An EFQM-Rembrandt excellence model based on the theory of displaced ideal

Madjid Tavana
Management Information Systems, La Salle University, Philadelphia, Pennsylvania, USA

Amir Karbassi Yazdi and Mehran Shiri
Department of Industrial Management, Islamic Azad University – South Tehran Branch, Tehran, Iran, and

Jack Rappaport
Management Department, La Salle University, Philadelphia, Pennsylvania, USA

Abstract

Purpose – This paper aims to propose a new benchmarking framework that uses a series of existing intuitive and analytical methods to systematically capture both objective data and subjective beliefs and preferences from a group of decision makers (DMs).

Design/methodology/approach – The proposed framework combines the excellence model developed by the European Foundation for Quality Management with the Rembrandt method, the entropy concept, the weighted-sum approach, and the theory of the displaced ideal. Hard data and personal judgments are synthesized to evaluate a set of business units (BUs) with two overall performance scores plotted in a four quadrant model.

Findings – The two performance scores are used to benchmark the performance of the BUs in accordance with their Euclidean distance from the "ideal" BU. Quadrants are used to classify the BUs as efficacious, productive ineffectual, proficient unproductive, and inefficacious. The efficacious BUs, referred to as "excellent", fall in the competency zone and have the shortest Euclidean distance from the ideal BU relative to their peers.

Originality/value – The benchmarking framework presented in this study has some obvious attractive features. First, the generic nature of the framework allows for the subjective and objective evaluation of a finite number of BUs by a group of DMs. Second, the information requirements of the framework are stratified hierarchically allowing DMs to focus on a small area of the large problem. Third, the framework does not dispel subjectivity; it calibrates the subjective weights with the objective weights determined through the entropy concept.

Keywords Multi-attribute decision making, Benchmarking, Rembrandt method, EFQM excellence model, Entropy concept, Weighted-sum approach, Theory of displaced ideal, European Foundation for Quality Management, Business excellence

Paper type Research paper

Benchmarking: An International Journal
Vol. 18 No. 5, 2011
pp. 644-667

1463-5771
DOI 10.1108/14635771111166802

The authors would like to thank the anonymous reviewers and the Editor for their insightful comments and suggestions.

1. Introduction

In this study, a benchmarking framework is proposed which uses:

- the excellence model developed by the European Foundation for Quality Management (EFQM, 2003) to identify the relevant decision criteria for evaluating a set of business units (BUs);
- the Rembrandt method to assign importance weights to these decision criteria;
- the entropy concept to revise these importance weights with the intrinsic weights derived from the performance scores;
- the weighted-sum approach to arrive at two weighted overall scores for each BU under consideration; and
- the theory of the displaced ideal to represent the degree of excellence for each BU.

The EFQM excellence model is a practical benchmarking tool to help organizations measure where they are on the path to excellence, help them understand the gaps, and then stimulate the best possible solutions (Karkoszka and Szewieczek, 2007; Michalska, 2008; Nazemi, 2010; Santos-Vijande and Alvarez-Gonzalez, 2007). This paper is organized into five sections. The following section presents the literature review. In Section 3, details of the proposed framework is outlined. In Section 4, there is an application of the applicability of the proposed framework to exhibit the efficacy of the procedures and algorithms. Section 5 consists of the conclusion and future research directions.

2. Literature review

Multi-criteria decision-making (MCDM) methods are frequently used to solve real-world problems with multiple, conflicting, and incommensurate criteria. Each method provides a different approach for selecting the best among several preselected alternatives (Janic and Reggiani, 2002). MCDM methods are generally categorized as continuous or discrete, depending on the domain of alternatives. Hwang and Yoon (1981) have classified the MCDM methods into two categories: multi-objective decision making (MODM) and multi-attribute decision making (MADM).

MODM has been widely studied by means of mathematical programming methods with well-formulated theoretical frameworks. MODM methods have decision variable values that are determined in a continuous or integer domain with either an infinite or a large number of alternative choices, the best of which should satisfy the decision-maker (DM) constraints and preference priorities (Hwang and Masud, 1979; Ehrgott and Wiecek, 2005).

Conversely, MADM methods have been used to solve problems with discrete decision spaces and a predetermined or a limited number of alternative choices. The MADM solution process requires inter- and intra-attribute comparisons and involves implicit or explicit tradeoffs (Hwang and Yoon, 1981). MADM methods are used for circumstances that necessitate the consideration of different options that cannot be measured in a single dimension. MADM methods assist DMs to learn about the issues they face, the value systems of their own and other parties, and the organizational values and objectives that will consequently guide them in identifying a preferred course of action. The primary goal in MADM is to provide a set of attribute-aggregation methodologies for considering the preferences and judgments of DMs (Doumpos and Zopounidis, 2002). Roy (1990) argues that solving MADM problems is not searching for an optimal solution, but rather helping DMs master the complex judgments and data involved in their problems and advance towards an acceptable solution. Multi-attributes analysis is not a template that can be applied to every problem and situation. The development of MADM models has often been dictated by real-life problems. Therefore, it is not surprising that methods have appeared in a rather diffuse manner, without any clear general methodology or basic theory (Vincke, 1992). The selection of a MADM framework or method should be done carefully according to the nature of the problem, types of choices, measurement scales, dependency among the attributes, type of uncertainty, expectations of the DMs, and the quantity and quality of the available data and judgments (Vincke, 1992). Finding the best MADM framework is an elusive goal that may never be reached (Triantaphyllou, 2000).

The EFQM model is a method used to evaluate organizations or BUs. The EFQM excellence model is a practical MCDM tool for self-assessment, a way to benchmark with other organizations, a guide to identify areas for improvement, or a structure for the organization's management system (Michalska, 2008). The EFQM excellence model is a practical benchmarking tool to help organizations measure where they are on the path to excellence, help them understand the gaps, and then stimulate the best possible solutions (Karkoszka and Szewieczek, 2007; Michalska, 2008; Nazemi, 2010; Santos-Vijande and Alvarez-Gonzalez, 2007).

The EFQM model has been widely used in health care throughout the world. Moeller *et al.* (2000) used the EFQM excellence model to address the increasing expenditure within health care and the increasing taxes and budget spending in Germany. Nabitz *et al.* (2000) showed the use of the EFQM excellence model as an auditing instrument on all levels of a health care organization in The Netherlands. Vogt (2001) showed how the EFQM model is used in German hospital laboratories to realize continuous improvement of service quality and at the same time reduce costs. Persaud (2002) discussed the application of the EFQM model for continuous quality improvement within the health care industry in the UK. Sánchez *et al.* (2005) described the implementation of the EFQM excellence model as a common framework for quality management in a regional health care service in Northern Spain. Vallejo *et al.* (2006) brought the EFQM fundamental concepts of excellence closer to health care by using a specific model for performance assessment developed for the World Health Organization Regional Office for Europe. Other authors have studied the application of the EFQM excellence model in health care for achieving clinical governance, quality improvement, benchmarking, and performance assessment (Jackson, 2000; Jackson and Bircher, 2002; Moeller and Sonntag, 2001; Stewart, 2003).

The model's framework is based on nine criteria as follow:

(1) leadership;

(2) people;

(3) policy and strategy;

(4) partnerships and resources;

(5) processes;

(6) people;

(7) customer;

(8) society; and

(9) key performance.

The first five criteria are "enablers" and the remaining four are "results". The "enablers" criteria cover what an organization does. The "results" criteria cover what an organization achieves. As shown in Figure 1, results are caused by enablers.

The criterion weights of the enablers and the results in the EFQM excellence model have always been an important part of the model. This is true with most of the other award models (Porter and Tanner, 1998). This importance stems from the fact that the award models are generally used to compare an organization with other organizations or to rate an organization against a commonly adopted benchmark (Lascelles and Peacock, 1996; Conti, 1997). Several researchers have studied criteria weights in the EFQM excellence model (Bemowski and Stratton, 1995; Coulambidou and Dale, 1995; Dahlgaard *et al.*, 1998; Dale and Ritchie, 2000; Donnelly, 2000; Eskildsen *et al.*, 2004, 2003, 1999; Juhl *et al.*, 2002; Malorny, 1996; Olson, 1996; Teo and Dale, 1997; Van der Wiele *et al.*, 1996).

The Rembrandt method (López and Monzón, 2010; Van den Honert and Lootsma, 2000; Lootsma, 1992, 1996; Olson *et al.*, 1995) is applied to calculate the weights of the enablers and results in sub-criteria in the EFQM excellence model. The Rembrandt method has been designed to address three criticized features of the analytic hierarchy process (AHP) (Saaty, 2000) including:

(1) different ratio input scales;

(2) alternative calculation of impact scores; and

(3) a different aggregation procedure.

The Rembrandt method is one of the best known attempts to retain the strengths of AHP while avoiding some of its objections (López and Monzón, 2010).

The Rembrandt weights are calibrated according to the entropy concept proposed by Zeleny (1982). The essential idea is that the overall importance of a criterion is a direct function of its subjective weights provided by the Rembrandt method

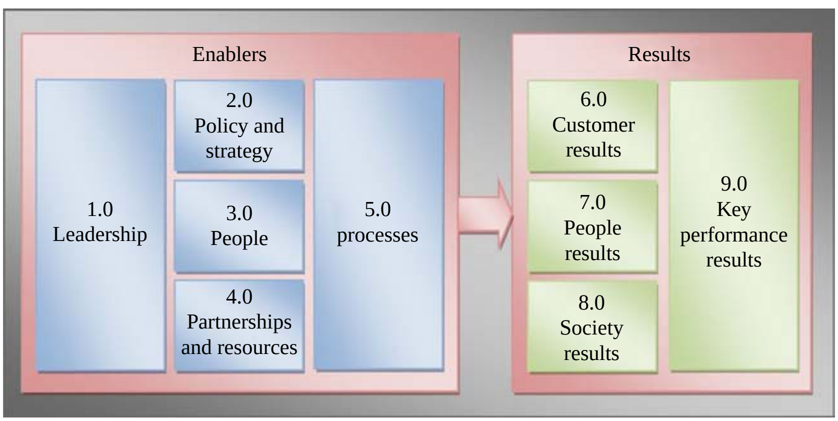

Figure 1. The EFQM excellence model

and the information conveyed by the criteria relative to the entire set of BUs. This means that the greater the dispersion in the performance score, the more important the criteria. Each criterion is an information source; therefore, the more information an EFQM enabler or result reveals, the more relevant it is. In other words, the most important criteria are those that have the greatest discriminating power between the BUs.

According to the entropy concept, a set of intrinsic weights associated with the enablers and results sub-criteria are determined without the direct involvement of the DMs. However, this is a complete contradiction to the notion that weights should represent the relative importance the DMs attach to the enablers and results sub-criteria. Therefore, the subjective weights representing the judgments of the DMs obtained by the Rembrandt method are multiplied by the values of weights (intrinsic weights) obtained by the entropy method. The final result, once normalized, will be used to find the overall enablers and results scores associated with the BUs under evaluation.

Pardalos and Hearn (2002) discuss the importance of exploring ways of combining criteria aggregation methodologies to enable the development of models that consider the DM's preferential system in complex problems. Belton and Stewart (2002) also argue the need for integrating frameworks in MCDM. A weighted-sum method is used to aggregate the results from the Rembrandt method and the entropy concept. This aggregation allows for the objective data and subjective judgments to be collected and used side-by-side in a weighted-sum model (Triantaphyllou, 2000).

3. The proposed framework

The proposed framework is comprised of five distinct phases: the EFQM phase, the Rembrandt phase, the entropy phase, the weighted-sum phase, and the theory of the displaced ideal phase.

3.1 The EFQM phase

Initially, the relevant factors to be used in the proposed benchmarking framework are identified according to the EFQM excellence model. These factors are essentially the sub-criteria for:

- the leadership, people, policy and strategy, partnerships and resources, and processes enablers; and
- the people, customer, society, and key performance results.

3.2 The Rembrandt phase

The Rembrandt method proposed by Van den Honert and Lootsma (2000) is used to ascertain the relative importance of the enabler and result criteria and their respective sub-criteria in the EFQM excellence model. Let us formulate the model by considering a group of g DMs ($g \geq 1$) which are charged with evaluating m criteria ($m \geq 1$). Assuming that criteria $C_i, i = 1, \ldots, m$, has an unknown subjective value V_i, and that V_i is the same for all DMs in the group, the Rembrandt method is used to estimate the $m-$ vector of V_i values from the DMs' verbal subjective judgments. Each DM is instructed to record his/her graded comparative judgment on pairs of criteria, C_l and C_j, in the decision matrix $D_{m \times n}$. The process requires between $(m-1)$ and $m(m-1)/2$ pairwise comparisons for a set of m criteria under evaluation. That is, each DM records his/her indifference between the two criteria as a weak, definite, strong, or very strong preference for one criteria over the other. Incomplete pairwise comparisons are handled using a general procedure proposed by Lootsma (1997, pp. 114-17). The subjective criteria weights are normalized because of the ratio information so that $\sum_i V_i = 1$.

The DM d's pairwise comparison judgments are captured on a category scale to limit the range of verbal responses. Each verbal response is converted into an integer-valued gradation index δ_{jld} using the scale presented in Table I.

The gradation index δ_{jld} is then converted into a value on a geometric scale, characterized by a scale parameter γ. Thus, r_{jld}, the numeric estimate of the preference ratio V_j/V_l given by DM d is defined as:

$$r_{jld} = \exp(\gamma\delta_{jld}); \quad j, l = 1, \ldots, m; \quad d = 1, \ldots, g. \tag{1}$$

Given that there is no unique scale for human judgment, a plausible value of γ for the group is $\ln\sqrt{2}$ implying a geometric scale with progression factor $\sqrt{2}$ (Lootsma, 1993). Then $\mathbf{V}$ is approximated by the normalized vector of $\mathbf{v}$ of group weights which minimizes:

$$\sum_{j<l}\sum_{d=1}^{g}(\ln r_{jld} - \ln v_j + \ln u_l)^2; \quad l = 2, \ldots, m. \tag{2}$$

Assume that all DMs offer a complete set of pariwise comparisons. Now, let $\rho_{jld} = \ln r_{jld} = \gamma\delta_{jld}$ and $w_j = \ln v_j$. Then the vector $\mathbf{v}$ is found by minimizing equation (3) as a fuction of $w_j (j = 1, \ldots, m)$:

$$\theta = \sum_{j<l}\sum_{d=1}^{g}\left(\rho_{jld} - w_j + w_l^2\right); \quad l = 2, \ldots, m. \tag{3}$$

The set dependence of normal questions are found from:

$$\frac{\partial\theta}{\partial w_j} = \sum_{j<l}\sum_{d=1}^{g}(\rho_{jld} - w_j + w_l) = 0; \quad j = 1, \ldots, m; \quad l = 2, \ldots, m. \tag{4}$$

Since $\rho_{jld} = \rho_{ljd}$ and $\rho_{jjd} = 0$ for any j, equation (4) may be written as:

Comparative judgment	Gradation index δ_{jkd}
Very strong preference for C_k over C_j	-8
Strong preference for C_k over C_j	-6
Definite preference for C_k over C_j	-4
Weak preference for C_k over C_j	-2
Indifference between C_j and C_k	0
Weak preference for C_j and C_k	$+2$
Definite preference for C_j and C_k	$+4$
Strong preference for C_j and C_k	$+6$
Very strong preference for C_j and C_k	$+8$

Table I. The Rembrandt category scale

$$\gamma\sum_{l=1}^{m}\sum_{d=1}^{g}\delta_{jld}=\sum_{l=1}^{m}\sum_{d=1}^{g}w_j-\sum_{l=1}^{m}\sum_{d=1}^{g}w_l;\quad j=1,\ldots,m. \tag{5}$$

There is no unique solution to this set of normal equations and the sum of the variables ($\sum w_l$) is set equal to zero for a particular solution and we reduce equation (5) to the following unnormalized solution:

$$w_j=\frac{1}{g}\,\frac{1}{m}\,\gamma\sum_{l=1}^{m}\sum_{d=1}^{g}\delta_{jld};\quad j=1,\ldots,m. \tag{6}$$

Therefore:

$$v_j=\exp(w_j)=\exp\left(\frac{1}{g}\,\frac{1}{m}\,\gamma\sum_{k=1}^{m}\sum_{d=1}^{g}\delta_{jld}\right) \tag{7}$$

and:

$$v_j={}^{\mathrm{m}}\sqrt{\prod_{l=1}^{m}\prod_{d=1}^{g}r_{jld}^{1/g}};\quad j=1,\ldots,m. \tag{8}$$

where equation (8) implies that the criteria weights of v_j are calculated by a sequence of geometric means.

The result of v_j in equation (8) are multiplied with the degree of freedom to determine the normalized solution vector $\mathbf{v}$. In addition, since $v_j=f(\exp(\gamma\delta_{jld}))$, the normalized criteria weights will depended on the scale parameter γ, without changing the rank ordering of v_j.

3.3 The entropy phase

The entropy method is used to determine the importance weights associated with the EFQM sub-criteria without the direct involvement of the DMs in terms of the performance score of the EFQM enablers or results of the BUs. Zeleny (1982) shows that this intrinsic information must be used in parallel with the subjective weights (Rembrandt weights) assigned to various enabler and result criteria and sub-criteria. In other words, the overall importance weight of an enabler or result criterion ($\bar{v}_j$) is directly related to the intrinsic weight of that enabler or result criterion ($\hat{v}_j$) and the Rembrandt weight of the enabler and result criteria. The more different the performance scores of an enabler or result criteria are with respect to a set of BUs, the larger is the contrast intensity of the criteria, and the greater is the amount of information transmitted by that criteria.

Assuming that the vector $p_{fj}=(p_{fj}^1,\ldots,p_{fj}^q)$ characterizes the performance of the qth BU according to the fth criterion and the jth sub-criterion, the entropy measure for a given fj and q is:

$$e(p_{fj})=-K\sum_{k=1}^{q}\frac{p_{fj}^q}{p_{fj}}\ln\frac{p_{fj}^q}{p_{fj}} \tag{9}$$

where $p_{fj}=\sum_{k=1}^{q}p_{fj}^k,\quad j=1,\ldots,m;\quad f=1,\ldots,h$ and $K\rangle 0$, ln is the natural logarithm, $0\leq p_{fj}^q\leq 1$, and $e(p_{fj})\geq 0$. When all p_{fj}^q are equal for a given fj and q, then $p^q_{fj}/p_{fj} = 1/r$, and $e(p_{fj})$ assumes its maximum value, which is $e_{\max} = \ln r$. By setting $K = 1/e_{\max}$, $0 \leq e(p_{fj}) \leq 1$ is achieved for all p_{fj}'s. This normalization is necessary for meaningful comparisons. In addition, the total entropy is defined as:

$$E = \sum_{f=1}^{h} \sum_{j=1}^{m} e(p_{fj}) \tag{10}$$

The smaller $e(p_{fj})$ is, the more information the *f*th criterion and the *j*th sub-criterion transmits and the larger $e(p_{fj})$ is, the less information it transmits. When $e(p_{fj}) = e_{\max} = \ln r$, the *f*th criterion and the *j*th sub-criterion transmits no useful information. Next, the intrinsic weight is calculated as:

$$\hat{v}_{fj} = \frac{1}{m - E}[1 - e(p_{fj})] \tag{11}$$

where m is the total number of sub-criteria.

Because $\hat{v}_{fj}$ is inversely related to $e(p_{fj})$, $1 - e(p_{fj})$ is used instead and normalized to ensure $0 \leq \hat{v}_{fj} \leq 1$ and $\sum_{f=1}^{h}\sum_{j=1}^{m}\hat{v}_{fj} = 1$. The higher $e(p_{fj})$, the less information content is provided by the *f*th criterion and the *j*th sub-criterion. When the information content of the *f*th criterion and the *j*th sub-criterion is low, the corresponding intrinsic weight ($\hat{v}_{fj}$) should be low. Thus, the intrinsic weight is assumed to be inversely related to the entropy. Therefore, $1 - e(p_{fj})$ is used in the definition of the intrinsic weight.

The more different the p^q_{fj} scores are, the larger $\hat{v}_{fj}$ is and the more important the *f*th criterion and the *j*th sub-criterion are. When all the scores, p^q_{fj}, are equal for the *f*th criterion and the *j*th sub-criterion, then, $\hat{v}_{fj} = 0$ for that criterion. However, this is not true if the scores p^q_{fj} are equal for all the *f*th criterion and the *j*th sub-criterion. In that case, the weights are assumed to be equal or $\hat{v}_{fj} = 1/m$ where m is the number of sub-criteria. Entropy multiplies the intrinsic weight ($\hat{v}_{fj}$) by the subjective (Rembrandt) weight (v_{fj}) and normalizes the product to calculate the overall importance weight of the *j*th criterion ($\bar{v}_{fj}$):

$$\bar{v}_{fj} = \frac{v_{fj}.\hat{v}_{fj}}{\sum_{f=1}^{h}\sum_{j=1}^{m} v_{fj}.\hat{v}_{fj}} \tag{12}$$

There are two other methods for calculating the intrinsic weights of the threats and the responses. Diakoulaki *et al.* (2000) proposes a method based on the correlation between the columns of the decision matrix. The other method measures the importance of each threat or response as a member of a coalition by means of the Shapley value (Grabisch and Roubens, 1999). The entropy method as suggested by Zeleny (1982, Chapter 7) is used as it is readily available in MCDM, provides consistent results, and easily is accepted by DMs (Pomero and Barba-Romero, 2000, Chapter 4).

3.4 The weighted-sum phase

Weighted-sum models are used to combine the weights of the EFQM criteria (w_f) with the overall weights of the EFQM sub-criteria ($\bar{v}_{fj}$) and the performance scores of the *k*th BU for the *f*th criterion and the *j*th sub-criterion (p^k_{fj}). The first model is used to find an overall "enablers score" for each BU (S^k_e) and the second model is used to find an overall "results score" for each BU (S^k_r):

$$S_e^k = \sum_{f=1}^{5}\sum_{j=1}^{m}\sum_{k=1}^{q} w_f.\bar{v}_{fj}.p_{fj}^k \tag{13}$$

$$S_r^k = \sum_{f=6}^{9}\sum_{j=1}^{m}\sum_{k=1}^{q} w_f.\bar{v}_{fj}.p_{fj}^k \tag{14}$$

Triantaphyllou (2000) has discussed the mathematical properties of weighted-sum MADM models. Many weighted-sum models have been developed to help DMs deal with the strategy evaluation process (Gouveia *et al.*, 2008; Leyva-Lopez and Fernandez-Gonzalez, 2003). Triantaphyllou and Baig (2005) have examined the use of four important weighted-sum MADM methods when advantages and disadvantages, corresponding to opportunities and threats, are used as conflicting criteria. They compared the simple weighted-sum model, the weighted-product model, and the AHP along with some of its variants, including the multiplicative AHP. Their extensive empirical analysis revealed some ranking inconsistencies among the four methods, especially, when the number of alternatives was high. Although they were not able to show which method results in the appropriate classification, they did prove multiplicative AHP is immune to ranking inconsistencies.

3.5 The theory of displaced ideal phase

The weighted-sum scores in this model are used to compare potential BUs among themselves and with the ideal BU. The concept of ideal state, an unattainable idea, serving as a norm or rationale facilitating human choice problem is not new (Tavana, 2002). The seminal work of Schelling (1960) introduced the concept. Subsequently, Festinger (1964) showed that an external, generally non-accessible choice assumes the important role of a point of reference against which choices are measured. Zeleny (1974, 1982) demonstrated how the highest achievable scores on all currently considered decision criteria form this composite ideal choice. As all choices are compared, those closer to the ideal are preferred to those farther away. Zeleny (1982) shows that the Euclidean measure can be used as a proxy measure of distance.

An ideal BU is the one with the highest possible enablers score ($S_e^k = 1$) and highest possible results score ($S_r^k = 1$). In this phase, the Euclidean distance of each BU from the ideal BU (D^k) is estimated. The Euclidean distance is the sum of the quadratic root of squared differences between the ideal BU and the kth BU. BUs with smaller D^k are closer to the ideal BU and are preferred to BUs with larger D^k which are further away from the ideal BU.

The BUs were then graphed and the x-axis represents the enablers scores and the y-axis represents the results scores. The position of the point corresponding to the kth BU has Cartesian coordinates (S_e^k, S_r^k) on the graph. The average enablers score ($\bar{S}_e^k$) and the average results score ($\bar{S}_r^k$) divide this graph into quadrants (competency, incapability, capability, and incompetency) as shown in Figure 2:

- *The efficacious zone*. This quadrant includes optimal BUs with above average enablers and results scores.
- *The productive ineffectual zone*. This quadrant includes BUs with above average results scores and below average enablers scores. These BUs produce results

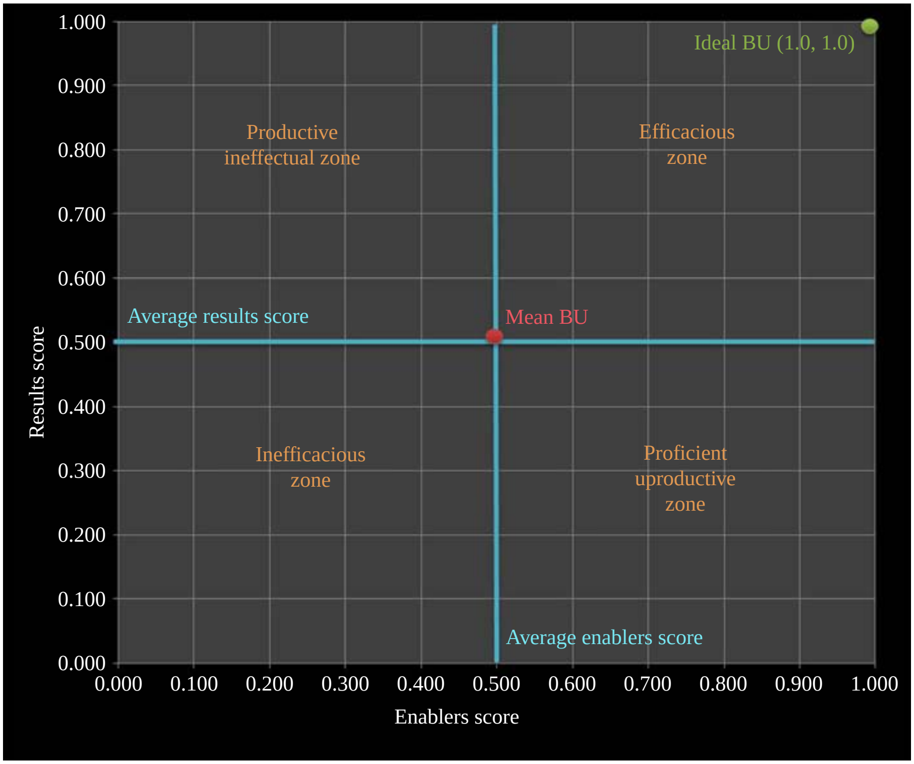

Figure 2.
The benchmarking model

in spite of their incapability. In most circumstances, improving capabilities may improve results.

- *The proficient unproductive zone*. This quadrant includes inefficient BUs with above average enablers score and below average results scores. These BUs are capable but cannot produce results. The BUs in this quadrant should redirect their capabilities to produce satisfactory results.
- *The inefficacious zone*. This quadrant includes BUs with below average scores with respect to the enablers and results scores. These BUs should not be considered.

The results of the model can be used to create a variety of benchmarking systems. In the macro-approach, the overall ideal BU can be considered as a benchmark with a score of 1.000 on both components (i.e. the enablers and the results component). Each BU is then benchmarked against this ideal BU using the Euclidean distance. Another potential model in the macro-approach is to benchmark each BU against the ideal for each component separately. For example, hospital B could have a results score of 1.000 (which implies a perfect benchmark value for this component), but a low enablers score of 0.439.

In the micro-approach, a hypothetical BU could be used as a benchmark which is comprised of the highest attained score by the BUs on each of the criteria.

This is similar to benchmarking against the ideal of 1.000, but it does represent a different benchmarking approach. In a similar approach, a benchmarking system can be used with just the enablers or the results criteria. For example, a hypothetical BU could be used as a benchmark comprised of the highest score that is attained by each of the BUs on each of the enablers criteria. Finally, a benchmarking system could be implemented where each of the criteria is used separately as a benchmark against the BU with the highest score on a particular criterion.

Once the model is developed, sensitivity analyses could be performed to determine the impact on the ranking of the BUs for changes in various model assumptions. Some sensitivity analyses that are usually of interest are on the weights and scores. The weights representing the relative importance of the criteria and sub-criteria are occasionally a point for discussion among the various DMs. In addition, scores which reflect the degree of performance of an uncertain criterion are sometimes a matter of contention.

4. An application of the methods in benchmarking

The results of a pilot benchmarking study conducted for the Office of Inspector General (OIG) – East Virginia Department of Health and Human Services[1] as given below. A total of 15 hospitals were selected by OIG to participate in this pilot study. All hospitals were not-for-profit with an average age of 16.2 years. The hospitals had an average bed capacity of 346 and an average occupancy rate of 74 percent. The hospital utilization rates for the state of East Virginia are generally lower than the US average. The average number of acute hospital days per 1,000 people for these 15 hospitals in 2009 was 497.2 while the admissions per 1,000 were 95.6. The OIG assigned five seasoned inspectors, referred to as DMs, to collect and synthesize all the necessary data and judgments. The DMs agreed to use the EFQM excellence model with five enablers, four results, and 41 sub-criteria presented in Table II to measure the strengths and areas for improvement of each hospital across all of their activities.

The five DMs started the process by evaluating the relative importance of the "enablers" and the "results" in the EFQM excellence model. Two pairwise comparison matrices were constructed by each DM to evaluate the importance of leadership, policy and strategy, people, partnerships and resources, and process enablers; and similarly, the importance of the consumer, the employee, society, and key performance results. Two sets of comparison matrices (one set of five enablers and one set of five results comparison matrices) were collected and synthesized for the group. The geometric mean was used to calculate the average scores presented in Tables III and IV.

As shown in these tables, the group generally felt that the five enablers and the four results were equally important.

Next, the DMs were asked to repeat the pairwise comparison process for the five leadership enabler sub-criteria, four policy and strategy enabler sub-criteria, five people enabler sub-criteria, five partnerships and resources enabler sub-criteria, and five process enabler sub-criteria. The judgments from the five DMs were combined and synthesized with the geometric means presented in Table V.

The DMs then repeated the pairwise comparison process for the four customer results sub-criteria, three people results sub-criteria, two society results sub-criteria, and the eight key performance results sub-criteria. The judgments from the five DMs were similarly combined and synthesized with the geometric means presented in Table VI.

Criteria and sub-criteria	Criteria and sub-criteria descriptions
1.0	*Leadership enablers*
1.a	Leaders develop the mission, vision and values and are role models of a culture of excellence
1.b	Leaders are personally involved in ensuring the organization's management system is developed, implemented and continuously improved
1.c	Leaders are involved with customers, partners and representatives of society
1.d	Leaders motivate, support and recognize the organization's people
1.e	Leaders provide guidance on tasks and check progress
2.0	*Policy and strategy enablers*
2.a	Policy and strategy are based on the present and future needs and expectations of stakeholders
2.b	Policy and strategy are based on information from performance measurement, research, learning and creativity related activities
2.c	Policy and strategy are developed, reviewed, updated and deployed through a framework of key processes
2.d	Policy and strategy are communicated and implemented
3.0	*People enablers*
3.a	People resources are planned, managed and improved
3.b	People's knowledge and competencies are identified, developed and sustained
3.c	People are involved and empowered
3.d	People and the organization have a dialogue
3.e	People are rewarded, recognized and cared for
4.0	*Partnerships and resources enablers*
4.a	External partnerships are managed
4.b	Finances are managed
4.c	Buildings, equipment and materials are managed
4.d	Technology is managed
4.e	Information and knowledge are managed
5.0	*Processes enablers*
5.a	Processes are systematically designed and managed
5.b	Processes are improved, as needed, using innovation in order to fully satisfy and generate increasing value for customers and other stakeholders
5.c	Products and services are designed and developed based on customer needs and expectations
5.d	Products and services are produced, delivered and serviced
5.e	Customer relationships are managed and enhanced
6.0	*Customer results*
6.a	Increased satisfaction
6.b	Increased loyalty
6.c	Improved quality
6.d	Reduced complaints
7.0	*People results*
7.a	Increased engagement
7.b	Training delivery
7.c	Increased productivity
8.0	*Society results*
8.a	Reduced waste
8.b	Reduced energy

(*continued*)

Table II.
The EFQM excellence model criteria and sub-criteria

Criteria and sub-criteria	Criteria and sub-criteria descriptions
9.0	*Key performance results*
9.a	Reduced turnover
9.b	Increased profit
9.c	Improved cost savings
9.d	Improved productivity
9.e	Improved quality
9.f	Improved efficiency
9.g	Improved availability of knowledge and information
9.h	Reduced errors/defects

Table II.

Enablers	Leadership	Policy and strategy	People	Partnerships and resources	Processes	...	v_j
Leadership	1	2.07	1.742	1.414	1.425	...	0.213
Policy and strategy	0.482	1	1.867	1.742	1.57	...	0.208
People	0.573	0.535	1	2.07	1	...	0.197
Partnerships and resources	0.706	0.573	0.482	1	1.89	...	0.193
Processes	0.615	0.636	0.84	1	1	...	0.189

Table III.
The enablers criteria pairwise comparison matrix and weights

Results	Customer results	Employee results	Society results	Key performance results	...	v_j
Customer results	1	1.07	1.231	1	...	0.254
People results	0.932	1	1	1.189	...	0.250
Society results	0.812	1	1	1.275	...	0.251
Key performance results	1	0.84	0.784	1	...	0.245

Table IV.
The results criteria pairwise comparison matrix and weights

Next, the DMs met collectively to discuss the appropriate metrics for evaluating the performance of the 15 hospitals on the basis of the 41 enablers and results sub-criteria presented in Table II. Some criteria required objective data while others required subjective judgments. For example, sub-criterion 9.b (increased profit) was measured by retrieving the profit figures from the forms 501(c)3 that hospitals and other non-profit organizations file with the Internal Revenue Service. Other sub-criteria such as sub-criterion 1.1 (leaders develop the mission, vision and values and are role models of a culture of excellence) were more subjective and a ten-point Likert scale (range 1-10) was used by each DM to subjectively assess the overall performance of this criterion in each hospital. The objective data were normalized to arrive at a 0-1 performance score of each criterion at each hospital. All subjective performance scores provided by the five DMs were averaged and then normalized to arrive at a 0-1 performance score similar to the objective data. The normalized performance score of each criterion at each hospital for all DMs ($0 \leq p_{fj}^k \leq 1$) are presented in Table VII.

Table V. The pairwise comparison matrices and weights for the enablers criteria

The leadership sub-criteria							
	1.a	1.b	1.c	1.d	1.e	…	v_j
1.a	1	1.274	1.867	1	1.414	…	0.208
1.b	0.784	1	2.380	1.035	1.933	…	0.212
1.c	0.535	0.420	1	1.071	1.570	…	0.194
1.d	1	0.965	0.932	1	2.145	…	0.205
1.e	0.706	0.517	0.637	0.466	1	…	0.181
The policy and strategy sub-criteria							
	2.a	2.b	2.c	2.d	…	v_j	
2.a	1	1.366	1	1	…	0.254	
2.b	0.731	1	1.803	1.274	…	0.260	
2.c	1	0.554	1	1	…	0.241	
2.d	1	0.784	1	1	…	0.245	
The people sub-criteria							
	3.a	3.b	3.c	3.d	3.e	…	v_j
3.a	1	2.380	1.319	1.365	1.275	…	0.214
3.b	0.420	1	1.190	1.933	1	…	0.202
3.c	0.757	0.840	1	1.035	1.109	…	0.196
3.d	0.731	0.517	0.965	1	1.189	…	0.193
3.e	0.784	1	0.901	0.840	1	…	0.195
The partnerships and resources sub-criteria							
	4.a	4.b	4.c	4.d	4.e	…	v_j
4.a	1	1.742	1.803	1.933	1.319	…	0.213
4.b	0.573	1	1.319	3.61	1.803	…	0.217
4.c	0.544	0.757	1	1.682	1.274	…	0.198
4.d	0.517	0.276	0.594	1	0.784	…	0.179
4.e	0.757	0.554	0.784	1.274	1	…	0.193
The processes sub-criteria							
	5.a	5.b	5.c	5.d	5.e	…	v_j
5.a	1	1	3.142	1.189	1	…	0.212
5.b	1	1	2.642	1.890	1.071	…	0.213
5.c	0.318	0.378	1	1.071	1.071	…	0.186
5.d	0.840	0.84	0.932	1	0.784	…	0.191
5.e	1	0.932	0.932	1.275	1	…	0.198

The Rembrandt method described in Section 3.1 is then used to calculate the enablers and results criteria weights (w_f) and the sub-criteria weights (v_{fj}) presented in Table VIII from the pairwise criterion matrices given in Tables III-VI.

The intrinsic weights ($\hat{v}_{fj}$) associated with the 41 enablers and results sub-criteria are calculated using the entropy process described in Section 3.2. The sub-criteria overall weights ($\bar{v}_{fj}$) presented in Tables VIII were calculated by multiplying the sub-criteria importance weights by their intrinsic weights and normalizing the product.

The weighted-sum approach described in Section 3.3 was used to combine the weights of the EFQM criteria (w_f) with the overall weights of the EFQM sub-criteria ($\bar{v}_{fj}$) and the performance scores of the kth hospital for the fth criterion and the jth sub-criterion (p^k_{fj}). One weighted-sum score represented the overall "enablers score" for each hospital (S^k_e) and the second one represented the overall "results score" for each hospital (S^k_r). An ideal hospital is the one with the highest possible enablers score ($S^k_e = 1$) and highest possible results score ($S^k_r = 1$). The Euclidean distance of each hospital from the ideal hospital (D^k) was ascertained. Hospitals with smaller D^k were

Table VI.
The pairwise comparison matrices and weights for the results criteria

The customer results sub-criteria

	6.a	6.b	6.c	6.d	...	v_j
6.a	1	1.189	1.274	1.035	...	0.279
6.b	0.840	1	1.109	1.23	...	0.258
6.c	0.784	0.901	1	1.109	...	0.334
6.d	0.966	0.813	0.901	1	...	0.229

The people results sub-criteria

	7.a	7.b	7.c	...	v_j
7.a	1	1.319	1.189	...	0.385
7.b	0.758	1	1.109	...	0.313
7.c	0.841	0.901	1	...	0.302

The society results sub-criteria

	8.a	8.b	...	v_j
8.a	1	3.368	...	0.771
8.b	0.296	1	...	0.229

The key performance results sub-criteria

	9.a	9.b	9.c	9.d	9.e	9.f	9.g	9.h	...	v_j
9.a	1	1.800	0.932	2.465	1.319	1.803	2.55	0.615	...	0.172
9.b	0.550	1	1.070	1.109	1.230	1.570	1.930	1.189	...	0.139
9.c	1.072	0.934	1	1.274	1.319	1.625	2.070	1.035	...	0.153
9.d	0.405	0.901	0.784	1	1.100	1.231	1.100	1.274	...	0.113
9.e	0.758	0.830	0.758	0.909	1	1.274	1.189	1.109	...	0.117
9.f	0.544	0.636	0.615	0.812	0.784	1	1.516	1.148	...	0.103
9.g	0.392	0.518	0.483	0.909	0.841	0.659	1	0.840	...	0.082
9.h	1.626	0.841	0.966	0.784	0.901	0.871	1.190	1	...	0.122

closer to the ideal hospital and were preferred to hospitals with larger D^k which were further away from the ideal hospital.

The hospitals were plotted on a graph where the x-axis represented the enablers scores and y-axis represented the results scores. The position of the point corresponding to the kth hospital has Cartesian coordinates (S_e^k, S_r^k) on the graph. The average enablers score $(\bar{S}_e^k = 0.427)$ and the average results score $(\bar{S}_r^k = 0.597)$ divided this graph into four quadrants as shown in Figure 3.

The competency zone included hospitals B, H, and K; the productive ineffectual zone included hospitals C, G, I, J, and O; the proficient unproductive zone included hospitals D, L, M, and N; and the inefficacious zone included hospitals A, E, and F. Hospitals B, H, and K in the efficacious zone had above average scores with respect to both the enablers and the results scores. These hospitals were considered as "excellent hospitals" by the East Virginia Department of Health and Human Services.

Finally, a rank order of the 15 hospitals according to their Euclidean distance from the ideal hospital is shown in Figure 3. The ideal hospital is a hypothetical hospital with an overall enablers score of 1.0 and an overall results score of 1.0. Hospitals with a smaller Euclidean distance are closer to the ideal hospital and are preferred to those with larger Euclidean distances.

We also carried out extensive sensitivity analysis on the most important aspects of the model. The Rembrandt method has been used infrequently in the literature so we decided to investigate some of the sensitivity issues pertaining to this model. The results of the sensitivity analysis showed that the Rembrandt method is quite

	Hospital scores $\left(p_{fj}^{k}\right)$														
Sub-criteria	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1.a	0.81	0.33	0.46	0.16	0.37	0.24	0.49	0.71	0.65	0.21	0.42	0.62	0.31	0.23	0.31
1.b	0.40	0.82	0.15	0.31	0.14	0.39	0.13	0.50	0.57	0.51	0.12	0.48	0.54	0.25	0.17
1.c	0.32	0.48	0.41	0.17	0.20	0.14	0.33	0.31	0.31	0.05	0.26	0.16	0.42	0.36	0.19
1.d	0.13	0.16	0.23	0.09	0.10	0.06	0.11	0.21	0.39	0.37	0.19	0.15	0.23	0.13	0.18
1.e	0.15	0.53	0.28	0.29	0.66	0.33	0.45	0.68	0.65	0.34	0.48	0.41	0.32	0.39	0.46
2.a	0.26	0.26	0.25	0.02	0.32	0.14	0.35	0.19	0.27	0.19	0.53	0.62	0.48	0.21	0.22
2.b	0.15	0.15	0.16	0.25	0.18	0.45	0.50	0.41	0.19	0.23	0.34	0.14	0.17	0.19	0.14
2.c	0.21	0.30	0.17	0.20	0.36	0.02	0.18	0.55	0.13	0.24	0.44	0.42	0.33	0.24	0.18
2.d	0.67	0.44	0.51	0.16	0.37	0.29	0.27	0.53	0.69	0.72	0.80	0.81	0.63	0.86	0.25
3.a	0.16	0.11	0.18	0.24	0.09	0.29	0.12	0.14	0.03	0.12	0.76	0.16	0.24	0.09	0.11
3.b	0.12	0.11	0.24	0.14	0.15	0.18	0.11	0.10	0.23	0.10	0.13	0.08	0.21	0.13	0.24
3.c	0.20	0.10	0.05	0.07	0.34	0.15	0.45	0.08	0.09	0.32	0.15	0.43	0.17	0.13	0.15
3.d	0.25	0.09	0.13	0.20	0.06	0.10	0.44	0.24	0.17	0.22	0.22	0.25	0.37	0.21	0.12
3.e	0.39	0.26	0.27	0.39	0.16	0.65	0.41	0.44	0.58	0.52	0.38	0.48	0.37	0.60	0.53
4.a	0.34	0.34	0.59	0.63	0.32	0.35	0.27	0.94	0.16	0.20	0.59	0.53	0.73	0.80	0.36
4.b	0.63	0.76	0.35	0.66	0.71	0.18	0.73	0.61	0.42	0.85	0.37	0.19	0.87	0.71	0.36
4.c	0.47	0.33	0.24	0.56	0.21	0.18	0.23	0.38	0.15	0.28	0.59	0.22	0.57	0.27	0.41
4.d	0.21	0.23	0.24	0.59	0.51	0.19	0.20	0.18	0.18	0.17	0.25	0.78	0.14	0.28	0.39
4.e	0.19	0.16	0.14	0.46	0.06	0.32	0.17	0.33	0.03	0.18	0.23	0.13	0.11	0.24	0.44
5.a	0.15	0.31	0.29	0.42	0.69	0.24	0.26	0.43	0.41	0.34	0.84	0.14	0.76	0.75	0.17
5.b	0.11	0.25	0.10	0.38	0.19	0.15	0.21	0.18	0.17	0.12	0.15	0.39	0.20	0.33	0.27
5.c	0.10	0.38	0.15	0.41	0.53	0.18	0.41	0.41	0.24	0.24	0.70	0.16	0.82	0.44	0.86
5.d	0.42	0.45	0.19	0.71	0.48	0.17	0.20	0.13	0.35	0.48	0.15	0.17	0.30	0.35	0.20
5.e	0.19	0.31	0.29	0.89	0.38	0.76	0.19	0.29	0.18	0.05	0.22	0.33	0.13	0.40	0.32
6.a	0.22	0.62	0.31	0.34	0.19	0.22	0.43	0.57	0.45	0.51	0.36	0.18	0.15	0.17	0.65
6.b	0.52	0.57	0.85	0.16	0.42	0.34	0.65	0.32	0.32	0.63	0.65	0.17	0.16	0.12	0.61
6.c	0.66	0.74	0.86	0.26	0.29	0.19	0.24	0.28	0.19	0.76	0.79	0.11	0.29	0.15	0.93
6.d	0.76	0.49	0.98	0.45	0.79	0.69	0.86	0.75	0.82	0.43	0.78	0.78	0.41	0.57	0.69
7.a	0.41	0.84	0.81	0.14	0.03	0.04	0.65	0.30	0.76	0.90	0.80	0.17	0.21	0.24	0.73
7.b	0.30	0.67	0.58	0.13	0.05	0.04	0.98	0.89	0.89	0.63	0.83	0.73	0.16	0.66	0.78
7.c	0.24	0.55	0.35	0.29	0.19	0.13	0.86	0.39	0.39	0.93	0.80	0.43	0.11	0.23	0.17
8.a	0.27	0.31	0.33	0.54	0.28	0.16	0.26	0.53	0.24	0.20	0.12	0.38	0.25	0.63	0.53
8.b	0.19	0.43	0.33	0.15	0.65	0.12	0.47	0.74	0.42	0.18	0.34	0.41	0.24	0.85	0.20
9.a	0.73	0.67	0.78	0.48	0.74	0.56	0.86	0.47	0.34	0.64	0.77	0.27	0.20	0.12	0.49
9.b	0.64	0.48	0.57	0.49	0.67	0.39	0.54	0.67	0.78	0.66	0.76	0.56	0.49	0.37	0.27
9.c	0.24	0.36	0.43	0.39	0.27	0.57	0.68	0.24	0.33	0.18	0.34	0.22	0.12	0.19	0.28
9.d	0.58	0.45	0.68	0.32	0.71	0.78	0.59	0.67	0.58	0.56	0.76	0.27	0.41	0.16	0.17
9.e	0.54	0.32	0.30	0.54	0.68	0.41	0.37	0.25	0.15	0.13	0.74	0.87	0.57	0.13	0.13
9.f	0.45	0.56	0.19	0.45	0.65	0.73	0.85	0.62	1.00	0.67	0.27	0.80	0.71	0.12	0.27
9.g	0.25	0.34	0.55	0.85	0.53	0.76	0.90	0.58	0.37	0.72	0.68	0.11	0.66	0.93	0.48
9.h	0.41	0.67	0.41	0.28	0.56	0.55	0.62	0.67	0.81	0.49	0.13	0.08	0.76	0.53	0.20

Table VII. The EFQM excellence model normalized performance scores

robust, similar to the results obtained from the general AHP. The criteria weights (w_f), as shown in Table VIII, were computed based on the pairwise comparisons in the Rembrandt method. A 10-50 percent increase (or decrease) in each pairwise comparison resulted in a very small change in the final criteria weights (less than 4.5 percent). A similar sensitivity analysis was applied to the calculation of the sub-criteria weights for the first criterion. Similarly, a fairly large increase or decrease in one of the pairwise comparisons resulted in a relatively small change in the final sub-criteria weights.

Table VIII.
The EFQM excellence model weights

Criteria	Criteria weight (w_f)	Sub-criteria	Sub-criteria importance weight (v_{fj})	Sub-criteria intrinsic weight ($\hat{v}_{fj}$)	Sub-criteria overall weight ($\bar{v}_{fj}$)
1.0	0.213	1.a	0.208	0.190	0.195
		1.b	0.212	0.289	0.302
		1.c	0.194	0.188	0.180
		1.d	0.205	0.216	0.218
		1.e	0.181	0.117	0.105
2.0	0.208	2.a	0.254	0.320	0.307
		2.b	0.260	0.225	0.228
		2.c	0.241	0.230	0.273
		2.d	0.245	0.226	0.192
3.0	0.197	3.a	0.214	0.411	0.414
		3.b	0.202	0.080	0.082
		3.c	0.196	0.280	0.278
		3.d	0.193	0.159	0.156
		3.e	0.195	0.070	0.070
4.0	0.193	4.a	0.213	0.179	0.193
		4.b	0.217	0.146	0.160
		4.c	0.198	0.149	0.149
		4.d	0.179	0.247	0.224
		4.e	0.193	0.280	0.274
5.0	0.189	5.a	0.212	0.209	0.223
		5.b	0.213	0.118	0.126
		5.c	0.186	0.230	0.216
		5.d	0.191	0.174	0.167
		5.e	0.198	0.270	0.269
6.0	0.254	6.a	0.279	0.225	0.214
		6.b	0.258	0.283	0.249
		6.c	0.334	0.427	0.487
		6.d	0.229	0.065	0.051
7.0	0.250	7.a	0.385	0.389	0.443
		7.b	0.313	0.318	0.295
		7.c	0.302	0.293	0.262
8.0	0.251	8.a	0.771	0.393	0.685
		8.b	0.229	0.607	0.315
9.0	0.245	9.a	0.172	0.120	0.168
		9.b	0.139	0.042	0.048
		9.c	0.153	0.116	0.144
		9.d	0.113	0.108	0.099
		9.e	0.117	0.207	0.197
		9.f	0.103	0.140	0.117
		9.g	0.082	0.115	0.076
		9.h	0.122	0.152	0.151

In addition, similar sensitivity analysis was carried out to show the effect of the changes in sub-criteria weights on the final result. Each of the sub-criteria for the first criterion was increased (or decreased) by 10-50 percent and the other sub-criteria weights were adjusted accordingly. The final result produced no significant effect on the final scores and no change in the overall ranking of the alternatives based on the Euclidean distance (Table IX).

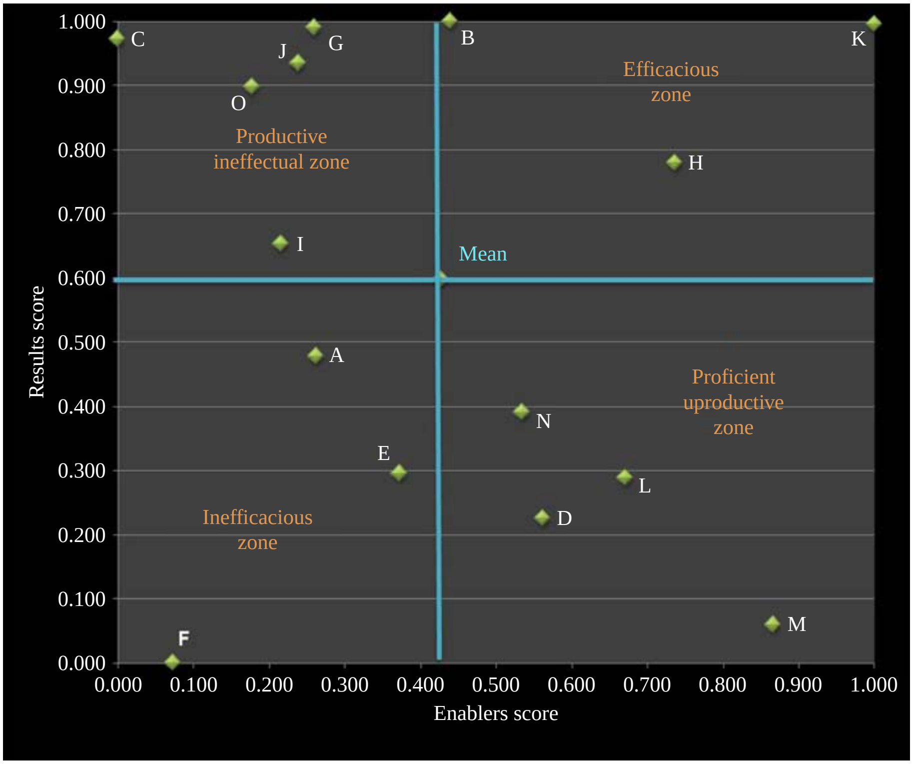

Figure 3.
The results of the benchmarking model

Although we cannot make any generalizations based on the numerical results from one application, our results seem to conform to the previous results obtained in the literature which shows that the pairwise comparison approach in AHP generally results in robust weights.

Some form of validation of the results would be useful from a research perspective. After obtaining the final rankings, we assembled a group of five different inspectors not involved in the implementation of the model. Each of them was familiar with the hospitals and had some opinions based upon their experience and familiarity with the hospitals' operations. After reviewing the results from the pilot study, they felt that the overall results were consistent with the results that they would have obtained using their experience and intuition. However, they also admitted that the structured approach proposed in this study was preferable to the methods that they have used in the past.

5. Conclusions and future research directions

The benchmarking framework presented in this study has some obvious attractive features. First, the generic nature of the framework allows for the subjective and objective evaluation of a finite number of BUs by a group of DMs. Second, the information requirements of the framework are stratified hierarchically allowing DMs to focus on a small area of the large problem. Third, the framework does not dispel subjectivity; it calibrates the subjective weights with the objective weights determined through

Table IX.
The EFQM excellence model overall scores and rankings

Hospital	Enablers score (S_e^k)	Results score (S_r^k)	Category	Euclidean distance (D^k)	Ranking
A	0.262	0.477	Inefficacious	0.55	1
B	0.439	1.000	Efficacious	0.85	7
C	0.000	0.972	Productive ineffectual	0.95	11
D	0.562	0.225	Proficient unproductive	0.78	5
E	0.372	0.296	Inefficacious	0.64	2
F	0.073	0.000	Inefficacious	0.75	3
G	0.259	0.990	Productive ineffectual	0.87	8
H	0.736	0.779	Efficacious	1.12	15
I	0.215	0.653	Productive ineffectual	1.03	14
J	0.238	0.935	Productive ineffectual	0.97	13
K	1.000	0.996	Efficacious	0.77	4
L	0.671	0.289	Proficient unproductive	0.91	9
M	0.866	0.060	Proficient unproductive	0.95	10
N	0.534	0.391	Proficient unproductive	0.96	12
O	0.177	0.898	Productive ineffectual	0.81	6
Mean	0.427	0.597			
Ideal point	1.000	1.000			

the entropy concept. As Russo and Schoemaker (1989) note, considerable research indicates that DMs can maximize their chances of making the best choice(s) if they find a systematic way to evaluate all the evidence favorable or unfavorable to each choice, such as the weighted-sum model described here. Still, in most applied settings, it is not possible to demonstrate the accuracy of the weighted-sum models. In contrast, where the same decision is made repeatedly, data on the outcomes of past decisions are available, and one expects the future to resemble the past. In this setting, objective linear models such as multiple regression can be used to determine the optimal set of predictors. For many decisions, including the one described here, there are no objective outcomes of past decisions. In such situations, rigorous subjective weighted-sum models such as the benchmarking framework proposed here are likely to provide the best hope for optimizing the quality of decisions and the acceptability of those decisions to organizational stakeholders and public.

The overall enablers and results scores in our framework depend heavily on the subjective judgments and ratings provided by the DMs. Therefore, it is imperative that these judgments and ratings be perceived as reasonably accurate and fair. If the rating process is viewed by the stakeholders and public as biased, inaccurate, or contaminated by self-serving motives, then the EFQM excellence model results will be viewed as unfair. In developing subjective judgments and ratings, two types of rating errors can occur. Some rating errors are unintentional. However, some rating errors are intentional and reflect self-serving or political motives. In this case, DMs may have the ability to make accurate judgments, but they are unwilling to do so. DMs can play political games and distort their judgments to achieve a desired goal (Kozlowski *et al.*, 1998). Kozlowski *et al.* (1998) have noted that politics and associated judgment distortions are more likely when:

- there is a direct link between the judgments and desired rewards as in the EFQM excellence model decisions;
- there is a lack of surveillance of DM behavior; and
- there is a widespread perception that others will distort their judgments.

Kozlowski *et al.* (1998) describe several actions that organizations can take to minimize the role of politics in judgments and ratings. These recommendations include:

- having key DMs serve as role models by providing fair evaluations and discouraging political game playing;
- allowing other DMs to suggest potential improvements to the system itself;
- ensuring that the evaluation criteria are widely viewed as relevant;
- using multiple DMs; and
- making DMs accountable for their evaluations by having to explain the reasons for their judgments.

When DMs are motivated to provide accurate judgments, training can enhance the accuracy of the judgments. Hauenstein (1998) reviewed the empirical research in this area and described key elements in successful DM accuracy training. A facilitator should also guide rating sessions to ensure that the same process is applied systematically to all BUs. In general, it has been shown that facilitation enhances the effectiveness of groups using group decision support systems (Khalifa *et al.*, 2002).

Finally, using a structured, step-by-step approach like the proposed benchmarking framework is not intended to imply a deterministic approach to the EFQM excellence model. Determining the degree of excellence in BUs is a complex problem. While the proposed framework enables DMs to crystallize their thoughts and organize data by simultaneously considering both inherently subjective criteria and more objective criteria, it should be used very carefully. As with any decision analysis model, the researchers and practicing managers must be aware of the limitations of subjective estimates.

For future research, it is suggested that researchers study and develop fuzzy MADM approaches when the decision data are unquantifiable or incomplete. The observed values in real-world problems are often imprecise or vague. Imprecise or vague data may be the result of unquantifiable, incomplete and non-obtainable information and can be expressed with fuzzy numbers. The benchmarking framework developed in this study can potentially lend itself to other areas of study. There is a particular interest in extending this model to real-world problems with imprecise, ambiguous or unknown data.

Note

1. All the names and data presented in this study are changed to protect the anonymity of the hospitals and the health care organizations who participated in this project.

References

Belton, V. and Stewart, T.J. (2002), *Multiple Criteria Decision Analysis: An Integrated Approach*, Kluwer Academic, Boston, MA.

Bemowski, K. and Stratton, B. (1995), "How do people use the Baldrige Award criteria?", *Quality Progress*, Vol. 28 No. 5, pp. 43-7.

Conti, T. (1997), *Organizational Self-assessment*, Chapman & Hall, London.

Coulambidou, L. and Dale, B.G. (1995), "The use of quality management self-assessment in the UK: a state of the art study", *Quality World Technical Supplement*, September, pp. 110-18.

Dahlgaard, J.J., Kristensen, K. and Kanji, G.K. (1998), *Fundamental of Total Quality Management*, Chapman & Hall, London.

Dale, B.G. and Ritchie, L. (2000), "An analysis of the self-assessment practices using the business model", *Proceedings of the Institution of Mechanical Engineers*, Vol. 204, B4, pp. 593-602.

Diakoulaki, D., Mavrotas, G. and Papayannakis, L. (2000), "Objective weights of criteria for interfirm comparisons", *Journées du groupe européen Aide Multicritére à la Décision*, Vol. 36e Luxembourg,.

Donnelly, M. (2000), "A radical scoring system for the European Foundation for Quality Management business excellence model", *Managerial Auditing Journal*, Vol. 15 Nos 1/2, pp. 8-11.

Doumpos, M. and Zopounidis, C. (2002), *Multicriteria Decision Aid Classification Methods*, Kluwer Academic, Boston, MA.

EFQM (2003), *European Foundation for Quality Management Excellence Model*, Public and Voluntary Sector, EFQM, Brussels.

Ehrgott, M. and Wiecek, M.M. (2005), "Multiobjective programming", in Figueira, J., Greco, S. and Ehrgott, M. (Eds), *Multiple Criteria Decision Analysis: State of the Art Surveys*, Springer, New York, NY, pp. 667-722.

Eskildsen, J.K., Kristensen, K. and Westlund, A.H. (2003), "The predictive power of intangibles", *Measuring Business Excellence*, Vol. 7 No. 2, pp. 46-54.

Eskildsen, J.K., Kristensen, K., Juhl, H.J. and Østergaard, P. (2004), "The drivers of customer satisfaction and loyalty", *Total Quality Management and Business Excellence*, Vol. 15 Nos 5/6, pp. 859-68.

Eskildsen, J.K., Martensen, A., Gronholdt, L. and Kristensen, K. (1999), "Benchmarking student satisfaction in higher education based on the ECSI methodology", in Baccaroni, C. (Ed.), *Proceedings of the TQM for Higher Education Institutions II, Verona, Italy*.

Festinger, L. (1964), *Conflict, Decision, and Dissonance*, Tavistock, London.

Gouveia, M.C., Dias, L.C. and Antunes, C.H. (2008), "Additive DEA based on MCDA with imprecise information", *The Journal of the Operational Research Society*, Vol. 59, pp. 54-63.

Grabisch, M. and Roubens, M. (1999), "An axiomatic approach to the concept of interaction among players in cooperative games", *International Journal of Game Theory*, Vol. 28 No. 4, pp. 547-65.

Hauenstein, N.M.A. (1998), "Training raters to increase the accuracy of appraisals and the usefulness of feedback", in Smither, J.W. (Ed.), *Performance Appraisal: State-of-the-art in Practice*, Jossey-Bass, San Francisco, CA.

Hwang, C.L. and Masud, A.S. (1979), *Multi Objective Decision Making, Methods and Applications*, Springer, Berlin.

Hwang, C.L. and Yoon, K. (1981), *Multiple Attribute Decision Making: Methods and Applications*, Springer, New York, NY.

Jackson, S. (2000), "Achieving clinical governance in women's services through the use of the EFQM excellence model", *International Journal of Health Care Quality Assurance*, Vol. 13 No. 4, pp. 182-90.

Jackson, S. and Bircher, R. (2002), "Transforming a run down general practice into a leading edge primary care organisation with the help of the EFQM excellence model", *International Journal of Health Care Quality Assurance*, Vol. 15 No. 6, pp. 255-67.

Janic, M. and Reggiani, A. (2002), "An application of the multiple criteria decision making (MCDM) analysis to the selection of a new hub airport", *European Journal of Transport and Infrastructure Research*, Vol. 2 No. 2, pp. 113-41.

Juhl, H.J., Kristensen, K. and Østergaard, P. (2002), "Customer satisfaction in European food retailing", *Journal of Retailing and Consumer Services*, Vol. 9, pp. 327-34.

Karkoszka, T. and Szewieczek, D. (2007), "Risk of the processes in the aspect of quality, natural environment and occupational safety", *Journal of Achievements in Materials and Manufacturing Engineering*, Vol. 20, pp. 539-42.

Khalifa, M., Davison, R. and Kwok, R.C.-W. (2002), "The effects of process and content facilitation restrictiveness on GSS-mediated collaborative learning", *Group Decision and Negotiation*, Vol. 11, pp. 345-61.

Kozlowski, S.W.J., Chao, G.T. and Morrison, R.F. (1998), "Games raters play: politics, strategies, and impression management in performance appraisal", in Smither, J.W. (Ed.), *Performance Appraisal: State-of-the-art in Practice*, Jossey-Bass, San Francisco, CA.

Lascelles, D. and Peacock, R. (1996), *Self-assessment for Business Excellence*, McGraw-Hill, London.

Leyva-Lopez, J.C. and Fernandez-Gonzalez, E. (2003), "A new method for group decision support based on ELECTRE III methodology", *European Journal of Operational Research*, Vol. 148, pp. 14-27.

Lootsma, F.A. (1992), "The REMBRANDT system for multicriteria decision analysis via pairwise comparisons or direct rating", Rep. No. 92-05, Faculty of Technical Mathematics and Informatics, Delft University of Technology, Delft.

Lootsma, F.A. (1993), "Scale sensitivity in the multiplicative AHP and SMART", *Journal of Multi-Criteria Decision Analysis*, Vol. 2 No. 2, pp. 87-110.

Lootsma, F.A. (1996), "A model for the relative importance of the criteria in the Multiplicative AHP and SMART", *European Journal of Operational Research*, Vol. 94 No. 3, pp. 467-76.

Lootsma, F.A. (1997), *Fuzzy Logic for Planning and Decision Making*, Kluwer Academic, Dordrecht.

López, E. and Monzón, A. (2010), "Integration of sustainability issues in strategic transportation planning: a multi-criteria model for the assessment of transport infrastructure plans", *Computer-Aided Civil and Infrastructure Engineering*, Vol. 25, pp. 440-51.

Malorny, C. (1996), *TQM umsetzen: der Weg zur business excellence*, Schäffer-Poeschel, Stuttgart.

Michalska, J. (2008), "Using the EFQM excellence model to the process assessment", *Journal of Achievements in Materials and Manufacturing Engineering*, Vol. 27 No. 2, pp. 203-6.

Moeller, J. and Sonntag, A.K. (2001), "Evaluation of health services organisations – German experiences with the EFQM excellence approach in healthcare", *The TQM Magazine*, Vol. 13 No. 5, pp. 361-7.

Moeller, J., Breinlinger-O'Reilly, J. and Elser, J. (2000), "Quality management in German health care – the EFQM excellence model", *International Journal of Health Care Quality Assurance*, Vol. 13 No. 6, pp. 254-8.

Nabitz, U., Klazing, N. and Walburg, J. (2000), "The EFQM excellence model: European and Dutch experiences with the EFQM approach in health care", *International Journal for Quality in Health Care*, Vol. 12 No. 3, pp. 191-201.

Nazemi, J. (2010), "A process model for improvement through EFQM", *World Applied Sciences Journal*, Vol. 8 No. 3, pp. 279-87.

Olson, D.L. (1996), *Decision AIDS for Selection Problems*, Springer, New York, NY.

Olson, D.L., Fliedner, G. and Currie, K. (1995), "Comparison of the REMBRANDT system with analytic hierarchy process", *European Journal of Operational Research*, Vol. 82 No. 3, pp. 522-39.

Pardalos, P.M. and Hearn, D. (2002), *Multicriteria Decision Aid Classification Methods*, Kluwer Academic, Boston, MA.

Persaud, A. (2002), "Using the EFQM excellence model within health care, a practical guide to success", *International Journal of Health Care Quality Assurance*, Vol. 15 No. 4, pp. 182-3.

Pomero, J.C. and Brba-Romero, S. (2000), *Multicriterion Decision in Management: Principles and Practice*, Kluwer Academic, Boston, MA.

Porter, L.J. and Tanner, S.J. (1998), *Assessing Business Excellence*, Butterworth Heinemann, Oxford.

Roy, B. (1973), "How outranking relation helps multiple criteria decision making", in Cochrane, J. and Zeleny, M. (Eds), *Topics in Multiple Criteria Decision Making*, University of South Carolina Press, Columbia, SC, pp. 179-201.

Russo, J.E. and Schoemaker, P.J.H. (1989), *Decision Traps*, Fireside, New York, NY.

Saaty, T.L. (2000), *Fundamentals of Decision Making and Priority Theory with the AHP*, 2nd ed., RWS Publications, Pittsburgh, PA.

Sánchez, E., Letona, J., González, R., García, M., Darpón, J. and Garay, J.I. (2005), "A descriptive study of the implementation of the EFQM excellence model and underlying tools in the Basque Health Service", *International Journal for Quality in Health Care*, Vol. 18 No. 1, pp. 58-65.

Santos-Vijande, M.L. and Alvarez-Gonzalez, L.I. (2007), "TQM and firms performance: an EFQM excellence model research based survey", *International Journal of Business Science and Applied Management*, Vol. 2 No. 2, pp. 21-41.

Schelling, T.C. (1960), *The Strategy of Conflict*, Harvard University Press, Cambridge, MA.

Stewart, A. (2003), "An investigation of the suitability of the EFQM excellence model for a pharmacy department within an NHS Trust", *International Journal of Health Care Quality Assurance*, Vol. 16 No. 2, pp. 65-76.

Tavana, M. (2002), "Euclid: strategic alternative assessment matrix", *Journal of Multi-Criteria Decision Analysis*, Vol. 11, pp. 75-96.

Teo, W.F. and Dale, B.G. (1997), "Self-assessment: methods, management and practice", *Proceedings of the Institution of Mechanical Engineers, Part B*, Vol. 211 No. 5, pp. 365-75.

Triantaphyllou, E. (2000), *Multi-criteria Decision Making Methods: A Comparative Study*, Kluwer Academic, Boston, MA.

Triantaphyllou, E. and Baig, K. (2005), "The impact of aggregating benefit and cost criteria in four MCDA methods", *IEEE Transactions on Engineering Management*, Vol. 52, pp. 213-26.

Vallejo, P., Saura, R.M., Sunol, R., Kazandjian, V., Ureña, V. and Mauri, J. (2006), "A proposed adaptation of the EFQM fundamental concepts of excellence to health care based on the PATH framework", *International Journal of Quality in Health Care*, Vol. 18 No. 5, pp. 327-35.

Van den Honert, R.C. and Lootsma, F.A. (2000), "Assessing the quality of negotiated proposals using the REMBRANDT system", *European Journal of Operational Research*, Vol. 120 No. 1, pp. 162-73.

Van der Wiele, A., Williams, A.R.T., Dale, B.G., Carter, G., Kolb, F., Luzon, D.M., Schmidt, A. and Wallace, M. (1996), "Self-assessment: a study of progress in Europe's leading organizations in quality management practices", *International Journal of Quality & Reliability Management*, Vol. 13 No. 1, pp. 84-104.

Vincke, P. (1992), *Multicriteria Decision Aid*, Wiley, New York, NY.

Vogt, W. (2001), "The German perspective of using the EFQM model in medical laboratories", *Accreditation and Quality Assurance*, Vol. 6 Nos 9/10, pp. 396-401.

Zeleny, M. (1982), *Multiple Criteria Decision Making*, McGraw-Hill, New York, NY.

Zeleny, M.A. (1974), "Concept of compromise solutions and the method of the disptaced ideal", *Computers & Operations Research*, Vol. 1 Nos 3/4, pp. 479-96.

About the authors

Madjid Tavana is a Professor of Management Information Systems and Decision Sciences and the Lindback Distinguished Chair of Information Systems at La Salle University where he served as Chairman of the Management Department and Director of the Center for Technology and Management. He has been a distinguished research fellow at NASA's Kennedy Space Center, NASA's Johnson Space Center, Naval Research Laboratory – Stennis Space Center, and Air Force Research Laboratory. He was awarded the prestigious Space Act Award by NASA in 2005. He holds an MBA, a PMIS, and a PhD in Management Information Systems. Dr Tavana received his post-doctoral diploma in strategic information systems from the Wharton School of the University of Pennsylvania. He is the Editor-in-Chief for the *International Journal of Strategic Decision Sciences*, the *International Journal of Enterprise Information Systems*, and the *International Journal of Applied Decision Sciences*. He has published in journals such as *Decision Sciences*, *Interfaces*, *Information Systems*, *Annals of Operations Research*, *Information and Management*, *Journal of the Operational Research Society*, *Computers and Operations Research*, and *Advances in Engineering Software*, among others. Madjid Tavana is the corresponding author and can be contacted at: tavana@lasalle.edu

Amir Karbassi Yazdi is a Lecturer at Payam Nour University in Tehran, Iran. He received his BSc in Industrial Engineering from the Azad University, North Tehran Branch and an MA in Industrial Management from the Azad University, South Tehran Branch. His research interests are in operations research, quality management, and strategy and organizational performance.

Mehran Shiri received his MA in Industrial Management from the Azad University, South Tehran Branch. His research interests are in quality management, strategic decision sciences, and organizational productivity and performance analysis.

Jack Rappaport is an Assistant Professor of Production and Operations Management at La Salle University where he teaches courses in Operations Management and Statistics. He holds a BA in mathematics from the University of Pennsylvania and an MS in mathematics from the Courant Institute at New York University. His research interests are in the use of quantitative techniques to solve business problems. He has published in journals such as the *International Journal of Operational Research*, *International Journal of Integrated Supply Management*, *Journal of Financial Research*, and *International Journal of Operations Management*.