Ann Oper Res (2006) 145:301–319
DOI 10.1007/s10479-006-0037-4

Benchmarking with quality-adjusted DEA (Q-DEA) to seek lower-cost high-quality service: Evidence from a U.S. bank application*

H. David Sherman · Joe Zhu

Published online: 27 June 2006

Abstract Benchmarking is a widely cited method to identify and adopt best-practices as a means to improve performance. Data envelopment analysis (DEA) has been demonstrated to be a powerful benchmarking methodology for situations where multiple inputs and outputs need to be assessed to identify best-practices and improve productivity in organizations. Most DEA benchmarking studies have excluded quality, even in service-sector applications such as health care where quality is a key element of performance. This limits the practical value of DEA in organizations where maintaining and improving service quality is critical to achieving performance objectives. In this paper, alternative methods incorporating quality in DEA benchmarking are demonstrated and evaluated. It is shown that simply treating the quality measures as DEA outputs does not help in discriminating the performance. Thus, the current study presents a new, more sensitive, quality-adjusted DEA (Q-DEA), which effectively deals with quality measures in benchmarking. We report the results of applying Q-DEA to a U.S. bank's 200-branch network that required a method for benchmarking to help manage operating costs and service quality. Q-DEA findings helped the bank achieve cost savings and improved operations while preserving service quality, a dimension critical to its mission. New insights about ways to improve branch operations based on the best-practice (high-quality low-cost) benchmarks identified with Q-DEA are also described in the paper. This demonstrates the practical need and potential benefits of Q-DEA and its efficacy in one application, and also suggests the need for further research on measuring and incorporating quality into DEA benchmarking.

Keywords Data envelopment analysis (DEA) · Efficiency · Quality · Benchmarking · Performance · Bank

*The review process of this paper was handled by the Edit-in-Chief Peter Hammer.

H. D. Sherman (✉)
Northeastern University, 404 Hayden Hall, Boston, MA 02468
e-mail: h.sherman@neu.edu

J. Zhu
Department of Management, Worcester Polytechnic Institute, Worcester, MA 01609
e-mail: jzhu@wpi.edu

Introduction

Benchmarking is a method of improving performance by increasing revenues, augmenting quality, and/or reducing operating costs. It compares methods of producing products, providing services, selling and marketing products and services, and managing internal business systems with practices in other businesses. A group of organization units are compared to identify best-practices that can be adopted to improve organization performance on dimensions such as quality, efficiency, and profitability. The comparison groups may include similar operating units and service providers within an organization, competing businesses, and even businesses with similar transactions in other industries (Camp, 1993). Benchmarking is particularly valuable when no objective or engineered standard is available to define efficient and effective performance. For this reason, it is often used in managing services because service standards are more difficult to define than manufacturing standards.

Data Envelopment Analysis (DEA) is a linear-programming-based methodology that has been demonstrated to be effective for certain types of benchmarking (see, e.g., Cooper, Seiford, and Tone, 2000; Zhu, 2003). DEA has been used for benchmarking in many services including healthcare (physicians, hospitals, nursing homes) (Chilingerian and Sherman, 1990; banking (Athanassopoulos and Giokas, 2000; Sherman and Ladino, 1995), and education (Bessent et al., 1984). DEA allows one to compare organizations that use multiple inputs to produce multiple outputs and to measure these outputs and inputs in their natural units, i.e., without converting resources used and outputs into monetary units.

For example, a hospital study can include many outputs such as the number of each type of patient illness treated and number of residents trained, without the need to convert these to common units such as dollars. DEA compares units in the benchmarking comparison set and (i) identifies best-practice units, which are those producing their volume and mix of outputs with the least amount of inputs; (ii) identifies less-productive units, which are using more resources than the best-practice units to provide their volume and mix of services; and (iii) suggests the amount and location of performance improvement potential if the best-practices were adopted by less-productive units. When the less-productive hospital is able to reduce resources by adopting best-practice methods identified with DEA, that hospital will also achieve cost savings which may not have been identifiable using the accounting cost data.

Under the multiple-performance measure context, because it requires very few assumptions for its uses, DEA has opened up possibilities for use in cases which were resistant to other benchmarking approaches because of the complex (often unknown) nature of the relations between the multiple inputs and multiple outputs involved. As pointed out in Cooper, Seiford, and Tone (2000), DEA has also been used to supply new insights into activities (and entities) that have previously been evaluated by other methods. For instance, studies of benchmarking practices with DEA have identified numerous sources of inefficiency in some of the most profitable firms—firms that served as benchmarks solely because of their profitability—and this has provided a vehicle for identifying better benchmarks in many applied studies.

Most DEA benchmarking studies do not consider the quality of the services or products (Callen, 1991). Excluding quality can result in adopting perceived best-practices that improve efficiency and reduce cost, but *also* reduce quality. Quality concerns are critical to many services and products (healthcare, investment management, home appliances, etc.) but are not addressed in most DEA benchmarking studies. Implicitly, these studies assume that the quality of the benchmarked units is equal, that quality is independent of efficiency, or that quality is not relevant to the analysis. One may argue that one can include the quality measures into DEA. The numerical example in the next section, however, shows that the

DEA efficiency usually improves as the number of inputs and outputs increases. We further illustrate that including quality may suggest a tradeoff between quality and productivity that can be contrary to management objectives. For example, spending more time with a customer can result in higher service quality and will reduce the productivity due to the added time servicing the customer and management may encourage this type of tradeoff. However, if the service provider increases his/her productivity by spending less time with the customer resulting in a low customer satisfaction quality rating, this tradeoff may be contrary to management's objectives of retaining customers to ensure continued demand for the business services.

From an accounting perspective, one benefit of DEA is that it enables one to look behind the financial results of an organization and focus on its ability to use a variety of resources to provide or produce a variety of services. For example, many U.S. and European banks measure the profitability of their bank branches (Athanassopoulas and Giokas, 2000). One branch may be more profitable because it has a very large deposit base that generates interest income while another branch may have lower cash deposits because it services a different type of customer. Low deposits are found both in very high-income areas where funds are kept in non-bank privately managed investment accounts and in low-income areas where there are fewer funds available. While high-profit branches are the most desired, high profit does not necessarily mean a branch is well managed. The high-profit branch may be quite inefficient in processing transactions, which may not be apparent from financial measures, and as a result, it could generate even higher profits if it were operating efficiently (Sherman and Ladino, 1995). DEA has been used to benchmark the operating efficiency of branches to identify ways a branch can improve its profitability by improving the way it processes transactions. For example, it can suggest whether alternative methods for using part-time and full-time employees are more or less efficient and whether centralized versus branch handling of night deposits and ATM machines are less costly. This ultimately leads to ways to increase accounting profits that would not have been identifiable by analyzing the accounting cost-and-profit data.

The need to consider quality in benchmarking can also be seen in this banking example. Without explicitly including quality in the analysis, DEA could identify best-practice branches that use fewer resources even if they achieve this by providing low-quality service. Emulating low-cost/low-quality branches could result in loss of valuable customers and reduced profitability over time, a result that bank managers are unlikely to find acceptable. Hence, if we seek high-quality low-cost best-practice benchmark branches, quality needs to be considered.

This paper explores the way DEA can incorporate quality in its benchmarking to help identify high-quality and high-efficiency best-practice benchmarks. (High productivity and low cost are used interchangeably in this paper, suggesting that efficient resource use reduces costs. An efficient unit may not be low cost if it does not pay competitive prices. The price issue will not be considered in this paper). Several alternative DEA benchmarking methods that consider quality are considered, including ones proposed in other DEA studies. A new method designated Q-DEA is suggested as more comprehensive and the results of applying Q-DEA to a bank branch network are reported. The advantages and weaknesses of Q-DEA are identified, providing a potential basis for developing more comprehensive and effective DEA benchmarking methods.

The paper proceeds as follows: The next section compares alternative ways DEA might incorporate quality and illustrates these alternatives through a benchmarking application to a simplified data set. We then describe the results of applying the suggested Q-DEA evaluation to a multi-state bank branch network and the reactions of management to these results. The benefits and limitations of using Q-DEA for benchmarking are reviewed and the final section identifies several key issues where additional research is warranted.

1. Incorporating quality into DEA benchmarking

In this section, we use a simplified version of the real data set to illustrate the approach. The example, presented in Table 1, is designed to illustrate and contrast DEA benchmarking results using several different analytical models. Specifically analysis of this data set will (i) illustrate the impact of excluding quality; (ii) illustrate the benefits and weaknesses of incorporating quality as suggested in several DEA studies; and (iii) describe the advantages and limitations of a new benchmarking approach that adjusts DEA for quality and will be referred to as Q-DEA.

The example in Table 1 has the following characteristics (a graph plotting resources used by each branch is given in Fig. 1):

(i) Each branch processes the same number of identical transactions. This simplified assumption is naturally expanded to multiple transactions in the application in Section 3. Although DEA can and does accommodate multiple outputs in natural units, we use one type of output and one output level to facilitate graphic illustration of the approach.
(ii) Branches use two types of resources: teller (labor) hours to complete the transactions and non-personnel expenses such as costs of supplies, utilities, etc., to support the tellers completing the transactions. The amount of hours used and expenses are reported for each branch.
(iii) The quality of service is measured for each branch. One typical quality measure is a mystery shopper score based on visits by artificial customers employed by survey companies. They complete a questionnaire measuring service on predetermined quality measures and provide an overall rating for each teller and the entire branch. The highest quality rating is 100. In this example, Branch I has the lowest quality rating of 60.
(iv) Branches A through E use the identical levels of resources as Branches G through K. For example, Branch A uses the same resources as G (20 hours and $300 of other expenses). Similarly, B and H, C and I, D and J, and E and K have the same resource use and the

Table 1 Bank branch example

Branch	Transactions	Quality (maximum = 100)	Teller hours	Non-personnel expenses
A	1000	100	20	300
B	1000	100	30	200
C	1000	100	40	100
D	1000	100	20	200
E	1000	100	10	400
F	1000	80	20	150
G	1000	90	20	300
H	1000	70	30	200
I	1000	60	40	100
J	1000	90	20	200
K	1000	80	10	400

Branches A & G, B & H, C & I, D & J, E & K use the same resources. The quality scores among these pairs of branches differ.

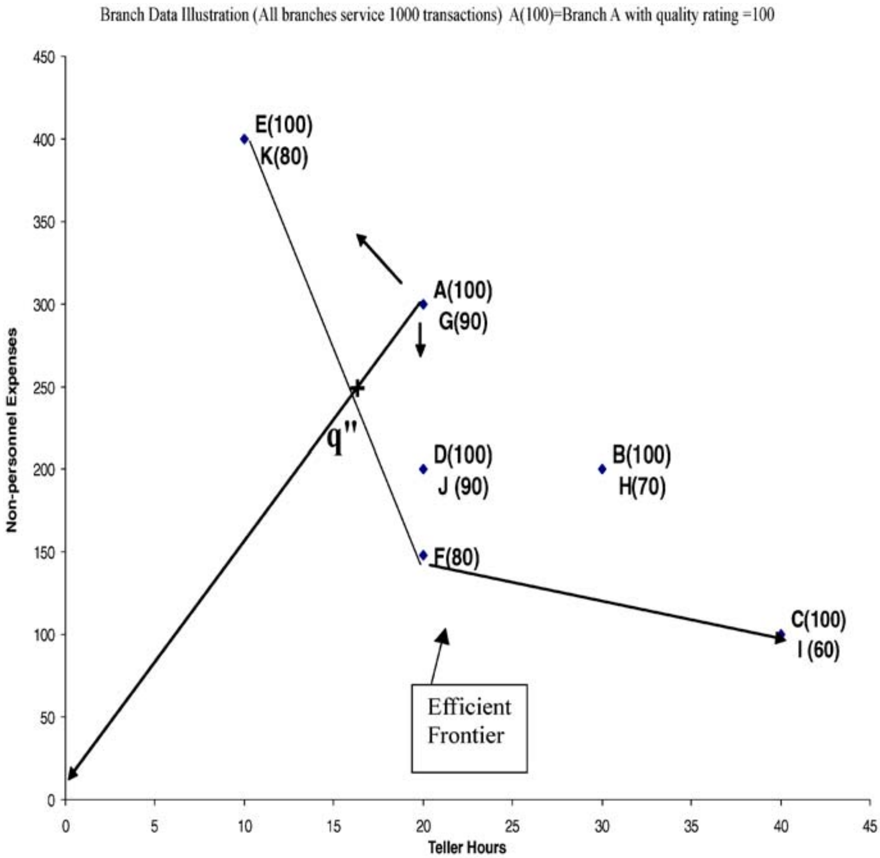


Fig. 1 Branch Data Illustration

same transaction levels. The only difference within these branch pairs is their quality level. For example, Branch A has a quality rating of 100 while Branch G has a quality rating of 90. Hence, they are equally productive but Branch A has higher service quality and would be considered a better model branch than Branch G.

(v) All the data are assumed to be accurate and represent the appropriate measures of quality, transactions, and resources used.

Questions a manager of this bank might ask include: How are these branches performing? Which are the best-practice branches? Could the network performance be improved if all branches emulated the best-practice branches? DEA-based benchmarking is one way to begin to address these questions.

Model I. Standard DEA Model

If one applied the standard DEA model, e.g., the CCR model of (Charnes, Cooper, and Rhodes, 1978), to benchmark these branches using the approach found in many DEA studies, one would run a model with one output (transactions), and two inputs (hours and other expenses). The result would be a DEA rating where the highest productivity rating is 1,

Table 2 Model I—Benchmarking productivity with DEA *excluding quality*

		Potential resource reductions		DEA benchmark reference set			
Branch name (quality rating)	DEA benchmark productivity rating	Teller hours	Non-personnel expenses	Reference branch (quality rating)	Reference branch weight	Reference branch (quality rating)	Reference branch weight
A (100)	0.81	4	56	E (100)	0.375	F (80)	0.625
B (100)	0.73	8	55	C (100)	0.091	F (80)	0.909
C (100)	1	0	0				
D (100)	0.93	1	14	E (100)	0.143	F (80)	0.857
E (100)	1	0	0				
F (80)	1	0	0				
G (90)	0.81	4	56	E (100)	0.375	F (80)	0.625
H (70)	0.73	8	55	C (100)	0.091	F (80)	0.909
I (60)	1	0	0				
J (90)	0.93	1	14	E (100)	0.143	F (80)	0.857
K (80)	1	0	0				
Total potential resource reduction		27	250				

signifying best-practice. A rating below 1 signifies that the unit is inefficient compared to other units in the study.

The results of this DEA benchmarking analysis are presented in Table 2, where Branches C, E, F, I and K are the best-practice benchmarks, indicated by the DEA benchmark productivity rating of 1. Note that branch F is a benchmark for all inefficient branches. The fact that F has a lower quality rating (80) than A and B (each is 100) is not considered in this analysis. If management were to have branches emulate the efficient Branch F, the network might adopt methods that reduce operating costs but also lower the branch service quality, which could ultimately be costly and dysfunctional.

A key characteristic of DEA as a benchmarking tool is the following: If any inefficient branch (DEA rating <1) adopts best-practice branch methods *without* damaging quality, that branch would reduce its resource use and operating costs, improving its profitability regardless of the cost per teller hour.

Model I, frequently found in DEA studies, illustrates one limitation in this DEA analysis: quality is not considered. Any effort to use this information without considering quality has the potential to reduce service quality, which may have significant effect on future costs.

In one DEA study of a bank-branch network in Greece, Athanassopoulos (1997) considers the relationship of the DEA productivity scores with quality. An independent quality measure is developed based on customer surveys and the statistical relationship between quality and the outputs in the DEA model, such as new accounts and loans, is measured. This suggests the extent to which quality is impacting sales of services. The DEA scores are calculated as above—model I–with *no* quality adjustment. A concept of effort effectiveness is proposed that would be a function of efficiency as measured with DEA and quality. This study does not actually combine these two key elements into this effort effectiveness measure. Rather, it acknowledges the void and calls for research to find ways to combine quality and efficiency—an objective of this paper.

Table 3 Model II—Benchmarking with quality as an output

Branch name (quality rating)	DEA benchmark productivity rating	DEA benchmark reference set					
		Reference branch (quality rating)	Reference branch weight	Reference branch (quality rating)	Reference branch weight	Reference branch (quality rating)	Reference branch weight
A (100)	0.86	D (100)	0.714	E (100)	0.286		
B (100)	0.86	C (100)	0.286	D (100)	0.714		
C (100)	1						
D (100)	1						
E (100)	1						
F (80)	1						
G (90)	0.82	D (100)	0.143	E (100)	0.357	F (80)	0.500
H (70)	0.73	F (80)	0.901	I (60)	0.091		
I (60)	1						
J (90)	0.96	D (100)	0.417	E (100)	0.083	F (80)	0.500
K (80)	1						

Model II. Quality as an Output in Standard DEA Model

We now demonstrate that directly adding quality as a DEA (output) measure does not help in discriminating the performance. If we include the quality measure as a second output in the DEA Model I, the efficiency of each branch will be improved, as evidenced in Table 3. Note that Branch D becomes efficient. It is theoretically true that the DEA score will not be decreased if additional output(s) and (or) input(s) are included.

Branches F and I are designated best-practice branches with model I and II even though they have low quality scores of 80 and 60, respectively. For some applications, this result may be problematic. The lower service-quality rating of 80 was not low enough to disqualify Branch F as a best-practice branch. Here, high productivity compensates for low quality. Should the manager of the service units being benchmarked accept this or any tradeoff between quality and productivity? If we can get this kind of tradeoff information, then we can use the cone ratio DEA model (Charnes et al., 1989) to refine the DEA efficiency. (One study that includes quality as an output in the context of airplane maintenance is Rouse, Putterill, and Ryan, 2002).

Service quality is a complex dimension in itself, however, and its importance varies in different industries. What should the tradeoff be between productivity and quality when selecting the surgeon to do a cardiac-bypass operation? Does the tradeoff between cost and quality of the physician change for less-severe types of medical care? What is the cost/quality tradeoff in selecting a public or private school? It is very difficult to quantify this type of tradeoff information.

On the other hand, quality/productivity and quality/cost tradeoffs are not readily acceptable in many applications. Even in this example, would one want to ask Branches A and B to emulate lower-quality Branch F, which could mean reduced costs but also reduced service quality? Would a manager advocate reducing service quality? Is there a clear enough measure of the value of service quality to be able to quantitatively define economically attractive tradeoffs?

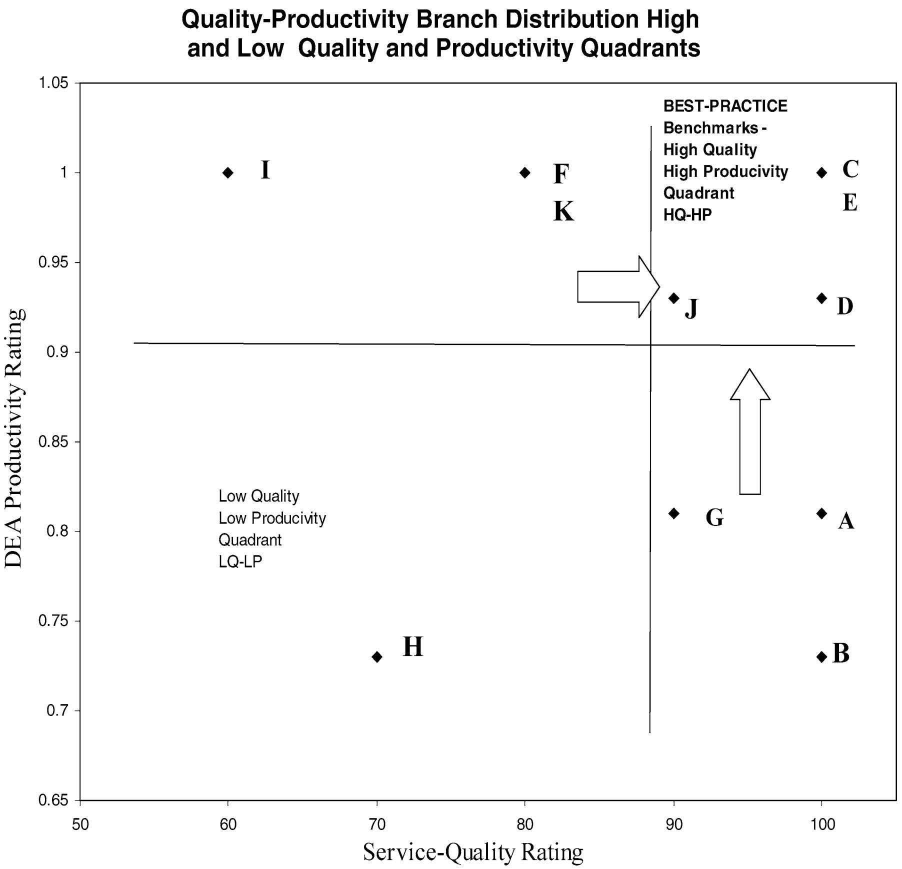


Fig. 2 Model III

Model III. Independent Quality and Productivity Dimensions

One model that avoids an automatic quality/cost tradeoff is to treat quality as a dimension independent from productivity and benchmark on these two dimensions simultaneously. The best-practice would be defined as high quality low cost or high quality high productivity (HQ-HP). This approach, Model III, maps the benchmarked units in terms of quality and productivity. The graph of these branches in Fig. 2 reflects the way quality and productivity are addressed in the other three DEA studies that consider quality, by Bessent et al. (1984) in public schools, by Chilingerian and Sherman (1990) in physicians, and by Soteriou and Zenios (1999) in banks.

This chart can be separated into segments to reflect the way a manager chooses to define high quality and high productivity. For example, in Fig. 2, the graph is broken into quadrants: high quality and high productivity (HQ-HP), low quality and low productivity (LQ-LP), low quality and high productivity, and high quality and low productivity. The advantage of model III is that it identifies branches in the HQ-HP quadrant that can serve as best-practice benchmarks for other branches. It culls out Branch I as the branch that is efficient but provides the lowest-quality service. The quality and productivity dimensions are measured independently in this analysis.

To identify the most appropriate benchmark branches, management would establish a range of quality and productivity that meets its operating objectives. In this example, quality levels of 90 and above are marked as acceptable benchmarks and productivity levels of 0.90 and above are marked as acceptable. This would identify Branches C, E, D, and J as best-practice benchmarks that are considered high-quality/high-productivity branches. Branches F, K, and I are no longer benchmark branches, in contrast to Model I where these were best-practice branches. These three branches might be challenged to improve their service quality even if this required added resources. Similarly, Branches G, B, and A might be expected to improve productivity without sacrificing their high service quality. Model III incorporates quality and does not allow for the quality/cost trade-off seen in Model II where quality is included as an output. For example, Branches F and K had high ratings with Model II because their productivity was high enough to offset and possibly disguise low quality that is apparent in this analysis. Model III singles out Branch H as the weakest performer, the only branch in the low-quality low-productivity quadrant and one that would likely be expected to improve both quality and productivity to move to the HQ-HP quadrant.

Model III has the advantage of making quality/productivity tradeoffs more visible and manageable. It allows management to identify benchmarks that will move the organization to higher quality and productivity.

Note that Model III includes low-quality/high-productivity branches in the set of reference branches which *under-values* the productivity of the other branches and overestimates the amount of resource savings that may be possible as the branches move to the best-practice benchmark levels. Adjusting the DEA productivity analysis to remove the impact of low-quality branches is the objective of Model IV—designated quality-adjusted DEA (Q-DEA).

Model IV. Quality-Adjusted DEA—Q-DEA

To filter out the impact of low-quality branches, a multi-stage DEA analysis is proposed, essentially reapplying DEA to an adjusted data set to achieve a particular objective. The objective here is to develop a DEA analysis where all the most productive branches (DEA rating of 1) are also high-quality branches, resulting in a set of high- quality/high-productivity benchmarks and a target productivity level that is not skewed by low-quality/high-productivity units.

We develop the following algorithm to execute the Q-DEA:

Step 1. Run DEA Model I.
Step 2. If the number of high-productivity/low-quality (HP-LQ) units = 0 then stop. Otherwise, remove the HP-LQ units and go to step 1.

In the branch illustration, the quality-adjusted DEA rating is achieved by removing the branches with high DEA rating of 1 and unacceptable quality levels. In this case, Branches I, F, and K are eliminated from the analysis, as they have DEA ratings of 1 but are below the minimum quality standard, in this case below 90. The result of this analysis is reflected in Tables 4 and 5.

Table 4 indicates that by removing the low-quality/high-productivity branches from the analysis, the remaining inefficient branches have higher productivity scores, and Branches D and J become best-practice branches. Each branch on the new best-practice frontier line now provides service that is above the target quality level of 90, i.e., all benchmark best-practice branches are high quality and low cost. Less-productive branches are identified based on a comparison with high-quality/low-cost benchmarks. Hence, a manager of a branch with

Table 4 DEA productivity ratings

Branch name	Quality	DEA unadjusted for quality (Model I)	Quality adjusted DEA (Model IV)	Branches with increased productivity rating with Q-DEA
A	100	0.81	0.86	←==
B	100	0.73	0.86	←==
C	100	1	1	
D	100	0.93	1	←==
E	100	1	1	
F	80	1	Low quality	
G	90	0.81	0.86	←==
H	70	0.73	0.86	←==
I	60	1	Low quality	
J	90	0.93	1	←==
K	80	1	Low quality	

High-productivity branches with low service quality identified in Model I are removed from the data set and DEA is rerun to determine if the new high-productivity benchmark branches are also high quality. If any of the high-productivity branches in Model IV, Branches D, C, E, and J, were low quality, they would be removed and the Q-DEA benchmarking would be rerun until this process generated a set of branches with a DEA rating of 1 and high-quality scores.

a low score cannot question the rating on the grounds that the reference set of benchmark branches are sacrificing quality to achieve lower costs. Bank management can reasonably question why branches with low Q-DEA benchmark scores cannot achieve the best-practice levels.

Table 5 indicates the expected savings and the benchmark reference branches for each branch in the Q-DEA analysis. Branch A is now identified as having a productivity rating of 0.86 compared with benchmark branches D and E, both of which have quality scores

Table 5 Q-DEA benchmarking

		Potential resource reductions		DEA benchmark reference set			
Branch name (quality rating)	DEA productivity Rating	Teller hours	Non-personnel expenses expenses	Reference branch (quality rating)	Reference branch weight	Reference branch (quality rating)	Reference branch weight
A (100)	0.86	3	43	D (100)	0.714	E (100)	0.286
B (100)	0.86	4	29	C (100)	0.286	D (100)	0.714
C (100)	1	0	0				
D (100)	1	0	0				
E (100)	1	0	0				
F (80)	Low quality						
G (90)	0.86	3	43	D (100)	0.714	E (100)	0.286
H (70)	0.86	4	29	C (100)	0.286	D (100)	0.714
I (60)	Low quality						
J (90)	1	10	0				
K (80)	Low quality						
Total potential resource reduction		24	143				

of 100. This is in contrast to Model I, where branch A was compared with Branches F and E, and where Branch F had a lower quality rating of 80 (see Table 2). In addition, the expected resource savings are measured based on the HQ-HP benchmarks. In this case, Q-DEA suggests that Branch A has the potential to reduce teller hours by 3 hours (versus 4 with no quality adjustment) and reduce expenses by $43 (versus $56 with no quality adjustment). The objective is to achieve lower-cost operations by reducing excess resources while maintaining or improving quality through benchmarking. The total potential savings suggested by the Q-DEA analysis for this branch set is 24 teller hours and $143 non-personnel expenses (Table 5), which is lower than the amount suggested with no quality adjustment in Model III (27 teller hours and $250 non-personnel expenses in Table 2). In this example, with only 11 DMUs, the Q-DEA algorithm offers new insights into which units have high productivity and high quality. However, this data set is designed to make these differences clear. Applying this algorithm to small data sets may result in less significant or even no added insights.

2. Q-DEA benchmarking with application to a bank branch network

A U.S. bank with over 200 branches distributed over five states was interested in rationalizing their branch network and reducing operating costs by adopting best-practices among their branch network. Prior to this study, the branch performance was evaluated within regions. No branch wide benchmarking was in use and management of branches was primarily the responsibility of regional managers. This study transpired at a time when there was pressure on the bank to reduce operating costs to boost earnings both to satisfy investment analyst demands and to maintain a strong position for acquisition of other banks. Service quality was considered very important due to intense competition in most of their geographic markets and the value placed on the branch network in retaining customers, generating fees, and supplying low-cost demand deposits. Having read publications about this technique, the network manager wanted to evaluate the possible use of DEA as a benchmarking tool to reduce operating costs.

Banks in the U.S. generally have three types of personnel in a branch: tellers, platform, and management. Tellers handle most high-volume standard transactions including deposits, withdrawals, and bank checks. Platform personnel handle more complicated and customized transactions such as loan applications, new and closed accounts, wire transfers, and retirement (IRA) and other investment accounts.

This bank had been using a staffing model that assigned tellers based on total teller transactions and peak demand periods. It was not sensitive to the mix of transactions. In addition, the staffing model applied only to teller transactions and did not consider total resources used by the branch. Another serious limitation of the staffing model was that it ignored the activities of branch management and platform personnel in the branch. The objective was to use DEA benchmarking to uncover new ways to improve branch operations.

The information systems of the bank were relatively advanced and captured branch transactions in detail. Branch performance measures included a profit-and-loss statement that focused primarily on expenses, net interest earned on deposit, and fees generated by the branch. Quality of service was measured based on evaluations by mystery shoppers posing as customers. The maximum score was 100. The issues of customer retention and limitations of mystery shopper scores were well understood by branch management. They believed that these mystery shopper scores were a good measure of quality and approved of this method of service evaluation. Other measures of service quality in banking are discussed in Athanassopoulos and Giokas (2000).

Table 6 Branch data used for Q-DEA benchmarking

Inputs:
1. Platform full-time equivalents (FTEs)
2. Teller FTEs
3. Management FTEs
4. Postage, supplies, telephone, travel expenses

Outputs:
1. Deposits, withdrawals, checks cashed
2. Bank checks
3. Bond transactions
4. Night deposits
5. Safe deposit visits
6. New accounts—time and demand deposits, certificates of deposit, IRA, safe deposit boxes
7. Mortgage and consumer loans—applications, closings
8. Automatic teller machines—serviced by branch

Quality Measure—Mystery shopper scores—branches below 90 out of maximum score of 100 were deemed disqualified to be best-practice models (cutoff was approximately the network average score).

Branch classification schemes:
Urban, rural, suburban
commercial, personal
Transaction volume—size classification
Deposit dollar—size classification
Vacation, retirement, community development (low-income urban), supermarket

Transactions weights were used to increase the power of the DEA analysis.

To develop the DEA model that would be used to benchmark the branches, the regional branch managers were asked to define the transactions handled by their branches, to indicate branches that had unusual activities or characteristics, to outline the types of resources used by the branches, and to identify branches that had unusual working hours and high-demand periods. Based on this information, the inputs and outputs to be used were established as summarized in Table 6. Various branches were excluded from the analysis because they were new, closing, or were special branches designed to handle unusual transactions or transaction volume. The first DEA analysis began with data on the customer-service quality, four inputs including personnel and operating expenses, and eight types of transaction outputs covering teller and platform activities for 225 branches.

Quality of service was a key issue, as management intended to challenge low-scoring branch managers to move to best-practice operating levels found in other branches or to justify their existing resource levels. The best-practice branches that would be the basis for this challenge needed to be high quality and high productivity (low cost). The idea was to move all branches to the high-quality/low-cost best-practices in the branch system. Several characteristics made the use of Q-DEA appear particularly appropriate beyond its ability to incorporate quality: The geographic separation and regional organization meant that some of the benchmark branches would not be familiar to other branch managers; and there were too many branches to allow management to qualitatively assess the interactions of quality and cost of operations. The DEA model used was the CCR-DEA model, which does not adjust for economies of scale. This means that large branches can be best-practice benchmarks for small branches and vice versa. There is no strong external evidence that branches have increasing or decreasing returns. To determine whether this was a potential issue in practice, separate DEA analyses of different branch types and sizes were completed before the results were used to influence branch operations. The only place where this altered the results was for small branches, where there are minimum staffing levels that are required to meet basic operational controls. The productivity ratings and target resource levels for small branches used for the small-branch analysis was used in guiding branch managers on the improvement targets. The small-branch analysis also leads to devising a new model for the efficient high-quality branch (O'Keefe, 1994).

2.1. Phase 1: Improve branch network quality

The data on 225 branches were accumulated and run with the inputs and outputs in Table 6 using the multi-stage Q-DEA process described in Section 2 of this paper. The input measures were the average number of FTEs for each type of employee and total expenses for the quarter and the output measures were the number of transactions for the quarter. Management was asked to select a cutoff point to disqualify benchmark branches with unacceptable quality. The average quality of the branch system was about 88, and management determined that any branch with quality scores below 90 would be disqualified as a benchmark branch. The Q-DEA result after several iterations culling out low-quality high-productivity reference branches was that 32 branches were identified as highly efficient but with low quality. If these were allowed to serve as benchmarks, there was the potential that the low-scoring branches would be challenged to achieve a target operating resource level based on branches that provided low quality service.

The initial reaction of the network manager was that this was a surprisingly large number of disqualified low-quality branches. His view was that these branches needed to improve their quality and be re-incorporated into the analysis. Essentially, the use of the benchmarking to reduce costs was deferred to deal with service quality, which was a higher priority. A program to focus on improved service quality was implemented and the branches were asked to focus on improving service quality.

2.2. Phase 2: Use Q-DEA to reduce branch network operating costs

Benchmarking was resumed about $\frac{1}{2}$ year after the program on quality improvement was initiated. Another data set was developed for a subsequent quarter using the same data elements with 229 branches. The net increase in branches reflects some branches closing and others that were just opening during Phase 1 but were fully operating and included in the second analysis. Average quality of the network increased to about 90 and branches with DEA efficiency scores of 1 with quality below 90 were disqualified as best practice branches. The results of the second Q-DEA analysis was that eight branches were low quality/high productivity. Management considered this to be more acceptable as a starting point for the benchmarking analysis and separately asked that these eight branches be studied to determine if their operating cost levels were too low and were causing the low service quality.

At this point, the bank had a listing of the best-practice branches in terms of high-quality and low-cost operations. The breakdown of branches is summarized in Table 7.

For each branch, the following information was generated: the DEA rating, the target resource savings, the benchmark branches that comprise the DEA efficient reference set of branches, and the weight assigned to each branch. This is similar to the information in Table 5. The Q-DEA results identified 46 best-practice high-quality/high-productivity branches, 8 low-quality/high-productivity branches and 175 branches with DEA ratings less than 1

Table 7 Q-DEA benchmarking applied to a U.S. branch network

	Phase 1: Raise quality			Phase 2: Increase profitability		
	# of branches	Average quality rating	Profitability ranking	# of branches	Average quality rating	Profitability ranking
Best practice	60	90.5	2	46	92.1	2
Less productive	133	89.1	3	175	90	3
Low-quality/ High-productivity	32	80.7	1	8	77.4	1
Total	225			229		

Note that the best-practice branches are more efficient, more profitable (partly due to efficiency) and have higher-average quality than the less-productive branches. The low-quality/high-productivity branches have the lowest average quality but are also the most profitable, possibly because personnel resources are not sufficient to provide high-quality service.
Management's reaction to Phase 1 was to focus on raising service quality. Phase 2 begins about 1/2 year after phase 1 and quality scores are higher with fewer low-quality/high-productivity branches.

Table 8 Q-DEA benchmarking distribution of productivity ratings in Phase 2 in the U.S. bank application

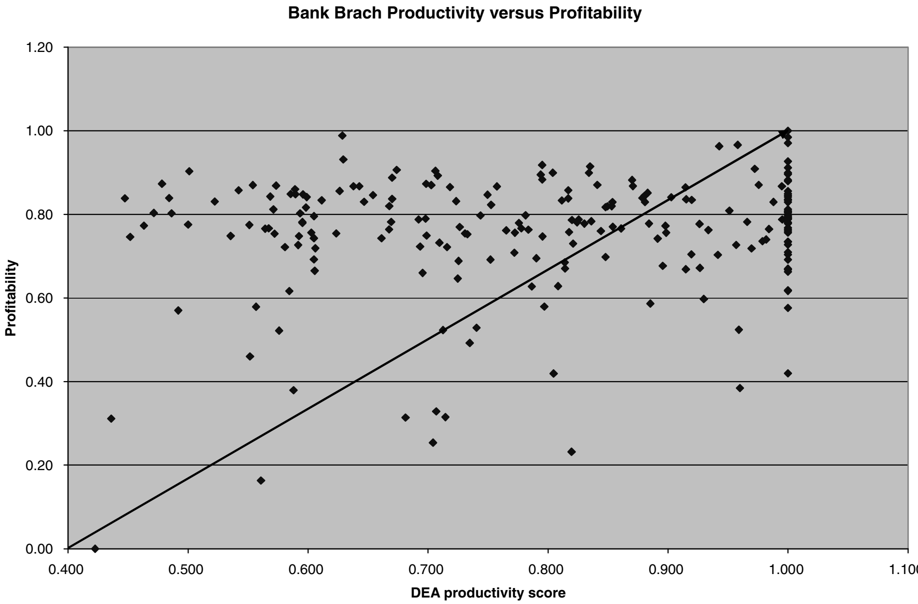
Q-DEA Productivity rating	# of branches	
1	46	Best practice—High-quality/High-productivity
		Less-productive branches
.90 – .99	28	
.80 – .89	41	147 branches
.70 – .79	37	with ratings below .90
.60 – .69	27	
.50 – .59	30	42 branches with ratings below .6
.40 – .49	12	Suggests that they are using 40% more resources than best-practice benchmark branches.

that had potential to reduce costs by adopting best-practice operating methods. Of the 175 branches with some inefficiencies, 147 had DEA ratings at or below 0.9 (see Table 8).

Various tests and analyses were completed by branch management to ensure that these results appeared to make sense. The 45 best-practice branches were reviewed to determine if there were any characteristics that would make them abnormally efficient or poor benchmarks. This included reviewing the results with regional managers and soliciting their qualitative reaction as to whether these were considered model branches. The least-efficient branches were also reviewed to determine if these were already known to be poor performers and whether there were characteristics that would handicap these branches.

The initial reaction was that Q-DEA provided a perspective that differed from the methods already in use and that the results were surprising in many ways. Examples of surprises in this initial stage included the fact that the 46 best-practice branches included a wide variety of branch types. Small branches, vacation community, low-income urban, and retirement community branches were among the best-practice branches. These branch types were thought to be potentially the least efficient because of seasonal staffing requirement, slower transaction times, low-income and multi-lingual customers, and difficulty in staffing small branches due to the need for a minimum staff level to maintain adequate financial controls over branch transactions. Similarly, many of the largest branches were identified as being inefficient and several of the branches that were considered to be the best performers in terms of profitability were among the lowest-rated branches.

The following chart shows a plot of profitability vs. DEA productivity scores.



There is no clear correlation between profitability as measured by the bank and productivity, which illustrates the ability of DEA to locate excess operating costs that are not visible from the accounting measures. The primary reason is that the bank measures branch profits based on revenues from fees and attributed net interest income from deposits generated by the branch less operating expenses. The operating costs are also included in the DEA analysis but these costs are compared with the activities of the branch in servicing customers such as deposits, visits to safe deposit boxes etc. and these are not reflected in the profitability. This is also the reason DEA complements the accounting profit analysis—both used together are most effective.

While these results were surprising, scrutiny of the data supported the Q-DEA results and management decided to proceed to the next step.

The Q-DEA results were distributed and explained to each regional manager in charge of a group of branches. The meaning of each element and the interpretation on a branch-by-branch basis was explained. The regional managers then met with each branch manager that had a DEA score of .90 or lower to focus on branches where the greatest excess resources are likely to be located. The branch managers were asked to consider the analysis, compare their operations with branches that were identified as best-practice benchmarks for their branch, and determine if and how they could alter their operations to move toward the types of branch procedures, staffing, and operating costs that characterize the best-practice benchmarks. One can visualize this process using the example in Fig. 1. Essentially, the managers of Branches A and G were challenged to reduce their resources while maintaining or improving their quality by moving toward Branch E, D, or J, or toward the point "q" based on what appears feasible for each branch manager knowing their business environment. This information and challenge was provided to each of the 147 branches with Q-DEA scores below 0.90.

2.3. Results of Q-DEA benchmarking

Management used the Q-DEA results to focus branches on identifying ways they could reduce operating costs to the best-practice levels. To rely on these analyses, management needed the confidence that there were real best-practice benchmark branches in their network that provide the same or higher-quality service with fewer resources than branches with low DEA scores. They used this to challenge the other branches to improve their performance. If the results were not quality adjusted, management would not have used these results.

The Q-DEA process generated substantial cost savings, a wide variety of responses and several centralized initiatives. Some of the responses are in reaction to very specific findings reflected in the DEA results that the bank had not identified with other performance measurement methods in use. Examples of the findings are described below.

The potential and actual resource savings from the Q-DEA benchmark analysis are summarized in the following table:

	FTEs	% of all branch FTEs	Non-personnel expenses	% of all branch non-personnel expenses
Potential resource savings (Q-DEA benchmark rating <1.0)	422	21%	$1.3 million	28%
Actual Savings (within 6 months of Q-DEA analysis)	149	7.4%	Not measured[a]	

[a] The largest expense impact was the identification of high telephone charges in two of the state branch networks compared to telephone charges in other states. This comparison was never made until the Q-DEA benchmarking. In one of these high cost states with a system of over 30 branches, new phone arrangements were negotiated triggered by this analysis. In the other high-cost telephone state, there were fewer than 20 branches and no changes in the phone system resulted.

FTEs were reduced by 149, or over 7% of the network staff. These reductions were all generated from the branch level by the branch managers working with their district managers (in charge of a group of branches). The changes in staffing were based on their analysis of the Q-DEA benchmark results. Savings were generated among all branch types including high- and low-deposit branches, and high- and low-transaction volume branches. The reductions realized were all believed to be feasible by the branch managers but were not identified before the Q-DEA benchmarking. Branch managers were using benchmarks to reduce excess resources. These savings are believed to be more likely to endure and less likely to damage service quality compared with more common across-the-board staff reductions. The endurance and long-term impact on service quality were not measured.

One group of branches where management was aware that Q-DEA potential staff reductions exceeded 60 FTEs reported that they could reduce their staff by only 6 FTEs. The regional manager reported that added savings were not possible due to changing business conditions and the need for these resources to build their business. *This was not challenged* even though it was not likely to be accurate, in part because these branches had the reputation as being the most profitable. Q-DEA analysis clearly identified significant excess resources in these branches. Bank politics prevented further challenge of this decision, though there was recognition that if the business growth did not materialize, the staffing would be re-evaluated.

Small branches. Of particular interest was the set of small-branch best-practice benchmarks. The bank was considering closing smaller branches and based on these findings began to develop a new small-transaction branch model that incorporated the part-time and flexible hours that the best-practice small branches had experimented with. A separate Q-DEA analysis was run benchmarking only small branches to refine the small-branch model being developed based on small best-practice branches. One of the senior bank managers was interviewed in a bank trade magazine and he commented on this Q-DEA process *before* the above savings were achieved. He noted that "... we found there's some very small productive branches, and we found a lot of large branches that were not very profitable."... "Many banks have small branches and they are looking for solutions other than just closing them." (O'Keefe, 1994). The way small branches were measured was altered as well. In addition to the deposit size, transaction size was considered. Much of the excess FTEs from small branches were found in small branches with low transaction levels that previously were automatically given the same FTEs as other small branches.

Large-deposit and high-transaction branches. Another bank-wide response to Q-DEA benchmarking was the recognition that large branches with high deposits appeared very profitable but were using more than 1/3 of the excess FTEs located in the study, suggesting that they should be even more profitable.

Part-time FTEs. Several best-practice branches were able to handle transactions efficiently with high quality service by allowing tellers and platform personnel to assist others during peak periods or when individuals were absent **and** by aggressive use of part-time personnel. It was determined that many low-productivity branches were not using part-time personnel and that this was due to limited ability to attract and hire this type of employee. The bank subsequently adopted a program to provide enriched health plans and other benefits to part-time employees to enable these branches to hire and retain part-time personnel with the objective of operating with a more flexible staff, resulting in a net decrease in FTEs.

Branch-by-branch changes. There were numerous other changes that occurred from individual branch analyses. Two examples of the kinds of insights generated from these benchmarks that had formerly escaped the attention of the branch managers, regional managers, and network management follow.

Two branches in geographically distant regions were compared based on the Q-DEA reference branch listing. It was noted that one branch had excess FTEs but was otherwise similar to the best-practice branch. The branch managers discussed their operations and determined that they both had drive-up teller windows that were detached from the branch. The best-practice branch had converted this to a drive-up ATM to eliminate the need to staff the drive-up window and this was received well by their customers. The idea had not been considered by the low-rated branch. Following this analysis, the low-rated branch changed its drive-up window to a drive-up ATM.

One low rated branch reflected no excess FTEs but excess operating expenses. The regional manager could not explain this at the introductory meeting. After meeting with the branch managers and investigating this expense, the bank identified billing errors dating back over one year. The branch was paying for its phones and another phone line in the same building that was used by another business. Among the 147 inefficient branches, there were many similar findings unique to individual branches.

3. Conclusion and future research

Many of the DEA applications are in services such as education and healthcare where quality is at least as important, and probably much more important than in the bank application. Even in the bank application, if service quality had not been explicitly included in the analysis, management would not have used the findings. The value of using DEA as a tool to improve profitability (or reduce operating costs for nonprofit users) can be increased by incorporating quality into the benchmarking analysis. The current paper proposes an integrated approach to incorporating quality measures into DEA efficiency. The resulting Q-DEA goes beyond the very few DEA studies that have explicitly studied quality and focuses on the importance of the quality dimension.

The quality measure used in the bank application was the one bank management chose to use and continues to rely on. This is only one dimension and one way of measuring quality. One recent study identifies three dimensions of quality in banking—approachability, location, and telephone service (Athanassopoulos, 1997). Others have identified retention of valuable customers as the key measure. How would the bank application results change if a different quality measure were used? If there are multiple quality measures, as suggested above and as is well recognized in health services, how would Q-DEA or other benchmarking models adapt? This is another exciting area for future research.

This paper suggests a way of enhancing the DEA methodology for benchmarking and illustrates the potential in a bank setting. While the insights were triggered by the Q-DEA analysis, the motivation to achieve cost savings was strong and other methods might have suggested other ways to achieve cost savings in the branch network. Hence, future research will also examine the enhancement of Q-DEA and other ways to incorporate quality into benchmarking with DEA, and whether other models and methods would be more effective.

Finally although the discussion is based upon the CCR DEA model that assumes constant returns to scale, the quality adjustments considered here can be adapted to other standard DEA models.

Appendix: DEA Model used

DEA linear programming model developed by Charnes, Cooper, and Rhodes (1978):

$$\begin{aligned} &\max E_k = \sum_{r=1}^{s} w_r Y_{rk} \\ \text{subject to} \quad &\sum_{i=1}^{m} \mu_i X_{ik} = 1 \\ &\sum_{r=1}^{s} w_r Y_{rj} - \sum_{i=1}^{m} \mu_i X_{ij} \le 0 \\ &w_r \ge \varepsilon, \mu_r \ge \varepsilon \end{aligned}$$

where, E_k : the measure of productivity or efficiency of branch "k," the branch in the set of j : 1,$\cdots$,225 branches rated relative to the others for the bank analysis; Y_{rk} : The amount of service output "r" produced by branch "k" during the period of observation; X_{ik} : The amount of resource input "*i*" used by branch "*k*" during the period of observation; Y_{rj} : The amount of service output "*r*" produced by branch "*j*" during the period of observation; X_{ij} : The amount of resource input "*i*" used by branch "*j*" during the period of observation; w_r : The coefficient or weight assigned to service output r computed in the solution to the DEA model; μ_i : The coefficient or weight assigned to resource input i computed in the solution to the DEA model; m : The number of resources or inputs used by the branches (four in the bank application); s : The number of services or outputs produced by the branches (ten in the bank application); ε : An infinitesimal positive number that constrains the input and output coefficients to be positive, eliminating the possibility that they will be given a zero relative value in the DEA results.

The objective function of this model maximizes the productivity or efficiency rating E for branch k. This is subject to the constraint that when the same set of w and coefficients are applied to all branches being compared, no branch will be more than 100 percent efficient and the coefficient values will be positive and non zero. For further description and references to other versions of the DEA model, see Cooper, Seiford and Tone (2002) and Zhu (2003).

Acknowledgments The authors are grateful to the two anonymous reviewers for their helpful comments on the earlier versions of the paper.

References

Athanassopoulos, A.D. and D. Giokas. (2000). "The Use of DEA in Banking Institutions: Evidence from the Commercial Bank of Greece." *Interfaces*, 30(2), 81–95.

Athanassopoulos, A.D. (1997). "Service Quality and Operating Efficiency Synergies for Management Control in the Provision of Financial Services: Evidence from Greek Bank Branches." *European Journal of Operations Research*, 98(2), 300–313.

Bessent, A., W. Bessent, J. Elam, and D. Long. (1984). "Educational Productivity Council Employs Management Science Methods to Improve Educational Quality." *Interfaces*, 14(6).

Callen, J.L. (1991). "Data Envelopment Analysis: Partial Survey and Applications for Management Accounting." *Journal of Management Accounting Research*, 3, 35–56.

Camp, R.C. (1993). A Bible for Benchmarking, by Xerox. *Financial Executive*, (July/August), 23–27.

Charnes, A., W.W. Cooper, and E. Rhodes. (1978). "Measuring the Efficiency of Decision Making Units." *European Journal of Operations Research*, 2(6), 95–112.

Charnes, A., W.W. Cooper, Z.M. Huang, and Q.L. Wei. (1989). "Cone Ratio Data Envelopment Analysis and Multi-Objective Programming." *International Journal of Systems Science*, 20, 1099–1118.

Chilingerian, J. and H.D. Sherman. (1990). "Managing Physician Efficient and Effectiveness in Providing Hospital Services." *Health Service Management Research*, 3(1), 3–15.

Cook, W.D. and M. Hababou. (2001). "Sales Performance Measurement in Bank Branches." *OMEGA*, 29, 299–307.

Cooper, W.W., L.M. Seiford, and K. Tone. (2000). *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*. Kluwer Academic Publishers, Boston.

O'Keefe, M. (1994). Different Perspectives—Cutting Through Benchmarking Jargon. *Bank Systems Technology*, November, 53.

Rouse, P., M. Putterill, and D. Ryan. (2002). "Integrated Performance Measurement Design: Insights from an Application in Aircraft Maintenance." *Management Accounting Research*, 13(2), 229–248.

Sherman, H.D. and G. Ladino. (1995). "Managing Bank Productivity using Data Envelopment Analysis (DEA)." *Interfaces*, 25(2), 60–73.

Soteriou, A. and S.A. Zenios. (1999). "Operations, Quality, and Profitability in the Provision of Banking Services." *Management Science*, 45(9), 1221–1238.

Zhu, J. (2003). *Quantitative Models for Performance Evaluation and Benchmarking: Data Envelopment Analysis with Spreadsheets and DEA Excel Solver*. Kluwer Academic Publishers, Boston.