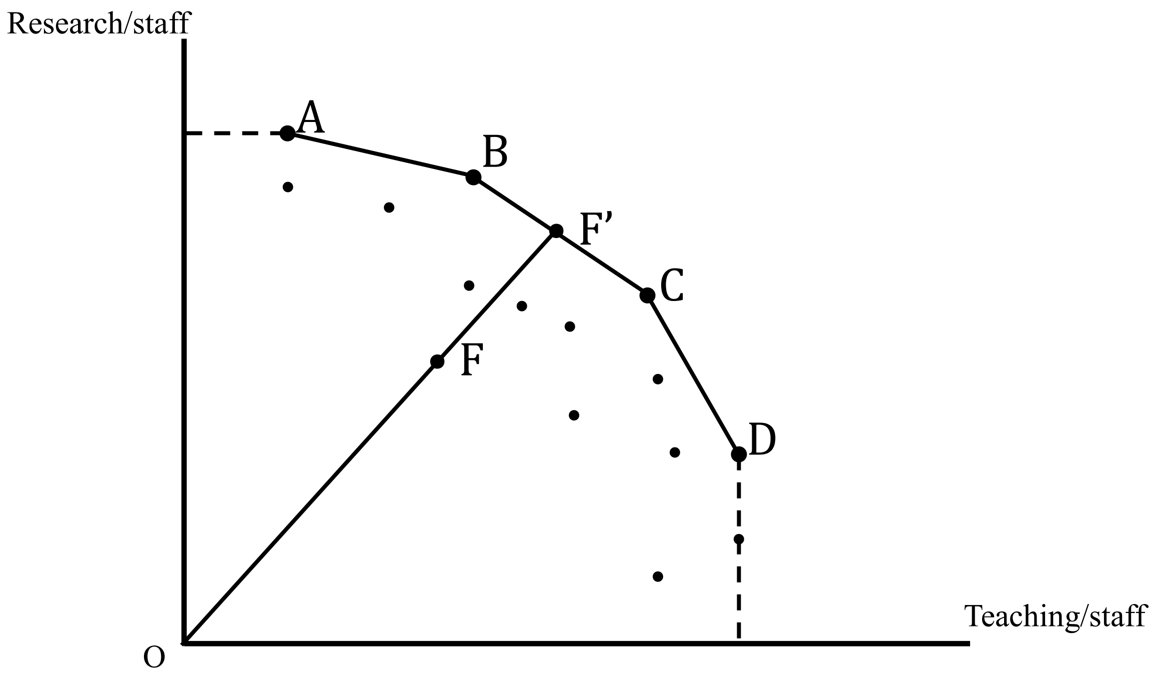


Figure 5: The PPF estimated by SFA

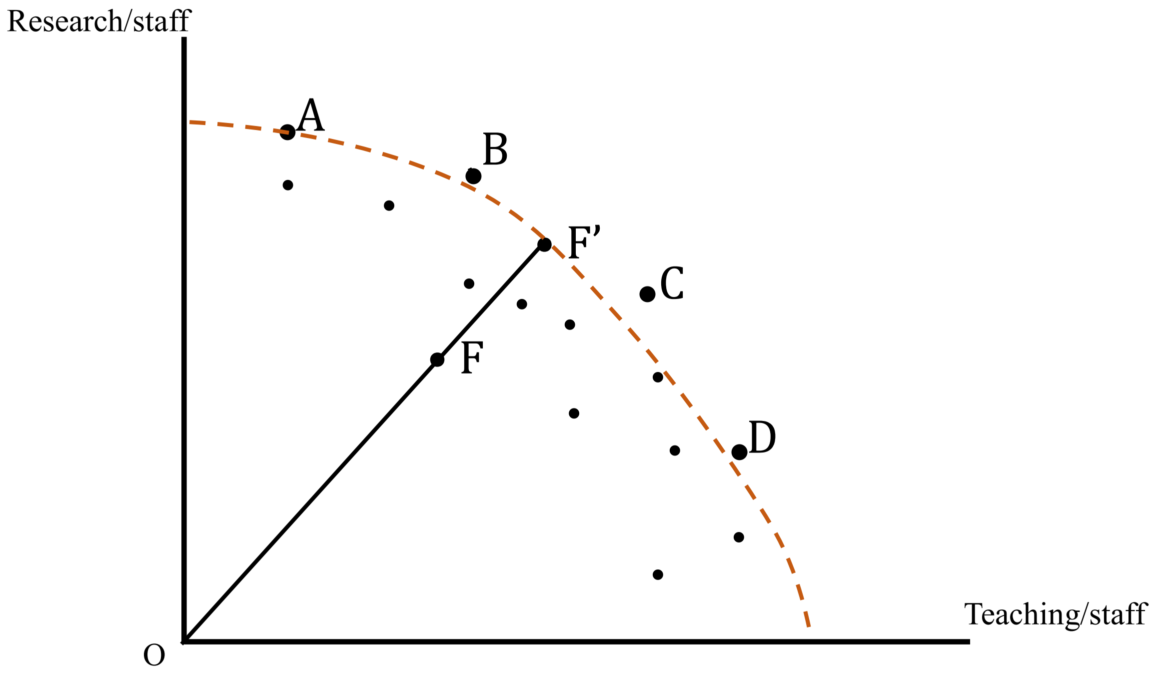
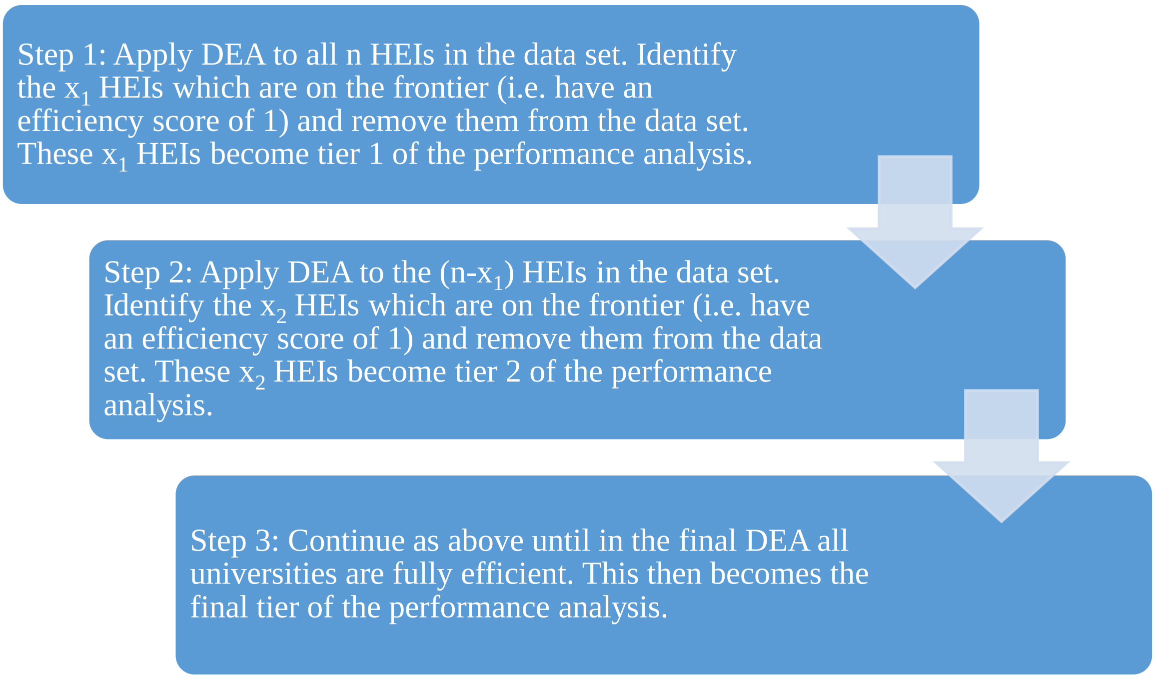


Figure 6: Peeling the DEA onion



References

Abbott, M. and Doucouliagos, C. (2009). 'Competition and efficiency: Overseas students and technical efficiency in Australian and New Zealand universities'. *Education Economics,* 17 (1), 31-57.

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

Bachan, R. (2015). 'Grade inflation in UK higher education'. *Studies in Higher Education,* DOI: 10.1080/03075079.2015.1019450.

Banker, R. D. (1996). 'Hypothesis tests using data envelopment analysis'. *Journal of Productivity Analysis,* 7 (2-3), 139-158.

Banker, R. D., Charnes, A. and Cooper, W. W. (1984). 'Some models for estimating technical and scale inefficiencies in data envelopment analysis'. *Management Science,* 30 (9), 1078-1092.

Barr, R. S., Durchholz, M. L. and Seiford, L. (2000). *Peeling the DEA onion: Layering and rank-ordering DMUs using tiered DEA*. Southern Methodist University Technical Report: Dallas, Texas, USA.

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

Beasley, J. E. (1995). 'Determining teaching and research efficiencies'. *Journal of the Operational Research Society,* 46 (4), 441-452.

Bougnol, M.-L. and Dula, J. H. (2006). 'Validating DEA as a ranking tool: An application of DEA to assess performance in higher education '. *Annals of Operations Research,* 145 (1), 339-365.

Casu, B., Shaw, D. and Thanassoulis, E. (2005). 'Using a group support system to aid input-output identification in DEA'. *Journal of the Operational Research Society,* 56 (12), 1363-1372.

Casu, B. and Thanassoulis, E. (2006). 'Evaluating cost efficiency in central administrative services in UK universities'. *Omega,* 34 (5), 417-426.

Charnes, A., Cooper, W. W. and Rhodes, E. (1978). 'Measuring the efficiency of decision making units'. *European Journal of Operational Research,* 2 (4), 429-444.

Charnes, A., Cooper, W. W. and Rhodes, E. (1979). 'Measuring the efficiency of decision making units: a short communication'. *European Journal of Operational Research,* 3 (4), 339.

Cherchye, L., Moesen, W., Rogge, N. and Van Puyenbroeck, T. (2007). 'An introduction to 'benefit of the doubt' composite indicators'. *Social Indicators Research,* 82 (1), 111-145.

Chizmar, J. F. and McCarney, B. J. (1984). 'An evaluation of 'trade-offs' implementation using canonical estimation of joint educational production functions'. *Journal of Economic Education,* 15 (1), 11-20.

Chizmar, J. F. and Zak, T. A. (1984). 'Canonical estimation of joint educational production functions'. *Economics of Education Review,* 3 (1), 37-43.

Colbert, A., Levary, R. R. and Shaner, M. C. (2000). 'Determining the relative efficiency of MBA programs using DEA'. *European Journal of Operational Research,* 125 (3), 656-669.

De Fraja, G. and Valbonesi, P. (2012). 'The design of the university system'. *Journal of Public Economics,* 96 (3-4), 317-330.

Department of Education and Science (1985). *The Development of Higher Education into the 1990s* (CMND 9524). London: H.M.S.O.

Dill, D. D. (2009). 'Convergence and diversity: The role and influence of university ranking'. In B. M. Kehm and B. Stensaker (Eds), *University Rankings, Diversity, and the new Landscape of Higher Education* (pp. 97-116). Rotterdam: Sense Publishers.

Dixon, M. (February 28th, 1976). 'Careers: More means better'. *The Financial Times*

Dixon, M. (November 14th, 1985). 'Jobs column: What happened to universities' graduates?'. *The Financial Times*
Dixon, M. (April 19th, 1989). 'Jobs column: Benefits, and risks, of trying for a degree'. *The Financial Times*
Drayson, R., Bone, E., Agombar, J. and Kemp, S. (2014). *Student attitudies towards and skills for sustainable devlopment*. York: Higher Education Academy.
Duh, R.-R., Chen, K.-T., Lin, R.-C. and Kuo, L.-C. (2014). 'Do internal controls improve operating efficiency of universities?'. *Annals of Operations Research,* 221 (1), 173-195.
Fandel, G. (2007). 'On the performance of universities in North Rhine-Westphalia, Germany: Government's redistribution of funds judged using DEA efficiency measures'. *European Journal of Operational Research,* 176 (1), 521-533.
Flegg, T. and Allen, D. (2007). 'Does expansion cause congestion? The case of the older British universities, 1994-2004'. *Education Economics,* 15 (1), 75-102.
Flegg, T., Allen, D., Field, K. and Thurlow, T. W. (2004). 'Measuring the efficiency of British universities: a multi-period data envelopment analysis'. *Education Economics,* 12 (3), 231-249.
Giménez, V. M. and Martínez, J. L. (2006). 'Cost efficiency in the university: A departmental evaluation model'. *Economics of Education Review,* 25, 543-553.
Giménez, V. M., Prior, D. and Thieme, C. (2007). 'Technical efficiency, managerial efficiency and objective-setting in the educational system: An international comparison'. *Journal of the Operational Research Society,* 58 (8), 996-1007.
Glass, J. C., McCallion, G., McKillop, D. G., Rasaratnam, S. and Stringer, K. S. (2006). 'Implications of variant efficiency measures for policy evaluation in UK higher education'. *Socio-Economic Planning Sciences,* 40, 119-142.
Harman, G. (2011). 'Competitors of rankings: New directions in quality assurance and accountability'. In J. C. Shin, R. K. Toutkoushian and U. Teichler (Eds), *University Rankings: Theoretical Basis, Methodology and Impacts on Global Higher Education* (pp. 35-54). Dordrecht, The Netherlands: Springer.
Hazelkorn, E. (2015). 'How the geo-politics of rankings is shaping behaviour'. *Higher Education in Russia and Beyond,* 2 (4), 6-7.
HEFCE (2008). *Counting what is measured or measuring what counts? League tables and their impact on higher education institutions in England*. Bristol: Higher Education Funding Council for England.
HEFCE. (2012). Collaborations, alliances and mergers in higher education: Consultation on lessons learned and guidance for institutions. London: Higher Education Funding Council for England.
Izadi, H., Johnes, G., Oskrochi, R. and Crouchley, R. (2002). 'Stochastic frontier estimation of a CES cost function: the case of higher education in Britain'. *Economics of Education Review,* 21 (1), 63-71.
Jarratt, A. (1985). *Report of the Steering Committee for Efficiency Studies in Universities*. London Committee of Vice-Chancellors and Principals.
Johnes, G. (1992). 'Performance indicators in higher education: A survey of recent work'. *Oxford Review of Economic Policy,* 8 (2), 19-34.
Johnes, G. (2004a). 'Standards and grade inflation'. In G. Johnes and J. Johnes (Eds), *International Handbook on the Economics of Education* (pp. 462-483). Cheltenham: Edward Elgar.
Johnes, G. and Johnes, J. (2009). 'Higher education institutions' costs and efficiency: taking the decomposition a further step'. *Economics of Education Review,* 28 (1), 107-113.
Johnes, G., Johnes, J. and Thanassoulis, E. (2008). 'An analysis of costs in institutions of higher education in England'. *Studies in Higher Education,* 33 (5), 527-549.

Johnes, G., Johnes, J., Thanassoulis, E., Lenton, P. and Emrouznejad, A. (2005). *An Exploratory Analysis of the Cost Structure of Higher Education in England*. (Vol. Research Report 641). London: Department for Education and Skills.

Johnes, G. and Schwarzenberger, A. (2011). 'Differences in cost structure and the evaluation fo efficiency: the case of German universities'. *Education Economics*, 19 (5), 487-499.

Johnes, G. and Soo, K. T. (2015). 'Grades across unievrsities over time'. *The Manchester School*, DOI: 10.1111/manc.12138.

Johnes, J. (1990). 'Unit costs: some explanations of the differences between UK universities'. *Applied Economics*, 22 (7), 853-862.

Johnes, J. (1996). 'Performance assessment in higher education in Britain'. *European Journal of Operational Research*, 89, 18-33.

Johnes, J. (1997). 'Inter-university variations in undergraduate non-completion rates: a statistical analysis by subject of study'. *Journal of Applied Statistics*, 24 (3), 343-361.

Johnes, J. (2004b). 'Efficiency measurement'. In G. Johnes and J. Johnes (Eds), *International Handbook on the Economics of Education* (pp. 613-742). Cheltenham: Edward Elgar.

Johnes, J. (2006). 'Data envelopment analysis and its application to the measurement of efficiency in higher education'. *Economics of Education Review*, 25 (3), 273-288.

Johnes, J. (2008). 'Efficiency and productivity change in the English higher education sector from 1996/97 to 2004/05'. *The Manchester School*, 76 (6), 653-674.

Johnes, J. (2014a). Efficiency and input substitutability in English higher education 1996/97 to 2008/09, *INFORMS*. San Francisco, USA.

Johnes, J. (2014b). 'Efficiency and mergers in English higher education 1996/97 to 2008/9: Parametric and non-parametric estimation of the multi-input multi-output distance function'. *The Manchester School*, 82 (4), 465-487.

Johnes, J. (2015). 'Operational Research in education'. *European Journal of Operational Research*, 243 (3), 683-696.

Johnes, J. and Taylor, J. (1989a). 'An evaluation of performance indicators based upon the first destination of university graduates'. *Studies in Higher Education*, 14 (2), 219-235.

Johnes, J. and Taylor, J. (1989b). 'The first destination of new graduates: comparisons between universities'. *Applied Economics*, 21 (3), 357-373.

Johnes, J. and Taylor, J. (1989c). 'Undergraduate non-completion rates: difference between UK universities'. *Higher Education*, 18 (2), 209-225.

Johnes, J. and Taylor, J. (1990a). 'Determinants of student wastage in higher education'. *Studies in Higher Education*, 15 (1), 87-99.

Johnes, J. and Taylor, J. (1990b). *Performance Indicators in Higher Education*. Buckingham: Society for Rresearch into Higher Education and Open University Press.

Johnes, J. and Taylor, J. (1990c). 'Undergraduate non-completion rates: a reply'. *Higher Education*, 19 (3), 385-390.

Jondrow, J., Lovell, C. A. K., Materov, I. S. and Schmidt, P. (1982). 'On the estimation of technical inefficiency in the stochastic frontier production function model'. *Journal of Econometrics*, 19 (2-3), 233-238.

Jump, P. (2014). 'Cut 50% of universities and bar undergarduates from Oxbridge'. *Times Higher Education*, 25th June 2014, 9.

Kao, C. and Hung, H.-T. (2008). 'Efficiency analysis of university departments: An empirical study'. *Omega*, 36 (4), 653-664.

Kao, C. and Liu, S.-T. (2000). 'Data envelopment analysis with missing data: An application to university libraries in Taiwan'. *Journal of the Operational Research Society*, 51 (8), 897-905.

Kelchtermans, S. and Verboven, F. (2010). 'Program duplication in higher education is not necessarily bad'. *Journal of Public Economics*, 94 (5-6), 397-409.

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

Longden, B. (2011). 'Ranking indicators and weights'. In J. C. Shin, R. K. Toutkoushian and U. Teichler (Eds), *University Rankings: Theoretical Basis, Methodology and Impacts on Global Higher Education* (pp. 73-104). Dordrecht, The Netherlands: Springer.

McMillan, M. L. and Chan, W. H. (2006). 'University efficiency: A comparison and consolidation of results from stochastic and non-stochastic methods'. *Education Economics,* 14 (1), 1-30.

Meeusen, W. and van den Broeck, J. (1977). 'Efficiency estimation from Cobb-Douglas production functions with composed error'. *International Economic Review,* 18 (2), 435-444.

Moreno, A. A. and Tadepalli, R. (2002). 'Assessing academic department efficiency at a public university'. *Managerial and Decision Economics*, 23 (7), 385-397.

Morphew, C. C. and Swanson, C. (2011). 'On the efficacy of raising your university's ranking'. In J. C. Shin, R. K. Toutkoushian and U. Teichler (Eds), *University Rankings: Theoretical Basis, Methodology and Impacts on Global Higher Education* (pp. 185-200). Dordrecht, The Netherlands: Springer.

Newman, M. (2008). 'Students urged to inflate national survey marks to improve job options'. *Times Higher Education* 15th May 2008, 7.

Pastor, J. T., Ruiz, J. L. and Sirvent, I. (2002). 'A statistical test for nested radial DEA models'. *Operations Research,* 50 (4), 728-735.

Pollard, E., Williams, M., Williams, J., Bertram, C. and Buzzeo, J. (2013). *How should we measure higher education? A fundamental review of the Performance Indicators, Part 2: The evidence report*. Institute for Employment Studies, Brighton.

Popov, S. V. and Bernhardt, D. (2013). 'University competition, grading standards, and grade inflation'. *Economic inquiry,* 51 (3), 1764-1778.

Ray, S. C. and Jeon, Y. (2008). 'Reputation and efficiency: A non-parametric assessment of America's top-rated MBA programs'. *European Journal of Operational Research,* 189 (1), 245-268.

Rolfe, H. (2003). 'University strategy in an age of uncertainty: The effect of higher education funding on old and new universities'. *Higher Education Quarterly,* 57 (1), 24-47.

Saisana, M., d'Hombres, B. and Saltelli, A. (2011). 'Rickety numbers: Volatility of university rankings and policy implications'. *Research Policy,* 40 (1), 165-177.

Saltelli, A., Nardo, M., Tarantola, S., Giovannini, E., Hoffman, A. and Saisana, M. (2005). *Handbook on Constructing Composite Indicators: Methodology and User Guide*: OECD.

Shin, J. C. and Toutkoushian, R. K. (2011). 'The past, present, and future of university rankings'. In J. C. Shin, R. K. Toutkoushian and U. Teichler (Eds), *University Rankings: Theoretical Basis, Methodology and Impacts on Global Higher Education* (pp. 1-18). Dordrecht, The Netherlands: Springer.

Simon, J., Simon, C. and Arias, A. (2011). 'Changes in productivity of Spanish university libraries '. *Omega,* 39 (5), 578-588.

Smith, J. P., McKnight, A. and Naylor, R. (2000). 'Graduate employability: Policy and performance in higher education in the UK '. *Economic Journal*, 110 (464), 382-411.

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

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

Usher, A. and Medow, J. (2009). 'A global survey of university rankings and league tables'. In B. M. Kehm and B. Stensaker (Eds), *University Rankings, Diversity, and the new Landscape of Higher Education* (pp. 3-18). Rotterdam: Sense Publishers.

Williams, G. (1992a). 'British higher education in the world league'. *Oxford Review of Economic Policy*, 8 (2), 146-158.

Williams, G. (1992b). *Changing Patterns of Finance in Higher Education*. Bristol, PA; Buckingham [England]: Society for Research into Higher Education.

Yorke, M. (1997). 'A good league table guide?'. *Quality Assurance in Education,* 5 (2), 61-72.