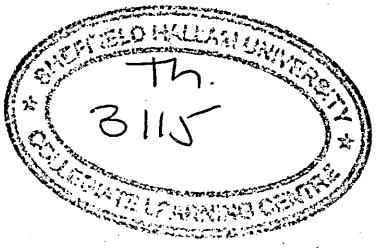
Performance Measurement of English Public Sport Facilities: Aggregate Analysis and its Practicability

Yi-De Liu

A thesis submitted in partial fulfilment of the requirement of
Sheffield Hallam University
for the degree of Doctor of Philosophy



October 2007

Contents

Abstract

The aim of this research is to adopt aggregate performance analyses to measure the performance of English public sport facilities and examine the practicability of these analyses. Based on the National Benchmarking Service for Sports Halls and Swimming Pools (NBS), provided for Sport England by the Sport Industry Research Centre at Sheffield Hallam University, four dimensions of performance are measured in this thesis, i.e. operational efficiency, sport equity, service quality and customer segmentation.

First, two aggregate performance analyses data envelopment analysis (DEA) and multivariate statistical analysis (MSA) are conducted to provide an insight into the industry's overall performance. Second, based on the two approaches above, an aggregate performance analysis framework is developed and tested at the individual facility level. Then its practicability is evaluated. The contributions of this research are twofold: (1) to present the value and critically evaluate the practicability of aggregate performance analysis; and (2) to expand the theoretical literature on performance measurement in the public leisure sector.

The research findings demonstrate that the 'convergence' strength of aggregate analysis can complement the 'inclusiveness' strength of partial measures adopted by the NBS, particularly in efficiency measurement and customer segmentation analysis. In addition, for practitioners in the public leisure sector, 'inclusiveness' and 'simplicity' are the most important criteria of a good performance measurement system. That is, not only does performance data need to be inclusive, the analytical process also needs to be simple and understandable. Finally, facility managers' analytical skills and motivations for benchmarking are two factors which determine

the feasibility of aggregate performance analysis in the public leisure sector.

Keywords: Performance measurement; Sport facility; National Benchmarking Services; Data envelopment analysis; Multivariate statistical analysis

Acknowledgment

My sincere thanks first go to my supervisors Professor Peter Taylor, Professor Simon Shibli and Dr. Jo Padmore. I could not have completed this thesis without their generous support. There is a Chinese proverb: "One who be your teacher one day, is the father for whole lifetime". That is my sincere feeling and the way that I want to express my gratitude. I am also most grateful to Professor Chris Gratton and Prōfessor Tony Bovaird for their helpful recommendation and encouragement during the viva.

I would like to express my gratitude to the three collaborating organisations in this research project. It was with their kind involvement and cooperation was I able to finish the action research. Also, I wish to thank the Taiwanese Government for sponsoring my master and doctoral studies in the UK.

Finally, a special and loving thank you to my parents, my wife and my son. Only with their enormous understanding and unfailing support had I been able to concentrate on my study. Hui-Hsien, you are my rock. Thank you, now and always!

List of Tables

List of Figures

List of Equations

CHAPTER 1

Introduction

1.1 Research Context

The delivery of UK local government services has changed considerably over the past two decades. The 'New Labour' administration in the UK embarked on an ambitious attempt to reform public-sector organisations, in order to improve the quality and cost effectiveness of the services which they provide. The centrepiece of these reforms is the introduction of the Best Value regime, which came into effect in England and Wales in April 2000. Under the provisions of the 1999 Local Government Act the requirement to submit defined activities to compulsory competitive tendering (CCT) was abolished in January 2000. The Best Value regime replaced CCT, and conferred on public-sector organisations a legal duty to provide Best Value services, service quality and value for money. In order to achieve Best Value in public-sector organisations, performance measurement and benchmarking are two of the proposed tools of government reform to assist in the monitoring and control of productivity and quality with a focus on internal and external stakeholders in mind (Bowerman and Ball, 2000; Bowerman *et al.*, 2001; Martin and Hartley, 2000; Ogden and Wilson, 2000).

Performance measurement against a strategic plan is consistent with requirements of the UK Audit Commission performance measurement framework that links performance with corporate vision and strategic objectives. It has also been

suggested that performance measurement will assist strategic planning by allowing better deployment of resources and management attention. Best Value involves the development of a corporate plan for the organisation, a review of the service currently being provided and, using consultation and performance measurement, the development and delivery of an action plan (performance improvement plan) for enhancing future service provision. At the end of the performance plan period the mode of service delivery will be reviewed based on performance indicators (McAdam and O'Neill; Kloot, 1999). More recently, Comprehensive Performance Assessment (CPA) that is the performance management framework used by the Audit Commission and Central Government to measure local authority performance and to drive improvement in the sector, further highlights the assessment of cultural services including sport (Audit Commission, 2005).

In England, sport contributes to a range of outcomes and priorities shared by local and central government and other public sector agencies. In particular, sport can promote: healthier communities, safer and stronger communities, raising standards, economic vitality and meeting the needs of children and young people. Public sports facilities including sports halls and swimming pools are an important aspect of local government provision and make a major contribution to these outcomes (Sport England, 2006). Under the regime of Best Value, local authorities are required to demonstrate that they are running public sport facilities in a manner ensuring that the community receives an effective and high quality service (Robinson and Taylor, 2003).

The National Benchmarking Service for Sports Halls and Swimming Pools (NBS), provided for Sport England by the Sport Industry Research Centre at Sheffield

Hallam University, is one of the most fully developed sport and leisure services performance measurement systems. In the CPA framework for 2005-2008, two sets of NBS performance indicators- 'value for money' and 'equity' were proposed to measure the performance of public sport facilities (Audit Commission, 2005). Although the DCMS has finally decided not to include the seven NBS indicators in the CPA, they are still a good indication of government priorities for performance of sports facilities.

It has been traditional in the UK public services to take a piecemeal approach to measure performance, relying on a set of performance indicators that capture particular aspects of performance. The NBS is an example that provides a wide range of performance indicators and benchmarks covering multiple performance dimensions. However, regulators are increasingly attracted to the development of global measures of organisational performance (Smith, 1990). For instance, one principle of CPA is to aggregate a range of existing judgements to provide a new means of ranking performance. These are aggregated by locating the different combinations of scores which summarise core service performance and provide an assessment of overall performance (Andrews, 2004; Broadbent, 2003).

To date, several techniques have been developed to aggregate various performance measures and reduce the complexity of a performance measurement model. According to previous research, techniques with the potential to enhance the NBS are Data Envelopment Analysis (DEA) and Multivariate Statistical Analysis (MSA). The characteristics of these two techniques will be discussed later. To sum up, the major functions of DEA and MSA lie in aggregating a series of performance indicators to provide a holistic overview of the performance of sport facilities.

Although there has been widespread interest in the application of DEA or MSA in the academic field, little attention has been paid to the question of the appropriateness and practicability of these aggregate performance analyses. The question of how aggregate performance analysis can contribute to the performance measurement of public sport facilities becomes the motivation of this research.

1.2 Research Objectives and Questions

The aim of this research is adopting aggregate performance analyses to measure the performance of English public sport facilities and examine the practicability of these analyses. According to the overall aim, two specific research objectives leading to two research questions are developed:

- ***Objective 1*** - Two aggregate performance analyses, i.e. DEA and MSA are conducted to provide an insight into the industry's overall performance.
- ***Objective 2*** - Based on the above two approaches, an aggregate performance analysis framework is developed and tested at the individual facility level, and then its practicability is evaluated.
- ***Question 1: Value*** - Whether aggregate performance analysis provides valuable information which is not readily available from NBS type evidence?
- ***Question 2: Practicability*** - Whether aggregate performance analysis can be transferred successfully from academic to practical domain?

The potential contributions of this research are twofold:

- Present the value and critically evaluate the practicability of aggregate performance analysis.

• Expand the theoretical literature on performance measurement in the public sector.

1.3 Research Framework

A conceptual map illustrating the research framework is presented in Figure 1.1. Two analytical approaches, i.e. DEA and MSA, are used to measure four dimensions of performance, i.e. operational efficiency, sport equity, service quality and customer segmentation. The selection of these four dimensions is principally in line with the current NBS framework. The reason why DEA and MSA were selected will be justified in Chapter 3 and Chapter 5.

Based on a secondary data analysis of the 2001 NBS database, the analysis at the industrial level aims to evaluate the industry's overall performance. On the other hand, taking action research as the research strategy, the practicability of aggregate performance analysis is examined in three public sport/leisure centres in England during 2005-2007.

The thesis begins with a literature review of the academic literature, covering the major domains of performance measurement theories and the measurement of specific performance dimensions to be studied in this research (Chapter 2, 3 and 4). The methodology chapter (Chapter 5) then explains the research philosophy, strategies and instruments. The analytical results at the industrial level are presented in Chapter 6 and 7, followed by the results at the individual facility level in Chapter 8. Finally conclusions are drawn in Chapter 9 to discuss the theoretical implications and contributions of the thesis. To facilitate reading, a 'road map' is also provided as

shown in Figure 1.2. The elements which each chapter deals with will be highlighted and this figure will be repeated in the beginning of every chapter.

Figure 1.1 Conceptual map of research framework

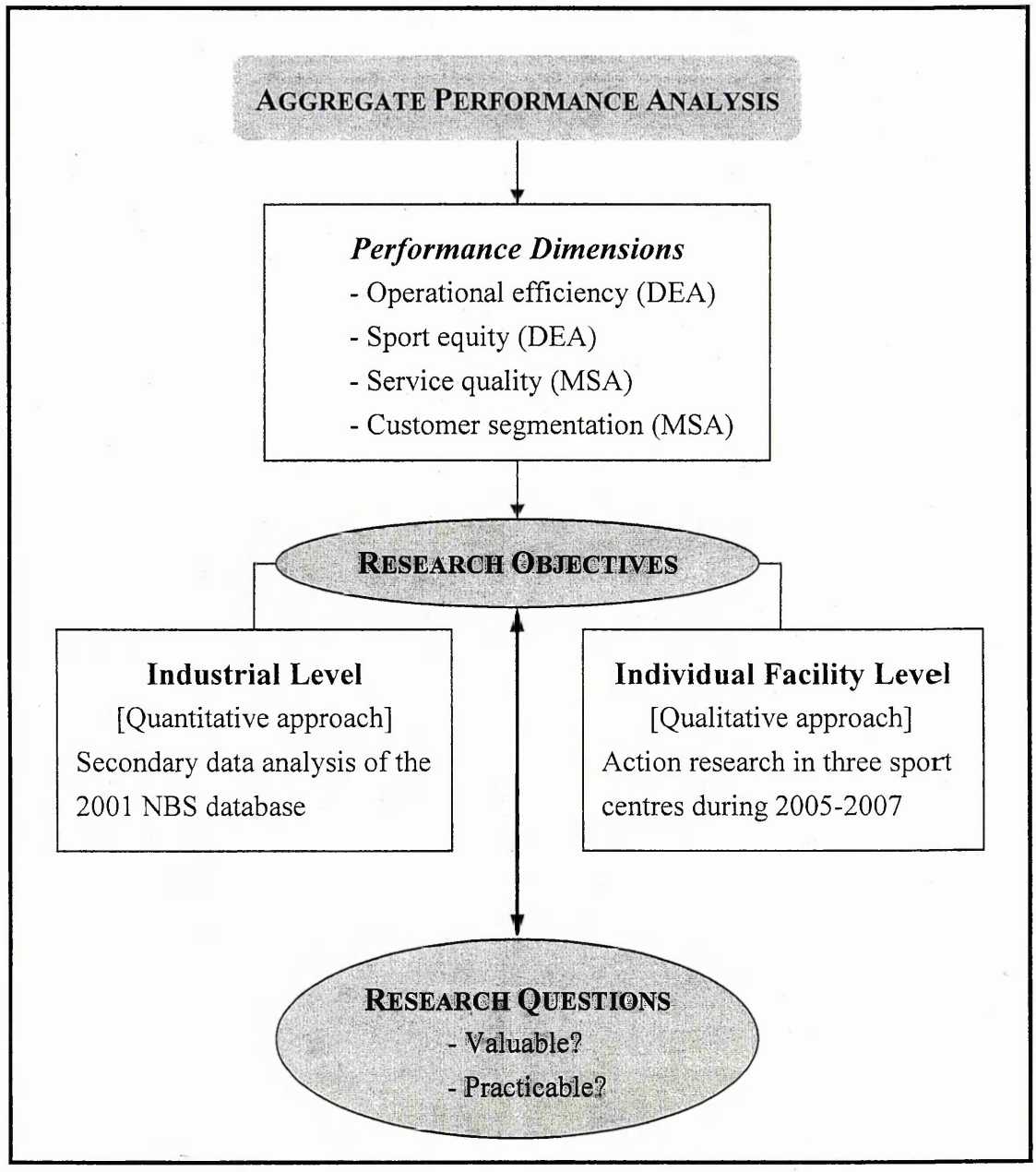


Figure 1.2 Road map of the thesis structure

Introduction	Theoretical context	Methodology	Results	Conclusion
Ch1 ▶	Ch2 Performance measurement overall Ch3 Efficiency & Equity Ch4 Customer service	▶ Ch5 ▶	Ch6 Efficiency & Equity Ch7 Customer service Ch8 Action research	▶ Ch9

CHAPTER 2

Theoretical Context of Performance Measurement

Road map of the thesis

Introduction	Theoretical context	Methodology	Results	Conclusion
Ch1 ▶	Ch2 Performance measurement overall Ch3 Efficiency & Equity Ch4 Customer service	▶ Ch5 ▶	Ch6 Efficiency & Equity Ch7 Customer service Ch8 Action research	▶ Ch9

In this chapter, a literature review of performance measurement theories assists the author in identifying the research gap in the field being studied. Performance measurement may refer to an organisation, but also enable comparison, i.e. 'benchmarking' between organisations (Bruijn, 2002). As such, in the first two sections, the focus is on the theoretical context of performance measurement and benchmarking. Then, the principles of performance measurement model design are discussed and the research rationale developed.

2.1 Performance Measurement

2.1.1 Definition of performance measurement

In performance measurement, it is often valuable to distinguish different types of 'performance information'. According to the classification of the Chartered Institute of Public Finance and Accountancy (CIPFA), it includes: "performance measures which look at the relationship between activities and the results associated with them; performance targets which state what an authority is required or aims to achieve in a future period; and performance standards which set out the minimum level of performance" (LIRC, 2002).

The most common method of measuring performance is through the use of 'performance indicators' (PIs) (Sport England, 2000a). Typically, PIs are expressed as ratios. Ratios are used rather than absolute numbers because they are standardised by the denominators and are therefore more appropriate for comparisons across service units. PIs are signposts highlighting areas that are performing well and areas that may require further exploration. Once the current performance level is known, it is important that realistic but challenging targets are set for PIs selected so as to improve services.

Furthermore, PIs' scores can be compared over time for the same organisation(s) or compared with other similar organisations (Taylor *et al.*, 2000). In his discussion of performance indicators, Craig (1995) noted that they have three key functions. First, they provide data by which an organisation can be judged. Second, performance indicators allow standards and targets to be generated, which tell people what to expect. Third, they could be used as a management tool to alert managers to potential

weaknesses and corrective actions. To conceptualise some common terms related to performance measurement, Table 2.1 provides a summary description of and a linkage between 'performance indicator', 'performance measurement' and 'performance measurement model'.

Table 2.1 Terms related to performance measurement

Terms	Definitions
Performance indicator	A variable that expresses quantitatively the effectiveness or efficiency or both, of a part of or a whole process, or system, against a given norm or target
Performance measurement	The activity of measuring performance using PIs.
Performance measurement model	A system to execute performance measurement in a consistent and complete way.

Source: Lohman *et al.* (2004)

According to Wisniewski and Stewart (2001), performance measurement can be broadly interpreted as the measure of '3Es' (economy, efficiency and effectiveness), including issues of quality of service and local accountability. A number of authors have also suggested that government performance needs to measure these 3Es (e.g. Palmer, 1993).

• ***Economy*** is defined as acquiring resources in appropriate quantity and at least cost.

• ***Efficiency***, which describes how well an organisation uses resources in producing services, that is, the relationship between the actual and optimal combination of inputs used to produce a given bundle of outputs (Worthington and

Dollery, 2000) Efficiency can be achieved either through maximising outputs for a given set of inputs or minimising inputs for a required output.

Together, economy and efficiency are consistent with notions of financial accountability to both central government and the local community. Economy and efficiency are usually measured in financial terms, and data such as costs, volume of service and productivity are relatively simple to measure (Palmer, 1993).

• ***Effectiveness*** is defined as the extent to which the defined task has been accomplished (Palmer, 1993; Jackson and Palmer, 1988) and is consistent with notions of non-financial accountability to the local community. Determining effectiveness is problematic, given the unmeasurable nature of the output of many local government services, such as human services. The problems include the inability to accurately measure output, difficulty in isolating the effects of the service (the outcomes) from other factors, lack of quantifiability of the effects of services and conflicting interpretations of results (Hasenfeld, 1983).

According to Worthington and Dollery (2000), effectiveness encompasses a number of different desired aspects of service linked to programme outcomes. These are:

(i) Accessibility (aspects such as affordability, representation amongst priority groups and physical accessibility);

(ii) Quality (the process of meeting required standards); and

(iii) Appropriateness (matching service to client needs).

From the typology proposed by Worthington and Dollery (2000), equity is subsumed under effectiveness as accessibility. However, as suggested by Crompton (1983), the distinctive public sector evaluation measure is equity which differentiates

mainly public and private service delivery. In this research, three dimensions of performance are measured: operational efficiency, sport equity and customer service. The latter two can also be defined as effectiveness if the typology of Worthington and Dollery (2000) is considered.

2.1.2 Partial and Global Measures

Traditionally, two broad approaches have been used to present the performance data, i.e. partial measures and global measures (Smith and Street, 2005). According to Chen (2003), a good performance measure approach should both have high 'inclusiveness' and high 'convergence'. Inclusiveness means that all aspects of the organisation should be considered and convergence means that consistent and simple information should be provided to facilitate decision making. In this section, the advantages and disadvantages of partial measures and global measures will be discussed on the basis of the two criteria- inclusiveness and convergence.

• ***Partial measures***

The word partial is used here since these measures do not capture performance based on all relevant outputs and inputs but deal only with one input and (or) output at a time. The partial measure is easy to interpret intuitively, but its main problem is a lack of inclusiveness (Chen, 2003). In order to consider all aspects of the organisation's performance, namely enhancing the inclusiveness, one way is to calculate a set of partial measures, or called 'family measure', which cover all aspects related to the organisation. Though listing a set of measures can cover all pertinent aspects, different partial measures may lead to different diagnoses and sometimes they

are even in the opposite directions (Bhargava *et al.* 1994; Lyons, 1995). The lack of convergence makes it hard to make decisions.

• ***Global measures***

In contrast to partial measures, global measures are designed to provide an indication of overall organisational performance by aggregating different aspects of performance. Whilst the calculation of a set of these partial measures is a relatively easy task, the aggregation of those measures can be quite complicated (Yeh, 1996). Even though economists have developed many different aggregating schemes: price-weighted sum, geometric weighted sum or other functional forms based on different assumptions (Grosskopf, 1993), there are still two barriers difficult to be conquered. First, if the price information is available, the aggregating process could convert all the original units into monetary units. However, the prices might change all the time, and more importantly, some inputs and outputs have no natural price, or because they are intangible and it is not easy to determine their values (Chen, 2003). Second, the aggregation process requires either *a priori* weights or the explicit functional relationship between inputs and outputs. It is usually difficult for a decision maker or analysts to specify the weights of different variables and define a suitable functional form (Agrell and West, 2001; Grosskopf, 1993).

2.1.3 Characteristics of Partial Measure

Partial measures (also known as 'ratio analysis') have become, over the years, a well-established technique that has found numerous applications in many areas of business (Athanassopoulos and Ballantine, 1995). One of the main advantages of ratio

analysis, according to Lawder (1989), is the ability to measure the relationship between two numbers in the financial statements. Not only can the nature of the relationship be expressed in absolute terms, but it is also possible to quantify the change in the relationship over time. Smith and Street (2005) and Smith (1990) also argue that, as a performance measurement tool, ratio analysis has much to recommend it. It focuses on specific aspects of performance, which are readily measured and validated, and easy to interpret. They might therefore be very useful from a local managerial perspective. In practice, ratio analysis continues to be the method of choice, at least for the UK government, in reporting performance in the delivery of publicly funded services (Thanassoulis *et al.*, 1996). Despite the widespread use of ratio analysis for assessing performance, the univariate nature of this approach leads to some limitations:

• ***Limitation of univariate analysis***

Ratio analysis typically involves the use of a number of performance indicators, i.e. a set of partial measures. One of the most fundamental limitations of this kind of univariate analysis is that only two dimensions of activity, represented by numerator and denominator, can be examined in any one indicator. In single-input, single-output contexts such a measure is a meaningful, easy to use, measure of performance. However, this is not the case where multiple non-commensurate inputs and/or outputs are involved (Thanassoulis *et al.*, 1996; Worthington, 1999). It is obvious that firms are multi-dimensional entities: a single measure is unlikely to reflect the complexity of decision-making or the scope of a firm's entire activities. Thus, a meaningful

evaluation tool has to take into consideration the multidimensional character of the performance construct (Athanassopoulos and Ballantine 1995).

• ***Difficult to evaluate overall performance***

Another problem with using ratio analysis is that there may be conflicting signals emerging from competing ratios while considering many ratios, as is usually the case. An organisation that appears to do well on one indicator may perform less successfully when considered on another. It is therefore not straightforward to draw conclusions about overall organisational performance from a range of performance indicators. This difficulty stems from the multitude of performance indicators and the fact that each performance indicator reflects only one input and one output level (Al-Shammari and Salimi, 1998; Smith and Street, 2005; Thanassoulis *et al.*, 1996). Thanassoulis *et al.* (1996) also argue that electing to use only some of the potential performance indicators will bias the assessment and inclusion of large numbers of variables and lack of an indicator to evaluate unit performance overall often frustrate management efforts to implement strategy. It is therefore important to consider the overall efficiency of decision making units as well as compare them on the basis of performance indicators which capture only one dimension of performance (Young, 1992).

• ***Problems of aggregating and weighting***

In order to evaluate the overall performance, aggregating a set of ratios and weighting these ratios in some fashion is usually necessary (Yeh, 1996). According to Metzger (1993), the weights for the inputs and outputs should be derived

endogenously, i.e. in an objective way and should not be influenced by the preferences and personal predispositions of policymakers. In contrast, the need for *a priori* specification of weights does presumably not lead to acceptable and credible benchmarking results, because of the dependence on negotiation processes. Assigning deterministic weights to each input and output variable upon which all decision makers finally agree could be an endless and resource consuming process.

2.1.4 Characteristics of Global Measure

Although it has been traditional in the UK public services to take a piecemeal approach to measuring the performance, relying on a set of partial measures that capture particular aspects of performance, regulators are increasingly attracted to the development of global measures of organisational performance (Smith, 1990). Examples in England include the system of CPA for local authorities and the performance ratings that are applied to National Health Service organisations (Audit Commission, 2003; Commission for Health Improvement, 2003). Moreover, the Public Services Productivity Panel has advocated the joint use of two of the most advanced relative efficiency measuring techniques— Data Envelopment Analysis (DEA) and Stochastic Frontier Analysis (SFA) to measure the efficiency of police forces in the UK (Spottiswoode, 2000), and suggested that the techniques would provide a systematic, comprehensive measure of relative police efficiency and allow differentiated efficiency (performance) targets to be set for the police.

According to Smith and Street (2005), the arguments pro and con to develop global measures are summarised in Table 2.2. The global measure model adopted by this research, i.e. DEA will be discussed in detail in Chapter 3.

Table 2.2 The arguments pro and con to develop global measures

Pro	Con
- offer a rounded assessment of system performance - offer local managers the freedom to set their own priorities - facilitate the publication of 'league tables' or rankings of entire organisations - support other objectives, e.g. allocating finance and identifying the priority of organisations	- Is it legitimate for policy makers to seek to develop global measures of organisational performance? - Is it legitimate for the central policy maker to attach a uniform set of objectives to all organisations and to apply a uniform set of weights to these objectives? - Why should policy makers have any interest in global measures of organisational efficiency?

Source: Smith and Street (2005)

2.2 Benchmarking

2.2.1 Definition of Benchmarking

Benchmarking can be seen as an important management tool of total quality management (TQM). It was first developed by Xerox Corporation in 1979. Given the prevailing emphasis on agency performance, customer focus and prevailing managerialism in many public sectors, administrative practitioners have taken to benchmarking as an instrument for assessing organisational performance and for facilitating management transfer and learning from other benchmarked organisations (Magd and Curry, 2002; Kouzmin *et al.*, 1999).

There have been a number of definitions of benchmarking. For example, Foot (1998), in her guide for the Inter-Authorities Group, describes it as: a process of measuring your service's processes and performance and systematically comparing them to the performance of others in order to seek best practice. Similarly, the Audit Commission (1995) defines benchmarking as: the process of comparing procedures and performance levels between and within organisations in order to identify where improvement is possible, how it might be achieved and how much benefit it might deliver.

Another relevant term is 'benchmark', which could be regarded as a reference point for comparisons. According to Taylor *et al.* (2000) and Robinson and Taylor (2003), it could be a performance indicator value for a particular facility at a specific time, against which future performance is compared, or a value for a group of similar organisations, so that an individual organisation can compare itself with the norm for such organisations.

2.2.2 Benchmarking in the Public Sector

Benchmarking is identified as one of an ever-growing number of management practices aimed at improving performance in public-sector organisations. In order to implement the Best Value framework, organisations must establish service performance measures and compare their achievements against those of the service leader in the industry. The concept of benchmarking has therefore been considered to be vital in achieving Best Value in the public sector (Magd and Curry, 2002; McAdam and O'Neill, 2002). However, the introduction of benchmarking into the public sector is still in its early stages. Benchmarking is also a relatively new practice

in the UK public-sector organisations. Technical problems, scepticism about usefulness and the appropriateness of transferring putative private sector competencies into public administration and the resistance in accepting organisational change as a necessary consequence of benchmarking exercises in the public sector, prevent the widespread acceptance and use of benchmarking in public sectors (Kouzmin *et al.*, 1999).

Benchmarking in the public sector is analogous to that in the private sector but the motivational forces and obstacles are somewhat different (Kouzmin *et al.*, 1999). The existing benchmarking research is also generally taken as providing a generic understanding how benchmarking might operate in the private sector, but little is related to evidence from the public sector. According to Bowerman *et al.* (2002), there are three characteristics that distinguish private sector benchmarking from that which takes place in the public sector.

- In the private sector, benchmarking is often undertaken in order to be the best. In contrast public sector organisations may strive, through benchmarking, to be 'good enough' or merely to demonstrate that they are not the worst.
- Information generated through benchmarking in the private sector is confidential. By contrast, public agencies have no competitive drawbacks to fear from passing information on to peer organisations.
- Private sector benchmarking is voluntary in nature. However, in the public sector, benchmarking is frequently conducted in response to central government requirements.

According to the above characteristics, a new public sector benchmarking typology was introduced by Bowerman *et al.* (2002) (as shown in Table 2.3). They

argued that voluntary benchmarking is similar to private sector benchmarking; that is with an emphasis on continual improvement and striving to be the best. In contrast compulsory benchmarking and defensive benchmarking are primarily to prove to an external agency that they are doing well or are not the worst.

Table 2.3 Typology of public sector benchmarking

Typology	Definition
Compulsory benchmarking	- Driven by external accountability requirements.
Defensive benchmarking	- To demonstrate that they are at or near 'good' levels of performance.
	- Benchmarking results become more important than acting on those results in order to close performance gaps.
Voluntary benchmarking	- Use of benchmarking for performance improvement and innovation.
	- Closest to received private-sector benchmarking wisdom.
	- Processes are examined in detail to understand performance differences.

Source: Bowerman *et al.* (2002).

2.2.3 Ranges of Benchmarking Activities

There is a wide range of benchmarking activities. For instance Foot (1998) and Ogden and Wilson (2000) suggest the following four types of benchmarking which public bodies can choose from, depending on what they are trying to tackle.

• ***Data benchmarking***: The numerical comparison of performance in key areas and identify performance gaps. Typically, performance indicators are used for measuring performance and monitoring progress against set targets.

• ***Process benchmarking***: The comparison and measurement of a specific process against a similar process in own or another organisation. It highlights the causes of differences in performance and generates ideas as to how to improve the processes.

• ***Functional benchmarking:*** The comparison of structure and performance of an entire function in the organisation with a comparable function elsewhere.

• ***Strategic benchmarking***: The comparison of strategic approaches or initiatives across organisations which can only be applied where strategic goals are comparable.

Some studies (e.g. Ball *et al.*, 2000; Ammons, 1999) have demonstrated that most UK public organisations tend to focus on league tables as comparators of performance. This tends to be caused by a lack of understanding of the purpose of benchmarking and the government focusing on or promoting the use of a specific type of benchmarking (Bowerman and Ball, 2000). According to Ogden and Wilson's research (2000), the evidence from the NHS suggests that only data benchmarking was carried out by exchanging data from 29 different benchmark organisations but they are planning to conduct some process benchmarking. Kouzmin *et al.* (1999) also pointed out that it is difficult for local authorities to have the degrees of freedom to implement functional benchmarking because information obtained from non-related public sector organisations cannot be used. Cox and Thompson (1998) found that strategic benchmarking is the least frequently used in public organisations because it is hard to do and the benefits of it are typically realised only in the long-term.

2.3 Performance Measurement Model

This research utilises data from the National Benchmarking Service (NBS) to explore performance measurement in public sector sport. The author attempts to develop a model which has the potential to enhance the NBS. Here, the characteristics of NBS are first introduced followed by discussing the principles about performance measurement model design.

2.3.1 National Benchmarking Service

The NBS could be regarded as a performance measurement model (or system). As a partial measure approach, the NBS provides a wide range of performance indicators and benchmarks covering multiple performance dimensions, which allow individual facilities and authorities with the choices appropriate to their different needs, circumstances and priorities (Taylor and Godfrey, 2003; Robinson and Taylor, 2003). The NBS is also an example of data benchmarking, identifying what performance is like relative to national benchmarks. The performance indicators / benchmarks provided by NBS fall into the following four groups (LIRC, 2002) and an overview of the NBS model can be found in Appendix 1.

- ***Access:*** representing the extent to which facilities are used by disadvantaged groups and new users. These are performance indicators of effectiveness, particularly in the context of social inclusion.
- ***Financial:*** representing subsidy, cost and income performance. These are indicators of efficiency and economy.

• ***Utilisation:*** representing the scale and nature of usage and non-usage of facilities. These are indicators of effectiveness and efficiency.

• ***Satisfaction and importance:*** representing the context to which users are satisfied with attributes of the facility and how important these attributes are to them. These are indicators of effectiveness. Six service dimensions are considered in the NBS: accessibility, availability, quality of facility, cleanliness, staff and value for money.

Furthermore, in order to prevent comparison against dissimilar organisations or against organisations with dissimilar customer profiles, the results of benchmarking are structured by four 'families' of centres representing major influences on performance: the type of centre, the socioeconomics of the centre's location, the size of the centre and the type of management. They demonstrate the validity and value of an extensive and 'like for like' measurement of performance (Robinson and Taylor, 2003; LIRC, 2002).

2.3.2 Performance Measurement Model Design

In recent years performance measurement has received considerable attention from academics, practitioners and policy makers. Some authors have focused attention on how organisations can design more appropriate measurement models (e.g. Bourne *et al.*, 2000; 2002; Neely *et al.*, 1997; 2000 etc.) Also, many models, such as the Balanced Scorecard (Kaplan and Norton, 1992) and the Performance Prism (Kennerley and Neely, 2000) have been developed. In the field of sport facility management, some frameworks such as NBS, CERM (developed by Centre for Environmental and Recreation Management at the University of South Australia) and

APSE Performance Networks etc. have been developed to help facility managers to assess performance. Finally, action research, consultancy experience and numerous processes have been developed that organisations can follow in order to design and implement performance measurement models.

However, it is necessary to recognise the changing nature of performance measurement. According to Neely and Bourne (2000), managers have become obsessed with measuring performance today, so they no longer have time to act on the performance data once they have been gathered. In the 1980s and early 1990s, the fundamental problem was measuring the wrong things. Now the problem is to measure too much. They suggested two criteria should be considered while developing a performance measurement model:

- The first criterion is associated with simplicity and automation. The trick is to measure as little as possible, but to ensure to measure the things that matter.
- It is also important to extract value from the performance measurement data because usually people are not aware of the tools and techniques that are available to help them understand the messages inside the performance data.

2.4 Development of Research Rationale

From the literature review, some gaps in the theory emerged, providing some potential research areas for this thesis.

First, the approach adopted by NBS is considered as partial measures. As mentioned in 2.1.3, a set of partial measures is considered good in inclusiveness but relatively weak in convergence (Chen, 2003). As argued by Young (1992), it is

important to consider the overall performance of organisations as well as compare them on the basis of performance indicators which capture only one dimension of performance. As such, partial and global measure should ideally be integrated in measuring the performance so that inclusiveness and convergence could be considered simultaneously. However, according to Smith and Street (2005), the weighting procedure for yielding a single global measure is not straightforward. If it is not legitimate to apply a uniform set of weights, a more flexible weighting system should be sought in order to reflect different circumstances or priorities of organisations.

Second, the NBS is an example of data benchmarking, identifying what performance is like relative to national benchmarks. However, the results of the service are the start of the process, not the end. Developing benchmarks is only the first stage in the benchmarking process. Data benchmarking without process benchmarking leads to a much more restricted appreciation of comparisons, because the reasons for relative performance are more difficult to deduce (Taylor and Godfrey, 2003; Ogden and Booth, 2001). Magd and Curry (2002) and Hinton *et al.* (2000) also suggested that the critical characteristic of benchmarking is the examination of processes, as it is only through a proper understanding of how inputs are transformed into outputs that will enable public-sector organisations to achieve superior results. The success of benchmarking is based on the desire to change processes as well as outputs and on organisational willingness to search for ideas outside the organisation. Consequently, data benchmarking and process benchmarking should ideally combine together in a synergistic relationship. First, data benchmarking indicates where performance could be improved. Second, process benchmarking may then be used to

improve performance by learning from innovation and new thinking elsewhere in the sector (Favret, 2000).

Furthermore, most of the existent performance measurement models aim to help organisations define a set of measures that reflects their objectives and assesses their performance appropriately. However, little attention and still less empirical research has been dedicated to help managers to extract the most useful information from the performance data.

The issues pointed out above prompted the author to:

- find a way to ensure the inclusiveness and convergence of performance measurement;
- facilitate a shift away from data benchmarking to process benchmarking; and
- make use of the existent NBS database to extract valuable performance information.

Finally, this research consists of the measurement of four dimensions of performance: operational efficiency, sport equity, customer service quality and customer segmentation. The theoretical contexts of these four dimensions are discussed in the following two chapters.

CHAPTER 3

Theoretical Context of Efficiency and Equity Measurement

Road map of the thesis

Introduction	Theoretical context	Methodology	Results	Conclusion
Ch1 ▶	Ch2 Performance measurement overall Ch3 Efficiency & Equity Ch4 Customer service	▶ Ch5 ▶	Ch6 Efficiency & Equity Ch7 Customer service Ch8 Action research	▶ Ch9

This chapter aims to review the literature related to operational efficiency and sport equity. Before discussing the measurement of these two dimensions, the industrial characteristics are first reviewed.

3.1 Industrial Context

3.1.1 Operational Efficiency

In England, over 60% of local authority leisure department net expenditure goes on indoor sports facilities (DCMS, 2002). UK Government legislation in the form of Best Value requires local authorities to demonstrate that they are running such facilities in a manner ensuring that the community receives an effective, high quality

service (Robinson and Taylor, 2003). As stated in the Local Government Act 1999, "a best value authority must make arrangements to secure continuous improvement in the way in which its functions are exercised, having regard to a combination of economy, efficiency and effectiveness" (Ogden and Booth, 2001). More recently, in the CPA framework for 2005-2008, two performance indicators- 'subsidy per visit' and 'visits per square metre' were proposed to measure the operational efficiency of sport facilities to ensure public sport services deliver value for money (Audit Commission, 2005).

According to Mintel (2006), the supply side of public sport facilities has enjoyed an unprecedented period of growth during the past decade, prompted by considerable investment in new provision funded largely by the National Lottery and, more recently, public-private partnerships. In the UK, there were 3,640 public sport facilities in 2001 and 4,215 in 2006, a growth rate of 16% over these five years. The provision of sport facilities relative to the growth in the population (2% in the same period 2001-2006) is an important factor in creating demand. Also, economic and demographic trends have driven the demand growth during the past five years. These include strong rises in real levels of personal disposable income and consumer spending, growth in the numbers of 15-24-year-olds, increased numbers of social classes AB (the two highest socio-economic groups in the UK classification) - consumers who are the most likely to be heavy users of such facilities - and the rising concern about the health of the nation, particularly relating to the dramatic increases in the numbers of overweight and obese people. However, the industry also faces dual threats from the looming problem of an ageing stock of buildings, many of which were constructed in the 1970s and need refurbishing or replacing, and the soaring cost

of energy, which is a substantial variable cost that has almost doubled in price during the past three years alone.

3.1.2 Sport Equity

In the UK, sport plays a potentially important role in promoting the inclusion of all groups in society, but inequalities have traditionally existed within sport across social class, age, ethnicity and disability. In the face of this evidence, an interventionist sports policy might hope to improve sporting opportunities for the socially excluded. In principle this has been a strong tradition in local authorities' sports provision, with 'Sport for all' policies tending to be interpreted as increasing opportunities for socially and recreationally disadvantaged groups (Gratton and Taylor, 2000). Recently, social inclusion became a major policy objective of the British Government, after 1997. After the publication of *Bringing Britain Together* and the 18 Policy Action Team (PAT) reports, including one on sport and the arts, the government required all the agencies and Lottery funds it oversaw to have a policy for combating social exclusion. The PAT for Sport and the Arts (PAT 10), reporting on the role of sport and the arts in contributing to neighbourhood regeneration and combating social exclusion, emphasises the importance of taking positive action to address inequalities and of measuring the outcomes (DCMS, 1999). Also, at the national level the strategy document of Sport England - *England, the Sporting Nation: A Strategy* has set targets to increase participation in sport and to reduce its inequities (Sport England, 1999b).

At the local level, swimming pools and sports halls are provided by local government at typically subsidised prices. This is either because of a social equity

concern to provide equal opportunities for all citizens irrespective of financial resources or because of presumed 'externalities' (associated social benefits) such as improved health or reduction in vandalism. In the case of social equity the assumption is that for many social groups the entrance cost is a major obstacle to participation and that the market would not provide a socially just distribution of opportunities (Coalter, 1998; 1993; Gratton and Taylor, 1988). In 1997, Sport England commissioned user surveys in 155 local authority sports halls and swimming pools, with a total 41,000 respondents. The results provide the most authoritative and conclusive evidence to date of some enduring inequities in participation in such facilities (Gratton and Taylor, 2000). For equality of opportunity to exist for all, it is important that barriers are removed and opportunities maximised. One way is to measure how representative facility usage is.

Recently, five NBS equity indicators were proposed for the CPA framework for 2005-2008: representativeness ratios for people from the most disadvantaged socio-economic groups (i.e. social classes DE); young people aged 11-19 years; people from black, Asian and other minority ethnic groups; people aged over 60 years; and the simpler share ratio of the percentage of facility use by disabled people aged under 60 years (Audit Commission, 2005). Although the DCMS has finally decided not to include these indicators in the CPA, they were included in the consultation process so they are a good indication of government priorities for performance of sports facilities.

3.2 Operational Efficiency

3.2.1 Parametric and Non-Parametric Methods

Operational efficiency, sometimes referred to as 'cost-effectiveness' or 'value for money', is defined, in this research, as the extent to which public sport facilities can maximise the outputs (operating income and visits) by a given set of inputs (i.e. operating cost, facility area and opening hours). From the literature, operational efficiency can be presented by partial or global measures and be measured via parametric or non-parametric approaches. Productivity research has developed two broad schools of analytic thought that is intended to inform the development of global efficiency measures (Stone, 2002). Relying on econometric techniques, the parametric approach (e.g. SFA) requires explicit specification of a functional form. By contrast, based on mathematical programming, the non-parametric approach (e.g. DEA) does not require any prior assumption about the relationship between inputs and outputs but allows data to determine the shape of the efficiency frontier. DEA is therefore less computationally intensive, easier to be applied and more widely used in efficiency measurement (Coelli *et al.*, 2005).

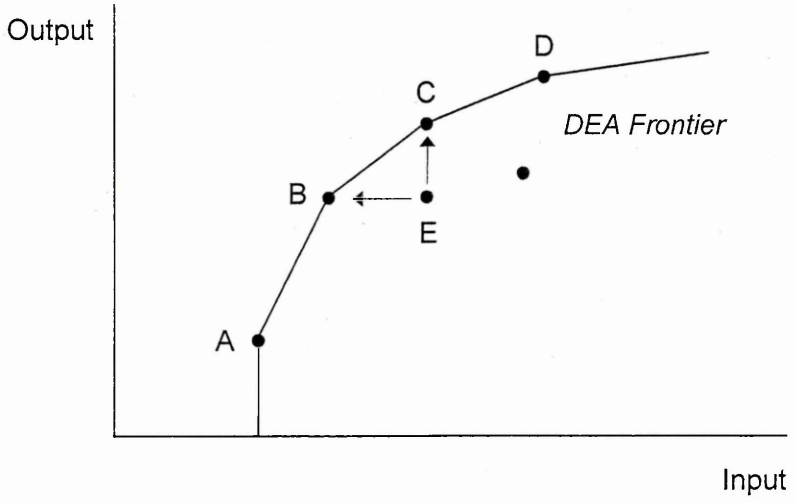
In addition, according to Metzger (1993), for the benchmarking results to be easy to communicate and close to practice, the benchmarks have to be real existing organisations not hypothetical or prescriptive ideals. Unlike parametric approaches which infer the efficient frontier from the average behaviour of all observed organisations, the evaluation provided via the DEA model conforms to the actual performance of the best performing organisations (Smith and Street, 2005). The above two arguments in favour of the non-parametric approach are the main reasons for

choosing DEA as the research model in this study. The advantages and disadvantages of DEA will be discussed further in the following section.

3.2.2 Principles of DEA

DEA was developed by Charnes *et al.* (1978) based on an idea of efficiency measurement first suggested by Farrell (1957). The definition of efficiency in DEA is derived from the engineering concept of total factor productivity and is specified as the ratio of the weighted sum of outputs to the weighted sum of inputs of a Decision Making Unit (DMU), the DMU in this study being a sport centre. The efficiency frontier constructed by DEA comprises linear segments that interpolate between those observations with the highest ratios of output to input. The resulting frontier thus 'envelops' all the observations. As shown in Figure 3.1, observations A, B, C and D are considered efficient, given the scale of their operations.

Figure 3.1 Illustration of DEA frontier



The inefficiency of observation E is indicated by either its vertical (EC) or its horizontal (EB) distance from the frontier- it uses more input to produce a similar level of output to observation B and, despite employing a similar amount of input to observation C, it produces less output. The enveloping surface is interpreted as the efficient technology frontier, so that the distance from the inefficient units to the frontier provides a measure of inefficiency. Also, the efficiency measure obtained through DEA is relative. Each unit is compared with those operating with a similar input and/or output value. This is done in order to determine its location on the frontier (provided that it is efficient), or to identify its reference set for a subsequent improvement in the results (in the case where it is considered inefficient) (Smith and Street, 2005; Doble, 1995). Finally, differing from regression analysis, the principle of linear programming underpinning DEA allows data to determine the shape of the efficiency frontier so the frontier line in Figure 3.1 is parabolic rather than straight.

The issue of assigning weights is tackled in DEA by assigning a unique set of weights for each DMU. The weights for a DMU are determined using mathematical programming as the weights which will maximize its efficiency subject to the condition that the efficiency of other DMUs (calculated using the same set of weights) is restricted between 0 and 1. Let there be N DMUs whose efficiencies have to be compared. Let us take one of the DMUs, say the m^{th} DMU, and maximize its efficiency as the definition above (Ramanathan, 2003). Then the mathematical programme of DEA can be presented as Equation 3.1. In this study, the efficiency measure E_m is equal to the weighted sum of two outputs (income and visits) divided by the weighted sum of three inputs (cost, opening hours and facility area).

Equation 3.1 Mathematical form of DEA

$$\max E_m = \frac{\sum_{j=1}^{J} v_{jm} y_{jm}}{\sum_{i=1}^{I} u_{im} x_{im}}$$

subject to

$$0 \le \frac{\sum_{j=1}^{J} v_{jm} y_{jn}}{\sum_{i=1}^{I} u_{im} x_{in}} \le 1; n = 1,2,\ldots,N$$

$$v_{jm}, u_{im} \ge 0; i = 1,2,\ldots,I; j = 1,2,\ldots,J$$

where

E_m is the efficiency of the m^{th} DMU,

y_{jm} is j^{th} output of the m^{th} DMU,

v_{jm} is the weight of that output,

x_{im} is i^{th} input of the m^{th} DMU,

u_{im} is the weight of that input, and

y_{jn} and x_{in} are j^{th} output and i^{th} input of the n^{th} DMU, $n = 1, 2, \ldots,$ (Note that here n includes m)

3.2.3 Functions of DEA

The usefulness of DEA lies in the way it can help to improve performance through information yielded and the fact that it outweighs ratio analysis and other parametric approaches. The functions and advantages of DEA are fivefold.

• ***Overall performance measurement***

The main characteristic of DEA is to integrate multiple inputs and outputs to yield a single indicator presenting the overall performance of a DMU. A DMU may be deemed as a good performer even when its performance on individual performance

indicators is not outstanding since its all-round performance is taken into account. DEA thus gives a more balanced approach to performance measurement (Charnes and Cooper, 1994; Athanassopoulos and Ballantine, 1995). DEA also enables management to identify best observed practices based on the overall performance rather than on the performance of individual performance indicators. The performance targets identified by DEA are therefore more realistic since multiple inputs and outputs are considered simultaneously. (Thanassoulis *et al.*, 1996).

Furthermore, as mentioned in 3.2.1, unlike statistical approaches, the evaluation provided via the DEA model conforms to the actual performance of the best performing (frontier) organizations (Bowlin, 1999). These best practices then become the benchmark for judging the performance of the other organizations. This is done by constructing a frontier comprising linear segments that interpolate between those observations where the highest ratios of output to input are constructed. The resulting DEA frontier thus 'envelops' all the observations (Smith and Street, 2005).

• ***Objective weighting method***

Stone (2002) and Smith and Street (2005) argue that, in the public services there is rarely a consensus about the choice of a set of weights reflecting the values that are attached to each input and output. In order to embrace multiple incommensurate inputs and outputs, a set of weights are required to be defined and this can be difficult, particularly if a common set of weights to be applied across the set of organisational units is sought. Under the principle of mathematical programming, this problem can be resolved by arguing that individual units may have their own particular value systems and therefore may legitimately define their own peculiar set of weights

(Boussofiance, 1991). DEA allows an individual DMU to select the most advantageous weights for maximising its performance score and each DMU's efficiency is judged against its individual weighting system which emphasises the particular strengths. Some constraints are that weights must be positive and weights of one DMU must not result in an efficiency score larger than 1 for any other company (Charnes and Cooper, 1994). Thus, the type of production function and the weights of each DMU are allowed to vary in such a way to show the performance of each unit in the best possible light and the resulting efficiency score is the best attainable efficiency level for each DMU (Bowlin, 1999).

• ***Virtual weight and its functional***

In running DEA, inputs and outputs are being multiplied by some sets of weights. The results of these multiplications are so-called virtual inputs and virtual outputs. The efficiency score is then obtained by dividing total virtual output by total virtual input while taking care that both of them are in some way normalized. The contribution of each input and output variable makes to the DEA rating for that unit is determined by taking the proportion that each input and output variables virtual weight is of the total of the virtual weights assigned to each unit (Despic, 2004). For example, one facility has the following optimal virtual weights: cost 0.00007, area 0.00014, and opening hours 0.00006. The total of these virtual weights is 0.00027. Thus, cost's proportion of and contribution to the DEA rating received by this facility is 25.9% (0.00007/ 0.0027). The virtual weight proportion is used instead of the absolute value of the weight because the absolute value is not comparable across business segments since it may be influenced by the size of the input or output

variable value similar in effect to the coefficient in regression analysis (Bowlin, 1999).

The inputs and outputs to which higher virtual weights are assigned are those which it wishes to be weighed most heavily in its comparison with other units. They thus give indications of particular areas of good practice (Boussofiance, 1991). The virtual weights also show whether changes of parameters in the appropriate direction will lead to an immediate improvement in efficiency – when multipliers are non-zero - or whether small changes in parameters values will lead to no immediate effect – when the multiplier is equal to zero. Thus, the question of what influences the efficiency position of an inefficient (in fact, any) DMU is best answered by looking at the virtual weights (Staat and Maik, 2000).

• ***Identifying benchmarking partners***

According to Metzger (1993), only units with a similar input-output structure should be compared to ensure that the performance evaluation is realistic and focuses not on extreme performance differences but on achievable best practice. The second function of virtual weights is to demonstrate the input/output orientation of the inefficient unit, and then select the appropriate benchmarking partners which have the similar input/output orientation as the inefficient unit but appear efficient. Given the fact that the peer units are relatively efficient and have a similar input/output orientation, they should provide examples of good operating practice for the inefficient unit to study (Boussofiance, 1991). Since DEA benchmarks only compare observations with a similar input-output structure, the often raised concern that benchmarking results in meaningless comparisons does not apply to DEA. Moreover,

the comparison of virtual weights need not refer to the DEA concept and so they may be suitable for the lay-person (Staat and Maik, 2000).

3.2.4 Deficiencies of DEA

Even though DEA is one of the most well-developed global measure approaches, it has some theoretical pitfalls. The following five deficiencies are highly related to the applicability of the integrating performance measurement approach proposed by this research.

• ***Problem of weighting flexibility***

In conventional DEA the weights are allowed to vary freely from unit to unit, so that each organization is evaluated in the best possible light. This flexibility in the choice of weights is both a strength and a weakness. It is a strength for if a unit turns out to be inefficient even when the most favourable weights have been incorporated in its efficiency measure then this is a strong statement and in particular the argument that the weights are inappropriate is not tenable. It is a weakness because a judicious choice of weights may allow a unit to be efficient, but there may be concern that this has more to do with the choice of weights than any inherent efficiency (Boussofiance, 1991).

Indeed we frequently find in unconstrained DEA that the highest efficiency score for an organization can be secured simply by assigning unreasonably low or excessively high values to the multipliers in an attempt to drive the efficiency rating for a particular DMU as high as possible. For instance, the efficiency may be achieved by placing all the weight on one or more inputs (outputs) on which it performs well

and assigning a zero weight to one or more inputs (outputs) on which it performs poorly (Allen *et al.*, 1997). An unbalanced set of weights is a signal that, although efficient, the organization may be achieving that efficiency by employing an unusual mix of inputs and outputs, and may be far from efficient in the allocative sense (Smith, 1990). It is also problematic because the efficiency ignores the performance of the unit on the remaining inputs and outputs, perhaps to the detriment of performance in other dimensions (Thanassoulis, 1996). Moreover, according to Smith and Street (2005), this flexibility is to the detriment of a common base for comparison. They argue that the lack of a single set of weights implies that it is never appropriate to rank DEA efficiency scores in a conventional 'league table' format.

The problem of weighting flexibility has led to the development of methods for using weights restrictions in DEA assessments to ensure that the DEA efficiencies estimated reflect the performance of units on all, and not just a subset, of the input and output variables (Thanassoulis, 1996). However, this is not a straightforward problem for the decision makers. It is difficult not only because the priorities might change all the time, but also because they lack the means to quantify the weights properly or a consensus is difficult to achieve in terms of developing a common set of weights (Smith and Street, 2005).

• ***Limitation on the number of inputs and outputs***

The selection of inputs and outputs can affect the discriminating powers of DEA as for effective discrimination the number selected needs to be small compared to the total number of units. This arises due to the flexibility in the choice of weights in determining the efficiency of each individual unit. In seeking to be seen to be efficient

a unit can allocate almost all its weight to a single input and output. DEA may, in turn, report nearly all DMUs as being efficient and lead to little discrimination between the various units (Thanassoulis *et al.*, 1996; Boussofiance, 1991). Therefore, the relationship between the number of input and output variables to the number of DMUs studied should not exceed a certain upper limit. In practice, this limit generally is a ratio of 1:2 (Richards, 2003).

• ***Target setting for DEA-efficient units***

According to Thanassoulis *et al.* (1996), DEA offers no view as to what improved input-output levels may be achievable by a DMU which is already relatively efficient. For DEA-efficient units optimal DEA weights offer some guidance as to where, but not the extent, to which, a unit might improve further its performance. Such guidance can be reinforced by the use of ratio performance indicators and this is one area where the methods can complement one another. In this context, performance indicators could enhance DEA on target setting by identifying specific aspects of operations on which an already efficient facility may be weak. It can be done in a straightforward manner by comparing a DEA efficient facility on specific performance indicator values with other facilities. The combined performance indicator and DEA information, reinforced by judgement, is useful for indicating the areas where a relatively efficient facility may strengthen its performance further, even if precise improved input and output levels cannot be estimated for a facility already DEA efficient

• ***Principle of radial movement***

DEA focuses on the single radial contraction level which is the maximum amount by which all controllable resources can be reduced simultaneously and keep the same resource mix; hence even if just one resource is being used at the right level, then the unit can escape identification as inefficient. Hence truly inefficient units can be identified in error as benchmarking partners for the other units, with possible accompanying losses if those processes are replicated (Bell and Morey, 1994).

• ***Effect of unusual or incorrect data***

The efficient boundary in DEA is established by comparing the observed input-output correspondences at various facilities. If the 'input/output mix' of an efficient facility is unusual and not found in many facilities, the facility's position on the efficient boundary will be reflecting its unusual input-output mix rather than efficient performance relative to any other facilities (Boussofiance, 1991). Also, DEA assumes no data errors or measurement errors in the observations. This raises the risk that incorrect outlying data points that form part of the frontier can adversely affect the results.

3.3 Sport Equity

3.3.1 What is Sport Equity?

According to Sport England (2006c), "sport equity is about fairness in sport, equality for access, recognising inequalities and taking steps to address them. It is about changing the structure of sport to ensure that it becomes equally accessible to all members of society, whatever their age, gender, race, ethnicity, sexuality, or

socio-economic status..." Theoretically, equity is a concept more complicated than such a policy guideline. As argued by Gratton and Taylor (2000, p.118), "the equity argument has always been a problem for economists because it involves essentially normative arguments and value judgements". Nicholls (2001) also emphasised that equity is clearly a subjective concept, open to multiple, sometimes competing, interpretations. It is therefore necessary to adopt a definition of equity in order to analyse it.

In the broader context of public service delivery, several typologies of equity have been suggested by different researchers. Crompton and Wicks (1988) attempted to integrate various approaches and offered a taxonomy of the following four equity models:

• ***Compensatory equity***: A compensatory, or need-based approach to equity implies that the unequal should be treated unequally. It involves allocating services so that economically disadvantaged groups receive extra increments of resources. The operational objective of this allocation guideline is to increase the compensatory role of public leisure services in order that opportunities for the economically disadvantaged may be improved.

• ***Equality***: Equity based on equality can be operationalised in one of two ways. The allocation of inputs to services in an equal manner, regardless of geographic area or the socio-economic characteristics of residents, is analogous to the notion of equal opportunity. Output equality, in contrast, requires equality of condition after receipt of service. It measures the benefits received by residents as a result of public service provision.

• ***Demand***: As an equity allocation guideline, the demand approach involves rewarding those who demonstrate an active interest in a service or facility, whether through use (the economic model) or vociferous advocacy (the political model). An economic model implies that the maximum number of citizens is likely to benefit if it is adopted. By contrast, a political model is often favoured by elected officials who see it as an appropriate means to improve their public accountability.

• ***Market equity***: This model illustrates the potential influence of market forces on service distribution. It may be operationalised in three ways. First, resources may be allocated on the basis of the amount of taxes paid. Second, direct prices are imposed and residents can buy as little or as much of a service as they wish, i.e. the concept of 'willingness to pay'. The third operationalisation of market equity is a least-cost approach which seeks to deliver a leisure service at the lowest cost.

Because subjective judgments are involved, there probably cannot be any 'right' or 'wrong' concepts of equity, only different concepts. If self-interest is the primary determinant of preference, it seems reasonable to expect wealthy citizens to favour market equity, middle income citizens to favour equal opportunity, and poor citizens to favour compensatory equity. Ideologically, preferences might be generalised to state that the conservative tendency is toward market equity, while the liberal tendency is toward compensatory equity (Crompton and Wicks, 1986). Veal (2002) also mentioned that the equity argument is at the heart of the difference between the left and the right wings in politics. The right believes that a considerable degree of inequality is equitable, because it reflects the rewards given for effort and risk-taking, whereas the left thinks that the level of inequality generally seen in Western societies is inequitable and therefore unacceptable.

In the UK, according to Gratton and Taylor (2000), it is Rawlsian equity (i.e. compensatory equity or need-based approach) – seeking to improve opportunities for the poorest in the society – that is dominating current policy in the form of 'social inclusion'. Rawlsian equity is therefore the primary basis on which this research is founded. Rawls (1971) suggested that service allocation decisions should be based on the 'difference principle' which argues for the redistribution of resources across society's members in order to improve the relative condition of those with the least resources (Crompton and Lue, 1992).

3.3.2 Sport Equity: Measurement and Benchmarking

Previous empirical equity analyses have been of two main kinds- 'outcome' and 'process' studies (Nicholls, 2001). 'Outcome' studies have focused on the distribution of various resources relative to the socio-economic characteristics of residents. 'Process' studies, in contrast, have been less concerned with who is or is not impacted by inequity, and more with the reasons underlying distributions of resources. In England, two measurement frameworks are currently available for the measurement and benchmarking of sport equity. The first is Sport Equity Index (SEI) which aims to measure the general sports participation rates of certain social groups. The second is the NBS which aims to measure how representative facility usage is by certain target groups. These two frameworks are both examples of 'data benchmarking', identifying what performance is like relative to national benchmarks. They are also both 'outcome' measurements, i.e. the outcome of the distribution of public leisure services is considered, in terms of variations between different socio-economic groups, rather than the processes underlying this distribution.

The SEI is formulated by Sport England to provide the evidence base to underpin sports equity policy and to assist those involved in sports development to better understand the levels of inequity that currently exist in sports participation. The analysis is based on national sports participation data taken from the General Household Survey and Sport England's Young People and Sport survey. Participation rates are converted into 'Sports equity indexes' that show the relative propensity of different socio-demographic groups in the population to take part in sport compared with the average participation rates (as shown in Table 3.1).

Table 3.1 Sport Equity Index: 1996 and 2002 [a]

Target groups	1996				2002			
	Casual		*Regular*		*Casual*		*Regular*	
	Rank	Index	Rank	Index	Rank	Index	Rank	Index
16-19 years	1	172	1	200	1	169	1	187
Norm (average)	.....							
Ethnic minority	9	89	9	98	11	79	11	79
Social class DE [b]	12	73	12	72	12	69	12	68
60-69 years	13	65	13	71	13	65	13	63
Disabled	14	64	14	62	14	61	14	58
+70 years	15	27	15	30	15	34	15	35

Notes: [a] extracted from Sport England (2002a).

[b] Social class DE in 1996 relates to the new National Statistics Socio-Economic classification (NSSEC) 6, 7 and8 in 2002.

For example, a group with an index of 169 means this group is 69% more likely to take part than the population as a whole. The SEI also provides a ranking scale to provide an indication of which groups in the population are participating at levels above and under expectation and thus need to be targeted in order to improve

participation levels. The ranks range from 1 to 15: the smaller the number the higher the participation rate is. Ranks from 1 to 8 denote the participation rate is higher than the norm (average participation rate); ranks from 9 to 15 denote the participation rate is lower than the norm. Finally, participation in sport is classified into 'casual' (at least one occasion in the last 4 weeks), 'regular' (at least once a week in the last 4 weeks) and 'frequent' (at least three occasions per week in the last 4 weeks) (Sport England, 2002a).

In the NBS, sport equity is measured through a set of PIs in a performance dimension labelled as 'access'. The access PIs are mainly 'representativeness' ratios, which comprise the percentage of visits by a certain type of user divided by the percentage of the facility's catchment population of that type. A PI score of less than 1.0 indicates that the group is under-represented and over 1.0 indicates over-represented in their use of the centre in comparison with the catchment population. The other type of access PI is a 'share' ratio (i.e. simply the percentage of visits of a type of users), which is used because the better alternative, the representativeness ratio, is not available. The percentage of total visits by disabled people is an example. This kind of share ratio needs to be interpreted carefully since the percentage of visits by the target group may be low simply because this group is a low proportion of the local community. Finally, one important part in the measurement of access is the determination of each facility's catchment area (Sport England, 2000a).

3.3.3 Target Groups: Rationale and Evidence

As mentioned in the first section, two sets of NBS PIs were proposed in the CPA: five for 'equity' and two for 'value for money'. The target groups selected for analysis in this research are in line with this designation. The rationale for and the evidence of inequity for the five target groups are discussed below. However, it is worth mentioning gender inequity. Participation statistics consistently reveal that women have a lower prevalence of taking part in sport than men. Consequently, women (and girls) are often targeted as a group in order to improve sport equity on gender grounds. However, analysis of the NBS data has shown that historically women are the majority user group of local authority sport centres generally and swimming pools specifically, so gender inequity is not justified as a major social inclusion PI for such facilities.

• ***Social class DE***

The PAT 10 report on Sport and Arts mentioned that people from lower socio-economic categories in general participate less in the arts/sport and therefore benefit less from Government support for the arts/sport (DCMS, 1999). Collins (2003) also emphasised that poverty is the core of social exclusion. As argued by Gratton and Taylor (1985), public sport facilities mainly benefit the higher socio-economic groups who are well informed and mobile enough to make use of them without social barriers. In addition, some previous research identified the relationship between occupational status and leisure participation. For example, Clarke (1956) and Burdge (1969) found that this relationship is especially significant with activities closely tied to social status and requiring financial resources. Kelly (1996) also pointed out that income and occupational status are related to opportunities and limitations to leisure participation.

In Britain, in 1996, those in the professional socio-economic group were almost three times as likely to participate in sport (excluding walking) as those in the unskilled manual socio-economic group, 63% compared with 24% (Sport England, 1999b). Examining the 2002 General Household Survey, Collins (2004) pointed out a major gap between the high participant professional and managerial social groups AB and the semi-skilled and unskilled groups DE. The SEI updates this - social groups AB are much more likely to take part in sport than social groups DE. The latter had a ranking of twelfth both in 2002 and 1996 (casual or regular participation) (see Table 3.1).

In terms of the use of public sport facilities, according to Collins (2003), despite the encouragement and grant aid of the Sports Council/Sport England to promote access/sport for all, the participation rate of the social classes DE in public sports halls and swimming pools did not improve a lot over more than twenty years (from 7% in the 1960s to 8% in the 1990s). During the same period, the disparity in the participation rates of social classes AB and DE widened. Furthermore, the 1997 Sport England survey showed that 10% of visitors came from the AB groups (which represent 6% of the population of England as a whole), while only 8% of visitors were from the DE groups (19% of the population as a whole). In addition, the visitors to pools were more unrepresentative of groups DE than those to halls (Sport England, 2000a).

• ***Young people***

According to the 1996 and 2002 SEI, the group most likely to take part in sport was those aged 16 to 19, ranked as first for both casual and regular participation (see

Table 3.1). Relatively high representativeness by 11-19 year olds is found in the usage of public sport facilities in the NBS statistics. It appears that the argument regarding youth as a target group is less because of the evidence of inequity but because of other reasons.

If we check trends in the 1990s, the decline in participation was most pronounced amongst the youngest age group, with participation in the previous four weeks (excluding walking) falling from 82% in 1990 to 72% in 2002 for those aged 16 to 19 years (Sport England, 2004). This causes particular concern and disappointment given the focus of public policy on young people and sport during the 1990's (Sport England, 2004). According to Sport England (2000b), many of the benefits of sport, for example, health, social regeneration, lifelong learning and reductions in criminal behaviour, can not be realised if appropriate skills and a positive attitude to sport are not developed at a young age. It is widely believed that young people in particular are liable to engage in antisocial activities in their leisure time, that is, activities which impose external cost on others. Such externalities may be short-term and immediate (e.g. vandalism and hooliganism) or long term (e.g. becoming involved in criminal subcultures or harmful drugs) (Bailey, 2005). Three recent UK policy-related reviews of the potential social value of sport (Collins *et al.*, 1999; DCMS, 1999; Sport England, 1999b) all list the prevention of youth crime as an issue to which sport can make a contribution, reflecting a widespread belief in the 'therapeutic' potential of sport. Veal (2002) argued that children and young people merit a special attention under the heading of equity because they are still dependent on their parents but their sport opportunities will not be entirely dependent on their parents' means.

• ***Ethnic minority groups***

Leisure policy for ethnic minority groups has tended to be oriented towards alleviating or compensating for general social deprivation or social discrimination (Veal, 2002). Two contrasting explanations, ethnicity and marginality hypotheses, have been the dominant frameworks for explaining differences between ethnic groups in terms of their leisure involvement (Lee *et al.*, 2001). The marginality hypothesis suggests that ethnic differences in leisure participation are a function of minority groups' subordinate socio-economic status stemming from historical discrimination. The ethnicity hypothesis states that ethnic differences in leisure participation stem from a culturally based value system, norms and leisure socialisation patterns, rather than socio-economic differences. Based on a comprehensive literature review on the issue of ethnicity and leisure, Floyd (1999) suggested that many studies to date have found greater support for the ethnicity hypothesis rather than the marginality hypothesis.

In England, a national survey conducted by Sport England during the period from 1998 to 2000 showed that, for ethnic minority groups, the overall participation rate in sport was 40% compared with a national average of 46% (Sport England, 2000b). Furthermore, there is evidence that this gap is growing - from 1996 to 2002 the indices for ethnic minorities have fallen 10 points (for casual participation) and 19 points (for regular participation) and the ranks have both fallen from ninth to eleventh (see Table 3.1). This reflects a substantial drop in participation rates of ethnic minority group over the last six years. However, in terms of the use of public sport facilities, ethnic minorities were well represented in the use of sports halls in 1997,

with 5.3% of users being from either a black or ethnic minority background. This compared to a population profile of 5.2%. However, black and ethnic minorities were underrepresented in their use of swimming pools in 1997, with only 2.8% of users classifying themselves as non-white (Sport England, 2000a).

• ***Older people***

According to Collins (2003), exclusion among older people comes from poor health, poverty and disability, all of which may be compounded by isolation and poor mobility. Veal (2002) argued that the net benefits to the community from providing leisure opportunities for the elderly are likely to be even greater than in providing for younger age-groups. However, participation in sport declines significantly with age.

According to the 2002 General Household Survey, participation in at least one activity (excluding walking) during the previous four weeks among 60-69 year olds was half that of 30-44 year olds (27% compared with 54% respectively) (Sport England, 2006a). From the SEI, the least likely to participate casually or regularly were those adults aged 70 and above. The group 60-69 year olds also had a low ranking- thirteenth both in 2002 and 1996 (casual or regular participation) (see Table 3.1). In terms of the use of public sport facilities, the 1997 Sport England survey showed clear under-representation of people over 45 in both halls and pools (26% and 31 % compared with 44% of the population as a whole). Almost 9 out of 10 facilities offered concessions for people aged over 60, but only 4% of hall users were senior citizens with discount cards and 6.7% of swimmers (Collins, 2003). Apart from the evidence of inequity, Sport England (2006a) proclaimed that to achieve the government's target of 70% of the population being active by 2020, one of the key

drivers is the ageing population, as by 2020 almost half of the UK population will be over 50 years old. Given the drop-off in participation associated with age, the ageing population phenomenon will have a significant negative effect on overall sports participation.

• ***People with disabilities***

Participation in sport has the potential to promote the social inclusion of disabled young people and increase their self-esteem (Sport England, 2001b). Sport and exercise offer the possibility of overcoming the stigma often associated with disability. Participation can provide the context within which young people exceed the expectations associated with their disability through demonstrations of physical skills or fitness, so emphasising an alternative, more positive, picture of the body and the self.

Two national studies of sport and disability were conducted by Sport England, one for children and young people (aged 6-16 years old) (Sport England, 2001b) and one for adults (aged 16-59 years old) (Sport England, 2002b). The first survey showed that both the overall rate of participation and the frequency with which children and young people with a disability take part in sport is lower than for young people in general. In terms of the second survey, similarly, sports participation rates for disabled adults were significantly lower than for non-disabled adults. The 2002 General Household Survey further demonstrates that, excluding walking, nearly twice as many respondents with no long-standing illness participate in sport and physical activities (51%) than those with a limiting long-standing illness (26%) during the 4 weeks before interview (Sport England, 2004). Shown by the SEI, the ranks of disabled were

very low (fourteenth) both in 1996 and 2002 (see Table 3.1). Inequities also exist in the use of public sport facilities - in 1997 only 7% of visitors in sports halls and 11% in swimming pools were disabled, compared with 22% of the population. Focusing on 16-44 year olds (to exclude age-related disability), in 1997 only 6% of visitors to pools had a disability compared with 15% of the population, while for halls the proportion with a disability was even lower at 4% (Sport England, 1999a).

3.4 Concluding Remarks

This chapter reviewed the literature related to efficiency and equity measurement as well as the theoretical concepts of DEA. DEA is mainly used for the measurement of efficiency. This study attempts to consider DEA in a broader context by adopting the DEA effectiveness model to evaluate the performance of sport equity. For instance, there may be a trade-off between the access of different target groups. DEA allows each facility to highlight its strength(s) to conclude with a best overall performance. The specific DEA models used and the research variables will be specified in Chapter 5. Then, the research findings of efficiency and equity measurement at the industrial and individual facility levels will be presented in Chapter 6 and Chapter 8, respectively.

CHAPTER 4

Theoretical Context of Customer Service Measurement

Road map of the thesis

Introduction	Theoretical context	Methodology	Results	Conclusion
Ch1 ▶	Ch2 Performance measurement overall Ch3 Efficiency & Equity Ch4 Customer service	▶ Ch5 ▶	Ch6 Efficiency & Equity Ch7 Customer service Ch8 Action research	▶ Ch9

This chapter aims to review the literature related to customer service quality and customer segmentation. Before discussing the analysis of these two dimensions, the industrial characteristics are first reviewed.

4.1 Industrial Context

4.1.1 Customer service quality

Customerism has been at the centre of the changing management of the public services in the UK (Walsh, 1994). As stated in the Citizen's Charter, the users of public services should be treated as customers who are entitled to expect high quality services and responsive to their needs (Citizen's Charter Unit, 1992). The introduction of

compulsory competitive tendering (CCT) into public leisure services in 1989 required local authorities to make the management of their sport and leisure facilities open to competitive tender. Organisations that wished to win the contract to manage these facilities had to demonstrate that they would be able to do so in an efficient and effective manner. The introduction of Best Value regime, in 1997, continued the need for quality management within public leisure services as the rationale for Best Value is not just about economy and efficiency, but also about effectiveness and the quality of local services.

However, most of the research related to service quality has been undertaken in the commercial sector. There is some evidence to indicate differences in the nature of public and commercial sector service delivery. Certainly, any attempt to conceptualise service quality in the context of public service has to recognise these differences. A primary factor which differentiates the public sector from the commercial sector is the absence of profit motive. Public services are less dominated by financial objectives. Producing revenue or raising funds is the means to an end, not the end in itself, by contrast with commercial firms, which are dependent upon profits for their continuation and success. As such, in the public sector, customer satisfaction is the ultimate goal, not the vehicle for achieving a profit (Crompton and MacKay, 1989).

More importantly, organisations operating in the public sector must ensure their services are soundly based on the needs and expectations of all stakeholders, i.e. not only customers but also broad communities (Wisniewski, 2001). Many of the current measurement models developed in the private sector aim to develop better and more systematic ways of measuring the quality of service, but the term 'service quality' in the public sector frequently refers to a broader concept which includes, at least, accessibility and availability, i.e. whether organisations provide services which are

accessible to target groups and whether a wide range of services are in place to cater for different needs. The Department for Environment, Transport and the Regions (DETR) expressly states that the use of quality management techniques will not be enough to deliver Best Value, rather local authorities will need to take account of issues such as equity and the balancing of interests (Robinson, 2002). As such, the challenge for the public sector is to consider different targets simultaneously, e.g. customer satisfaction, accessibility and financial performance, and balance distinct and often competing expectations. This context also makes the measurement of public service delivery more complicated.

In England, public sport facilities differ significantly from other leisure services, such as parks, libraries or galleries, as customers are charged to use these facilities. Moreover, these charges are usually more than an attempt to simply cover costs and for some services reflect market rates (e.g. health and fitness, function room hire) (Robinson, 2003; 2004). Over the past two decades this revenue earning capacity has led to the emergence of a quasi-commercial operating context for public sport facilities. Although often significantly subsidised by the local authority, facilities such as health and fitness suites compete directly with the commercial sector. In addition, these facilities are affected by customer choice, i.e. the general public does not have to use the leisure services offered by the public sector. This means that in order to generate revenue, sport and leisure facilities need to be an attractive provider of leisure opportunities.

The introduction of Best Value regime, in 1997, highlighted the need for quality management within public leisure services as the rationale for Best Value is not just about economy and efficiency, but also about effectiveness and the quality of local services. Since one of the main aims of Best Value is to produce customer-focused

services, it can only be achieved successfully by consulting with the customer on a regular basis and thereby ensure local accountability and continuous improvement (Robinson, 2003; Brysland and Curry, 2001). These factors make the measurement of service quality particularly important since it may help facility managers to deliver quality services so as to compete directly with those offered by the commercial sector, and respond to the requirements of central government in a more proactive way.

4.1.2 Customer segmentation

The concept of marketing has been perceived among public sector professionals as an overtly commercial concept in terms of both its origin and its nature; thus to be of limited relevance to public sector organisations concerned with the delivery of 'public-good' services (Caruana *et al.*, 1997; Laing, 2002; Walsh, 1991). Sceptics of the appropriateness of marketing concept in the sport/leisure field argue that "its application distorts a public leisure agency's objectives, is antithetical to its social service ethic and invites inappropriate commercialisation of its services" (Novatorov and Crompton, 2001, p.61). However, the adoption of private-sector-based approaches to the organisation of public services in many post-modern western economies has forced a reconsideration of the potential contribution of marketing to the delivery of public services. In the UK, the efforts of successive governments, both Conservative and Labour, to introduce private sector management practices into key areas of public service provision has been central to the renewed public sector interest in marketing. Marketing has been seen as providing some of the critical tools required by managers and professionals operating in such new environments (Day *et al.*, 1998; Kearsey and Varey, 1998; Laing, 2002).

There is also a general consensus that marketing has an important role to play in the public leisure sector. For instance, Borrie *et al.* (2002) argued that the use of market segmentation- a theoretical cornerstone of marketing- allows public leisure providers to better understand their clients' needs and to tailor their services to the diversity of those needs. Doherty *et al.* (1998) also indicated that, for the public leisure services sector to make greater, and more effective, use of marketing mix opportunities and to develop appropriate marketing strategies to guide longer term development, it is essential to develop the ability of segmenting their customer base.

Market segmentation can help facility managers to enhance their competitive advantages relative to private sector by developing more precise marketing strategies. Also, the success of customer orientation will only come to the organisation that best determines the perceptions, needs and wants of target markets and satisfies them through the design, communication, pricing and delivery of appropriate and competitively viable offerings (Kotler and Andreasen, 1991; Walsh, 1994). The aim of this research is therefore to help facility managers identifying the priorities and characteristics of different customer segments.

4.2 Customer Service Quality

The most common definition of service quality is the traditional notion that views quality as the customers' perception of service excellence. That is to say, quality is defined by the customer's impression (Parasuraman *et al.*, 1985). Nevertheless, there are a number of disagreements and differences in the measurement of service quality; therefore, this section first examines the operational merits and limitations of different measurement models. Then, terminologies relevant to the measurement of service

quality and service dimensions related to the performance of sport facilities are discussed.

4.2.1 Service quality measurement models

• Gap analysis

Service quality has most commonly been understood on the basis of the disconfirmation paradigm. Disconfirmation is typically measured as the gap (difference) between expectation and performance. Negative disconfirmation occurs when performance is less than expectation, and positive disconfirmation occurs when performance is greater than expectation (Parasuraman *et al.*, 1994). According to Zeithaml *et al.* (1993), the difference between customers' expectation and their perception of service quality is known as a service gap. Based on the disconfirmation paradigm, researchers have adopted a variety of service quality models including both inferred and direct disconfirmation models.

The inferred approach (also known as gap analysis) involves deducing separate data sets relating to customers' expectations and perceived performance. The scores for performance are then subtracted from those of expectations to form the third variable, the confirmation-disconfirmation (or difference) score. This produces a relative measure of how well the service has performed relative to what the consumer expected. The direct approach, by contrast, requires the use of summary-judgment scales to measure confirmation and disconfirmation (e.g. a Likert-type scale of 'better than expected' to 'worse than expected'). The researcher avoids the necessity of calculating difference scores, since the respondents can be asked directly the extent to which the service experience exceeded, met, or fell short of expectations. This approach thereby provides

an absolute measure of performance (O'Neill and Palmer, 2004; Yüksel and Rimmington, 1998).

One of the most widely used inferred disconfirmation models is the SERVQUAL, which is developed by Parasuraman *et al.* in 1985. Parasuraman *et al.* (1985) proposed that service quality is a function of the differences between expectation and performance along the quality dimensions from a consumer's perspective. The advantage of gap analysis is to help managers identify where performance improvement can best be targeted. Equally, if gap scores in some areas do turn out to be positive, this allows managers to review whether they may be over-supplying this particular feature of the service and whether there is potential for re-deployment of resources into features which are under-performing (Wisniewski, 2001). While gap analysis has attracted a lot of attention for its conceptualisation of quality measurement issues, it has also attracted considerable criticism.

The first concern with the use of gap analysis lies in the construct of a mathematical difference between expectation and performance. Different numerical scores can give the same gap scores, e.g. 3-5 = 2-4, but whether or not identical scores calculated from different values indicates the same perceived quality is debatable. Also, it is evident that a negative gap score (expectation > performance) is a negative evaluation, but a positive gap score (expectation < performance) or a gap score of zero (expectation = performance) may mean different things, depending on management priorities. A fiscally conservative manager may conclude that a gap score of zero is the best of the three ratings, because a positive gap score may indicate that too many resources are being invested in an unimportant service or facility. A more customer-oriented manager may seek to exceed customers' expectations in as many ways as possible (Burns *et al.*, 2003; Williams, 1998).

The second issue relates to the measurement of expectations. Some have argued that if expectations are measured after or even simultaneously with the service experience, those expectations will be biased by the experience. For instance, Cronin and Taylor (1992) argued that an accurate expectation measure can only be obtained prior to the service encounter. Additionally, from a practical perspective, a difficulty of measuring expectations is that such measurement leads to consistently high expectation ratings. That is, respondents may feel motivated to demonstrate an 'I-have-high-expectations' social norm and also indicate that they desire the maximum amount of quality on every attribute. The level of expectations, therefore, may exceed the experienced or existing level for no other reason than guests' tendency to proclaim high expectations. In this sense, Crompton and Love (1995) argued that if these scores are almost constant, then there is little point in including them on an instrument, since they will not give responses significantly different from using the perception scores alone (O'Neill and Palmer, 2004; Yüksel and Rimmington, 1998). The debates surrounding the measurement of expectation have led to the development of performance-only analysis.

• Performance-only analysis

As discussed above, the use of gap analysis has been challenged on conceptual and methodological grounds with many researchers advocating the measurement of perceptions of service quality only might provide a better indication of service quality than measuring the difference between expectation and perception. Cronin and Taylor (1992) investigated the conceptualisation and measurement of service quality and its relationship with consumer satisfaction and purchase intention. Their research showed that the performance dimensions alone predict customers' overall satisfaction at least as

well as the complete expectancy-disconfirmation model. This suggests that including the confirmation- disconfirmation calculation as an intervening variable is unnecessary and perceived performance only seems to be more straightforward and convenient.

On the other hand, even though acknowledging that it is possible for researchers to infer consumers' disconfirmation through arithmetic means (the P-E gap), Cronin and Taylor (1994) argued that consumer perceptions, not calculations, govern behaviour. This approach also overcomes some of the problems raised regarding gap analysis, i.e. survey fatigue due to two parts of the questionnaire, the statistical properties of difference scores and the ambiguity that occurs when customers indicate their expectations, i.e. customers may not discern a difference between a 'desired level' and an 'existing level' of services (O'Neill and Palmer, 2004; Burns *et al.*, 2003; Yüksel and Rimmington, 1998). SERVPERF, developed by Cronin and Taylor (1992), is one the most popular performance-only measurement models.

However, researchers have vacillated between the use of gap scores and performance-only measures. For example, Crompton and Love (1995) noted that even though performance-only measures are better predictors of satisfaction than disconfirmation measures (gap scores), the discrepancy measures provide managers with a degree of richness that is not attainable in the performance-only measures (Burns *et al.*, 2003). Lentell (2000) also argued that service managers are not usually interested in using measurement to predict overall customer satisfaction; rather, they tend to focus on the diagnostic usefulness of service quality measurements in helping them to decide priorities for service improvement. Measurement of expectations is necessary because it enables managers to direct their efforts to those service attributes where there are large gaps between customer expectations and perceptions of service. In addition, collecting both pieces of information allows for importance-performance analysis (Burns *et al.*,

2003). In sum, taking a single measure of service performance is seen to circumvent some tricky issues, but from an operational point of view, much useful information is lost when performance-only measures are taken (O'Neill and Palmer, 2004).

• Weighted gap analysis

Another criticism of gap analysis is that information about importance is not integrated in the calculation of the quality score. Crompton and MacKay (1989) stated that measuring expectations and perceptions of quality is not enough in determining satisfaction, but that the importance of individual attributes must be identified so that management resources can be properly allocated. Without considering attribute importance, one has no indication of the relative importance that respondents attach to particular aspects of a service performance. Some researchers thus have advocated that attribute importance be used to weight confirmation-disconfirmation scores in assessing service quality, and argued that it is this additional information which makes gap analysis more suited to the task of directing improvement based on what is deemed most important by consumers (Yüksel and Rimmington, 1998; O'Neill and Palmer, 2004). This kind of weighted gap analysis has been adopted by some researchers. For instance, Carman (1990) suggested that the original gap model should have been expressed as gap score multiplied by importance score. Haywood-Farmer and Stuart (1990) measured perceptions and weighted them with an importance score.

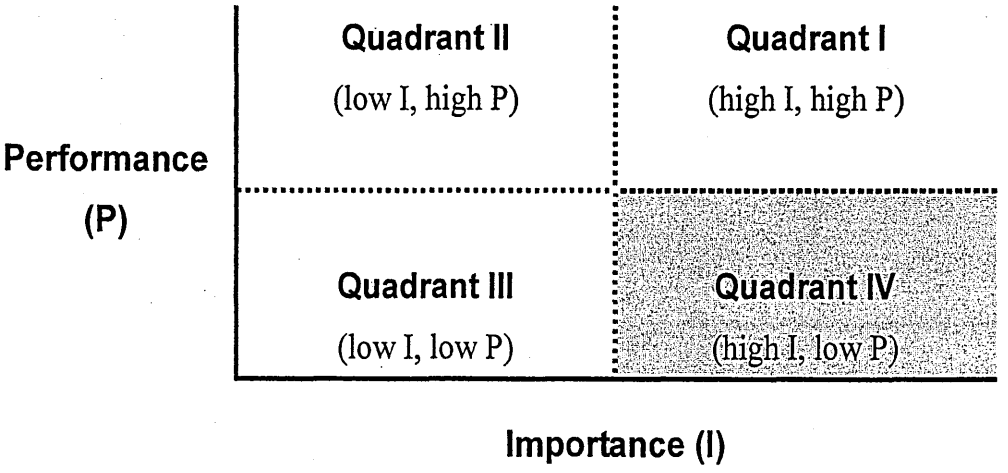
However, the multiplication approach has some shortcomings. For example, this approach is not capable of distinguishing between different numerical scores which give the same weighted gap scores, e.g. (4-5)*5 = (2.75-4)*4. Although these two combinations yield the same result, the implications in terms of satisfaction are substantially different. In this regard, scholars continue to discuss whether to include

importance in measuring attitudes. Oh and Parks (1997), for instance, commented that those who advocate inclusion mainly focus on the conceptual and realistic role of the relative importance of different attributes in human-decision processes, whereas those who dismiss inclusion tend to emphasise statistical and methodological efficiency.

• Importance- performance grid

Another key tool that has been widely used is importance-performance grid (also know as importance-performance analysis, IPA) proposed by Martilla and James (1977) as a means to develop firms' management strategies. In its essence, the importance-performance grid combines measures of attribute importance and performance into a two-dimensional grid in an effort to ease data interpretation and derive practical suggestions. This approach divides the factors into four types, with those that fall into the priorities for improvement quadrant being the focus for action. An illustrative example is shown in Figure 4.1. The average satisfaction and importance scores are shown as lines that split the grid into four quadrants. The two left hand quadrants contain services that are considered relatively less important. The top right hand quadrant contains those services that are considered important and which users are also relatively satisfied with. The bottom right hand quadrant shows those services users are less satisfied with but which are also relatively important – these can be seen as the priorities for improvement. The importance-performance grid has gained popularity over recent years for its simplicity, ease of application and diagnostic value (Oh, 2001; O'Neill and Palmer, 2004; Mori, 2002). However, this approach also raised a lot of debates and suffered from various criticisms.

Figure 4.1 Illustration of grid analysis



First, the placement of the axes creating the boundaries of four quadrants has a decisive impact on the strategic interpretation of the analysis. As Martilla and James (1977) prescribed, the majority of researchers have used the mean values of observed actual importance and performance ratings when determining the cross-hair point in the grid approach. They also recommended use of the median, instead of mean, when there is an insufficient amount of variance or when the ratings show a non-normal distribution pattern. Other researchers, however, suggested the use of scale means rather than actual means to draw the axes. The argument is that using scale means is likely to provide easy and valid interpretation of the data since the actual means are likely to differ in most cases (Oh, 2001). Decisions regarding the placement of two axes will always be somewhat subjective and the results and interpretations could be dramatically different depending on which criteria are used. Consequently, researchers need to caution readers about this hidden problem (Joseph *et al.*, 2005).

Second, according to Oh (2001), many studies have not considered potential relationships between importance and performance. The problem is that importance

may possess some interactive effects on performance ratings. Mathematically, a high positive correlation will cause the attributes to spread along the solid line stretching through Quadrants 1 and 3, whereas a high negative correlation between the two variables causes the attributes to scatter along the dotted line. This suggests that a positive correlation between the two variables tends to result in prescribing more suggestions of Quadrant 1 or Quadrant 3. In contrast, a negative correlation will tend to cause over-prescribing Quadrant 2 or Quadrant 4. As such, the suggestions yielded from the grid approach may tend to be affected by the causal relationship between importance and performance and probably offer serious misinformation.

Third, the traditional grid approach does not distinguish between attributes falling within the same quadrant. The question of validity arises because the grid approach assigns one of four marketing strategies to every attribute, but attributes falling close to the axes are treated in the same way as attributes falling neatly within the respective quadrant. Tarrant and Smith (2002) therefore proposed a modified version of the grid approach that includes, in addition to mean values, a measure of standard error. The proposed model becomes sensitive to the variance in response and to the number of respondents in the study, and also increases the validity of IPA to distinguish attributes that truly fall within a single quadrant.

Finally, Martilla and James (1977) views organisations' 'better-than-wanted' performance as 'overkill'. However, some service attributes, e.g. friendliness of staff, can often enhance customer satisfaction without committing the overkill of resources. On the other hand, from a practical standpoint, it is difficult for an organisation to always perform exactly at the importance level of the attribute in order to prevent possible overkill or to reduce needs for additional resource allocation to the attribute. Particularly in the competitive business environment, performance that exceeds the

level of importance may be desirable when extra performance often does not require corresponding input resources (Oh, 2001).

The approach adopted by the NBS is gap analysis. The shortcomings of gap analysis have been revealed in the above discussions, and alternative models which have the potential of overcome these shortcomings have also been introduced. However, the existence of different models is causing confusion among practitioners, as no consensus has been reached on which framework is best suited to assess service quality. There also remains an uncertainty as to the reliability and validity of the alternative models. It is important since the success of quality improvement programmes relies heavily on reliable and valid information (Yüksel and Rimmington, 1998). Consequently, the objectives of this research are twofold: (i) providing an insight into the industry's performance in customer service quality; and (ii) examining whether alternative service quality measurement models yield consistent results and discuss its implications.

4.2.2 Concepts relevant to service quality measurement

Many of the early research in this area attempted to distinguish and clarify two terms- 'service quality' and 'customer satisfaction'. There is also a great deal of discussion and disagreement in the literature about the distinction between these two concepts. For example, Crompton and MacKay (1989) have argued that customer satisfaction (a psychological outcome) is distinguished from service quality (an attribute of the service) by the service experience which can not be restricted to service quality dimensions. According to Johnston and Heineke (1998), the model underpinning the development of satisfaction formation is the disconfirmation theory, which holds that

customer satisfaction with a service is related to the size of the disconfirmation related to initial expectations.

On the other hand, service quality is usually expressed as a function of customers' expectations of the service to be provided compared with their perceptions of the actual service experience. According to Tsitskari *et al.* (2006) and Tomas *et al.* (2002), service quality is generally accepted as a long run, overall evaluation of the service and customer satisfaction is a transaction-specific measure of the experience. Liljander and Strandvik (1997) argued that customers' evaluation of service quality can be described as a cognitive process where customers consider the goodness/badness of different components of the service. Satisfaction, on the other hand, is believed to contain an affective dimension too. However, some authors have concluded that there is no distinction between satisfaction and service quality. For example, Parasuraman *et al.* (1988) agreed with both of these concepts but their research findings indicated that the same evaluation judgment occurs in each case. Also, according to Kouthouris and Alexandris (2005), a number of studies in the services marketing literature have reported that these two concepts are strongly related.

A second issue of concern with this research is the use of the term 'importance' or 'expectation'. Similar to the satisfaction versus service quality debate, little agreement has been reached, and the definitional issues are still being tested and debated. A number of researchers have often used the two concepts interchangeably when measuring and interpreting importance. This mixed use of the two concepts seems to have originated in the similarity between importance-performance analysis and SERVQUAL that involve similar concepts. Some evidence exists to support the conceptual difference of the two concepts. In his review of service quality research, Oh (2001), for example, distinguished importance from expectation such that the former is

a desired outcome and the latter a tolerated outcome in consumer perceptions of service quality.

In this research, one objective is to examine whether alternative measurement models yield consistent results, and the extent to which users are 'satisfied' with attributes of the facility as well as how 'important' these attributes are to them are both considered. It is therefore argue that adopting the term 'service quality' rather than 'customer satisfaction' can reflect better the research context. Furthermore, the NBS uses importance-performance gap scores, in preference to expectations- performance, on the ground that customers understand importance much more clearly than expectations.

4.2.3 Service quality dimensions

Similar to the disagreements between various service quality measurement models and terminologies, the determination of service quality dimensions and attributes is an ongoing debate. Determining appropriate dimensions in a service quality measurement model is important as they may cause service providers to direct their scarce resources to weak service dimensions. As mentioned above, the *a priori* NBS survey consists of six service dimensions, where not only the traditional service dimensions are measured (i.e. quality of facility, cleanliness, staff and value for money), but also the priorities of public service are taken into account (i.e. accessibility and availability). Since public sport facilities include a range of different attributes compared with other services, here, the NBS dimensions are compared with three similar models so as to clarify the service dimensions selected in this study.

First, the Centre for Environmental and Recreation Management (CERM) at the University of South Australia has developed prototypes for performance indicators of

efficiency and effectiveness that can be applied to sports and leisure centre management. Effectiveness indicators, based on the principles of customer service quality, measure customers' expectations compared to their perceptions of the centre's actual performance. According to Howard *et al.* (1996), the CERM customer service quality questionnaire includes fifteen core attributes, which could be grouped statistically into the following four dimensions: 'core service' (i.e. programme information, start/finish time, activity range, organisation, facility comfort, value for money and equipment quality); 'staff quality' (i.e. staff responsiveness, staff presentation, staff knowledge and officials); 'general facility' (safe parking and facility cleanliness) and 'secondary service' (food/drink and child minding). Since the CERM model focuses on leisure industry sectors, particularly public sports and leisure centres, many attributes emerged from the CERM model are comparable with those of the NBS.

Alternatively, according to Lentell (2000), the '3Ps' (physical evidence, process and participants) model suggested by Booms and Bitner (1981) can be used to outline the major service dimensions of sport facilities. 'Physical evidence' is comprised of the facility itself and the equipment in it. This includes not only the activity areas and equipment in them, but also supports areas such as reception, changing areas etc. The service quality literature tends to include these items in the category of 'tangibles', i.e. touchable items, a category which also embraces any physical products which are consumed as part of the service. Examples of the latter items would be a meal in the cafeteria. 'Process', such as bookings, tuition of a class or serving customers in the bar/cafeteria is directed at customers and requires their active participation. Finally, 'participants' include all service personnel who have contact with customers.

Grönroos (1984) also proposed that customers make service quality evaluations by assessing two aspects of service, which he called 'technical' and 'functional'

dimensions. The technical dimension is associated with what is actually provided, while the functional dimension relates to the way it is provided. Grönroos (1984) argued that technical quality is a necessary but not sufficient condition for higher levels of service quality and that functional quality is likely to be more important than technical quality, if the latter is at least of a sufficient standard. For instance, even if the facility is clean and the programme delivered meets high operational standards (technical quality) a negative impression can be generated through poor staff-customer interaction (functional quality).

A comparison between the NBS service quality dimensions and the above-mentioned frameworks is demonstrated in Table 4.1.

Table 4.1 Service dimensions of the NBS and similar models

NBS	CERM CSQ (Howat *et al.*, 1996)	3Ps (Booms and Bitner, 1981)	Grönroos (1984)
Accessibility	core service	process	functional
Availability	general facility; core and secondary services	physical evidence	technical
Quality of facility	core and secondary services	physical evidence	technical
Cleanliness	general facility	physical evidence	technical
Staff	staff quality	participants	functional
Value for money	core and secondary services	physical evidence	technical

This comparison is based on certain attributes rather than all attributes in one dimension. For instance, 'activity available at convenient time' (accessibility) in the

NBS is similar to 'start/finish time' in CERM CSQ; however, no item in CERM CSQ is equivalent to 'activity charges/fees' (accessibility) in the NBS. From this comparison, we can argue that the service quality dimensions are contextual-based and not universally applicable. Since it is difficult to say one approach is clearly superior to the other, from the author's point of view, different approaches should be considered while deciding appropriate service dimensions. The determination of service dimensions and attributes in the original NBS framework is made through consultation with industry representatives. Based on the customer survey data, this study attempted to extract meaningful factors via multivariate statistical analysis.

4.3 Customer Segmentation

Market segmentation has long been considered as one of the most fundamental concepts of modern marketing. Segmentation is generally defined as "the process of dividing a large, heterogeneous market into more homogeneous groups of people who have similar wants, needs, or demographic profiles, to whom a product may be targeted" (Mullin *et al.*, 2000, p.102). This section will discuss and justify the segmentation method and variable applied, follow by specifying the underlying service dimensions used for segmentation. Then, a comparison between this research with other existing sports segmentation studies helps to demonstrate what new this study adds to the existing sport marketing literature.

4.3.1 Segmentation methods

The delineation of existing market segments within the marketplace usually necessitates the use of two market segmentation methods: *a priori* (or conceptual) and *a*

posteriori (or data-driven) market segmentation (Calantone and Mazanec, 1991; Mazanec 1992). With *a priori* segmentation method, study subjects are partitioned into groups by attributes selected based on researchers' prior knowledge of the segments. On the other hand, when *a posteriori* approach is used, the starting point is typically an empirical data set. Quantitative analysis (normally a factor-cluster statistical analysis) is then applied to this data in order to identify the sizes and number of visitor segments that are previously unknown. Dolnicar (2002) revealed that data-driven segmentation has received increased attention in the last decades. As Ketchen and Shook (1996) illustrated in their surveys on the use of cluster analysis for market segmentation, the number of studies has increased dramatically ever since the market segmentation concept gained wide popularity in the early 70's. The number of empirical studies conducted is increasing and so is the number of taxonomies constructed with the goal of identifying the optimally suited target markets. With *a priori* segmentation approaches not having much potential for competitive advantage any more, attention has been drawn to the construction of multivariate taxonomies (Dolnicar, 2002).

In this research, given its exploratory nature, *a posteriori* segmentation approach is adopted since the characteristics of customer segments are previously unknown. Another underlying argument is that, according to a literature review conducted by Hsu and Lee (2002), most *a priori* studies use demographic variables as the bases of segmentation. By contrast, *a posteriori* segmentation studies employed primarily benefits and motivation in grouping respondents. As the segmentation variable (which will be discussed later) in this paper is based on needs/benefits, *a posteriori* approach seems to be more suitable. In addition, according to Chen (2003), two analytical stages in sequence are often found in data-driven segmentation studies. In the first stage (known as 'segment revelation'), researchers identify various segments from a

heterogeneous population by using a segmentation method. The second stage (known as 'segment diagnoses') further profiles the distinct characteristics of the derived segments, and descriptive and inferential statistics are often utilised. The resultant data might help practitioners draw further managerial strategies. This paper will be basically in line with this two-stage analytical approach.

4.3.2 Segmentation variables

The key to identifying the customer segments is to select suitable variables which discriminate people having different response characteristics to a product or service. A number of different segmentation variables have been utilised in the market segmentation literature. For example, according to Kotler (1994), major variables that might be used include geographic, demographic, psychographic and behavioural characteristics etc. Several researchers (e.g. Frank *et al.*, 1972; Sharma and Lambert, 1994; Brassington and Pettit, 2000) have proposed dichotomies for classifying these variables. Basically segmentation variables can be grouped into two categories: objective (or physical, identifiable) and subjective (or behavioural, needs/benefits) attributes. The former includes demographic, socio-economic or geographical factors etc., and the latter includes benefits sought, expectations, preferences, attitudes or perceptions etc.

Traditionally, marketers use demographic data to segment markets, because they are standard and readily available (Bagozzi *et al.*, 1998). However, the use of demographic data for segmentation has been criticised by several researchers. For instance, Wind (1978) questioned the 'actionability' of demographic variables. He doubted the capability of demographic variables in translating into application-oriented strategies for practitioners, for example, segmentation based on age or gender variables

does not suggest how the marketing strategies should be formulated. Crawford-Welch (1991) also criticised the use of demographics and noted that "descriptive data, by their very nature, are of little analytical worth in that they are not capable of implying causality and are, in turn, poor predictors of behaviour" (p.301). As a consequence, the use of other descriptors has been advocated.

In recent years, segmentation based on customer service needs or benefits sought has emerged as a more effective approach to segmentation. The underlying assumption is that with fast-changing consumption patterns, it is a great challenge to keep up with the expectations of customers. To be cost-effective, one way is to prioritise services to meet the needs of core groups of clients. Sharma and Lambert (1994) also indicated that one of the areas where additional research is required is the segmentation of markets based on customer service needs (Chen., 2003). Similarly, Loker and Perdue (1992) argued that the benefits which people are seeking in consuming a product or service are the basic reasons for the existence of true market segments. It has also been reported that benefits predict behaviour better than demographic and geographic segmentation because, while other segmentation variables are descriptive, benefit segmentation is causal (Chung *et al.*, 2004; Haley, 1985).

In terms of the public leisure service, Guest and Taylor (1999) argued that the customer is the ultimate judge of the adequacy of public sector leisure provision, so it is important to ensure the preferences of customers are considered while determining the provision of services. In her research about customer expectations of sport organizations, Robinson (2006) also noticed that it is necessary to understand the level of expectations that customers have, as it is generally accepted that customers use their expectations of a service to determine whether the service received is of an acceptable level of quality.

Considering the advantages of needs/benefits segmentation and knowledge of what are important to customers allows service priorities to be identified. The importance of various attributes related to the service provided by public sport facilities is selected as the segmentation variable in this paper. It is important to point out that relevant studies have been undertaken using either customer expectations (e.g. Díaz-Martín *et al.*, 2000; Pitt *et al.*, 1996; Webster, 1989) or the importance of service quality attributes (e.g. Cha *et al.*, 1995; McDougall and Levesque, 1994; Sharma and Lambert, 1994) as bases for needs/benefits segmentation. However, some researchers argued that 'expectation' and 'importance' are distinct concepts (e.g. Woo, 1998; Oh and Parks, 1997), while others used these two terms interchangeably. Although recognising this debate, the database this research relying on only allows the author to use customers' perceptions of the importance of service attributes as the segmentation variable.

4.4 Concluding Remarks

In order to develop customer oriented services and maintain competitiveness, the facility managers need not only to improve the service quality from the supply side but also to clarify customer service needs from the demand side. This chapter reviewed the literature related to customer service quality and customer segmentation. The next chapter will discuss the approach adopted for the analysis of these two dimensions, i.e. multivariate statistical analysis (MSA). Then, the analyses at the industrial and individual facility levels will be presented in Chapter 7 and Chapter 8, respectively.

CHAPTER 5

Methodology

Road map of the thesis

Introduction	Theoretical context	Methodology	Results	Conclusion
Ch1 ▶	Ch2 Performance measurement overall Ch3 Efficiency & Equity Ch4 Customer service	▶ Ch5 ▶	Ch6 Efficiency & Equity Ch7 Customer service Ch8 Action research	▶ Ch9

The way in which research is conducted may be conceived of the research philosophy subscribed to, the research strategy employed and the research instruments utilised in the pursuit of the research objectives. The research objectives have been outlined in Chapter 1. The purpose of this chapter is to:

- discuss the research philosophy;
- expound the research strategy; and
- introduce the research instruments utilised.

5.1 Research Philosophy

A research philosophy is a belief about the way in which data about a phenomenon should be gathered, analysed and used. Two major research

philosophies have been identified in the Western tradition of science: positivist and interpretivist (Galliers, 1991). Positivists believe that reality is stable and can be observed and described from an objective viewpoint (Levin, 1988), i.e. without interfering with the phenomena being studied. They contend that phenomena should be isolated and that observations should be repeatable. Interpretivists contend that only through the subjective interpretation of and intervention in reality can that reality be fully understood. The study of phenomena in their natural environment is the key of interpretivist philosophy, together with the acknowledgement that scientists cannot avoid affecting those phenomena they study.

Based on the two philosophies, research methodologies can be broadly classified into two distinct approaches: quantitative (generally fit in with the positivist stance) and qualitative (generally fit in with the interpretivist stance) research methods respectively (Burns, 1997). A comparison between these two research methods on five dimensions is given in Table 5.1 (Bryman, 1989).

It is believed that no single research methodology is intrinsically better than any other methodology. All methods are valuable if used appropriately and research can include elements of both the positivist and interpretivist approaches, if managed carefully. The concern here is that the research undertaken should be both relevant to the research objectives/questions set out in Chapter 1 and rigorous in its operationalisation. Overall it is believed that both quantitative and qualitative methods are required for achieving the research objectives. The rationales are as follows:

Table 5.1 Comparison between qualitative and quantitative research

Dimensions	Quantitative research	Qualitative research
Aim of the research	Mostly used to test hypotheses; sometimes may also used for exploratory study	Mostly used to discover patterns and linkages of theoretical importance
Approach to collecting data	Model based	Unstructured
Sources of data	Questionnaires are most commonly used or quantitative measurement such as counting are also used	Interview transcripts, field notes etc.
Linkages among parameters	Only static relationships can be identified	Dynamic relationships can be explored
Researcher's role	The researcher may have virtually no involvement in the organisation	The research adopts an stance of insider of the organisation

Source: Bryman (1989)

• ***At the industrial level.*** Quantitative methods normally involve a large randomised sample and the merit is that the results are generalisable. The first objective of this thesis is to provide an insight into the industry's overall performance by conducting a series of secondary analyses of the NBS data. The adoption of quantitative methods allows the author to make generalisable and objective statements on the phenomenon observed. Also, statistical significance can be sought in terms of the variations among observations (e.g. performance differences between facilities)

• *At the individual facility level.* While quantitative methods are an outstanding medium for gathering a breadth of information, qualitative research is the best research method for discovering underlying motivations, attitudes, and perceptions (Burns, 1997). The second research objective is to evaluate the practicability of aggregate performance analysis at the individual facility level, and more depth, rather than breadth, of information is sought. The adoption of qualitative research can help the author to explore, in depth, the attitudes and opinions of facility managers regarding the practicability of aggregate performance analysis.

After introducing the underlying philosophies of this research, the research strategies and instruments utilised in the pursuit of the two research objectives will be discussed in the following two sections.

5.2 Quantitative Methods

At the industrial level, two quantitative methods – Data Envelopment Analysis (DEA) and Multivariate Statistical Analysis (MSA) are used to analyse the performance in four dimensions: operational efficiency, sport equity, service quality and customer segmentation. The following section will first examine the general features of the research data, and then, the research models measuring the above four dimensions will be specified and justified.

5.2.1 Research Data

Research data are derived from the 2001 NBS database, where:

- For operational efficiency, financial information was generated by a survey of 105 centres;
- For sport equity, the information about user profiles was generated by user surveys at 100 centres during 1999-2001. The catchment population profiles, using National Census Data, were provided by Planning Data Management Services of Edinburgh University. A standard catchment area was used for each centre, defined as a 15-minute drive time (Robinson and Taylor, 2003); and
- For service quality and customer segmentation, data from the user surveys from the same 100 centres were used, although only 72 centres which provided complete customer ratings on the 25 service attributes were included in the research sample.

A summary description of the research data and the methods adopted are shown in Table 5.2. Public sport facilities, for the purposes of this study, are defined as swimming pools, sports halls or multi-purpose facilities that combine both 'wet' and 'dry' activities. As mentioned in Chapter 2, In order to prevent comparison against dissimilar organisations or against organisations with dissimilar customer profiles, the results of NBS are structured by the following four 'families' of centres representing major influences on performance.

Table 5.2 Research data and methods

Dimension	Method	Observation [a]	Variable
Operational Efficiency	DEA [b]	105 centres	Inputs: cost, opening hours and facility area Outputs: income and visits
Sport Equity	DEA	100 centres	5 access PIs: social class DE, 11-19 years, ethnic minority, 60+ years and disabled <60 years
Service Quality	MSA [c]	72 centres	Satisfaction and importance ratings on 24 service attributes
Customer Segmentation	MSA	23,329 customers	Importance ratings on 25 service attributes

Notes: [a] Data are all derived from the 2001 NBS database.

[b] DEA= Data Envelopment Analysis.

[c] MSA= Multivariate Statistical Analysis.

• Facility type: a three-way categorisation was chosen: dry (with and without outdoor facilities); mixed, i.e. wet and dry (with and without outdoor facilities) and wet.

• Facility location: a three-way categorisation was chosen. Percentage of centre catchment area population in DE social classes: <15%; 15% to <20% and 20%+. These thresholds are pragmatic to ensure an even distribution of the centres across the three categories. Social classes DE were felt to be representative of an array of social exclusion characteristics in a socioeconomic classification in the UK, with D indicated semi-skilled manual employment and E being unskilled manual employment.

• Facility size: a three-way categorisation of the internal floor space of centres was chosen: < 1500 square meters, 1500 to < 3000 square meters and 3000+ square meters. Again, pragmatic thresholds ensure even distribution of the centres across the three categorisations.

• Management type: a four-way categorisation was chosen: in-house/direct service organisation (DSO); commercial contracted; trust/Non-profit Distributing Organisation (NPDO) and other (e.g. school-based/joint-use).

Finally, the analyses are conducted by adopting the following software: DEA Excel Solver developed by Zhu (2002), Efficiency Measurement System (EMS) developed by Scheel (2000) and Statistical Package for the Social Sciences (SPSS).

5.2.2 Operational Efficiency

The basic concepts of DEA have been discussed in Chapter 3. Here, the focus in on the specific DEA models utilised.

In running DEA, there are four basic model options: input-oriented or output-oriented and constant returns-to-scale (CRS) proposed by Charnes *et al.* (1978) or variable returns-to-scale (VRS) proposed by Banker *et al.* (1984). First, the input-oriented model (also known as input minimisation) examines the extent to which inputs can be reduced while maintaining output levels. Alternatively, the output-oriented model (also known as output maximisation) investigates the extent outputs can be raised given current input levels (Coelli *et al.*, 2005). According to De Borger and Kerstens (1996), the choice of orientation should be inspired by the postulated underlying behavioural mode. If one assumes that local authorities take outputs as exogenous and have substantial control over inputs, then an

input-oriented measure seems appropriate. Input measures can then detect failures to minimise costs resulting from discretionary power and incomplete monitoring, and provide an indication of possible cost reductions. If on the other hand local authorities have limited control over inputs and face fixed budgets, then an output-oriented approach may be quite informative. Output measurement can then identify local authorities that fail to maximise the quantity of the local public services subject to the budget they face, and provide indications of the increase in outputs that could potentially be realised (Worthington and Dollery, 2000). In the following analysis, the output-oriented model was chosen since it is argued that the priority of public service provision is to maximise the benefits to the local community based on given resources, e.g. maximising the visits in the catchment area.

Next, the analyst is often concerned with the nature of returns to scale that would best reflect the operations of the DMUs in the sample. Even in a homogeneous sample some facilities may be operating at CRS while others would be operating at VRS. CRS implies a proportionate rise in outputs when inputs are increased. That is, the scale of operations does not influence the efficiency of the unit. Conversely, VRS implies a disproportionate rise or fall in outputs when inputs are increased. That is, as a unit grows in size, its efficiency would either fall or rise (Avkiran, 1999). Charnes, Cooper and Rhodes (1978) first proposed a model assuming CRS (named as CCR model). The CCR model is appropriate when all organisations are operating at an optimal scale. However, imperfect comparison, government regulations, constraints of finance etc., may cause an organisation to be not operating at optimal scale, namely the inefficiency may derive from scale

inefficiency. Banker, Charnes and Cooper (1984) thus suggested adjusting the CRS DEA model to account for VRS situations (named as BCC model). The use of CCR model results in measures of 'technical efficiency' (also known as gross efficiency) which comprises 'pure technical efficiency' and 'scale efficiency'. Pure technical efficiency describes the efficiency in converting inputs to outputs, while scale efficiency recognises that economy of scale cannot be attained at all scales of production, and that there is one most productive scale size, where the scale efficiency is maximum at 100%. The BCC model takes into consideration of the variation of efficiency with respect to scale of operation, and hence measures pure technical efficiency. Then, the measure of scale efficiency, caused purely by the fact one unit is not operating at the most productive scale size, can be calculated by dividing the measure yielded from CCR model (i.e. technical efficiency) by the measure yielded from BCC model (i.e. pure technical efficiency) (Coelli *et al.*, 2005; Ramanathan, 2003). Since considering CCR and BCC models can distinguish two different kinds of efficiencies and have an insight into the source of inefficiencies, both of them are run in this study.

In addition, standard DEA models assume that all inputs and outputs are discretionary, i.e. controlled by the management of each DMU and varied at its discretion. Thus, 'non-discretionary' variables, which are beyond the control of the facility managers but are likely to affect the performance of facilities, need to be excluded or treated as normal discretionary factors. It may lead to a biased view of efficiency (Syrjänen, 2004). A number of different approaches have been developed to overcome this weakness. The first approach to account for differences in non-discretionary factors was introduced by Banker and Morey (1986). Currently,

this approach can be considered a standard approach for the inclusion of non-discretionary variables and has been the most widely used in the literature (Syrjänen, 2004). In this research, one input item in the efficiency measurement model, i.e. facility area, should be treated as a non-discretionary variable.

Apart from DEA, this part also applies statistical tests, e.g. ANOVA, *t*-test and chi-squared test in order to examine whether performance differences exist between facilities with different type, location, size and management type. Furthermore, according to Sueyoshi and Aoki (2001) and Brockett and Golany (1996), DEA is a non-parametric method, assuming the distributions of effectiveness scores yielded are not normal, so non-parametric Kruskal-Wallis test and Mann-Whitney U test are used to examine the differences between sub-samples.

5.2.3 Sport Equity

The aim of this part of analysis is to aggregate five NBS equity PIs (i.e. social class DE, 11-19 years, ethnic minority, 60+ years and disabled <60 years) into a compositive index representing the overall performance of each facility. The motivation derives from the concern that, while assessing the overall performance of a sports centre, one PI may be high, but another PI may be very low. Therefore, it is difficult to conclude on the overall performance of that centre. One way is to assign weights on each PI, but the major problem is how these PIs should be weighted. An arbitrary attribution of weight usually ends up as a victim of subjectivity (Chang *et al.*, 1995). As argued by previous researchers, DEA has a potential to overcome the abovementioned problem.

In the DEA effectiveness model, the focus is on outputs in the form of five sport equity PIs and inputs are ignored by assuming that they are considered equal at level 1 for all the facilities. Lovell and Pastor (1997) have proved mathematically that an output-oriented DEA model with a single constant input is equivalent to an output-oriented model without inputs. Moreover, considering a single constant input is equivalent to considering multiple constant inputs. Considering the problem from a mathematical point of view (as shown in Equation 5.1), the N facilities under consideration produce a vector of outputs R_i in the form of the mentioned previously identified PIs. The matrix of outputs R_i (with i= 1, 2..., m) is known for each facility n (with n= 1, 2..., N). The n variables to be determined are a set of weights λ ($\lambda = \lambda_1, \lambda_2..., \lambda_K$) placed on each of the facilities in forming the frontier for facility $_0$ and an effectiveness measure Z_0. The effectiveness score for each facility is given by $1 / Z_0$, and it is positive and less than or equal to 1.

Equation 5.1 Mathematical form of DEA effectiveness model

$$\text{Max } Z_0$$

$$\text{s.t. } \sum_{n=1}^{N} \lambda_n R_{in} \geq z_0 R_{i0}; \quad i = 1, 2, ..., m$$

$$\sum_{n=1}^{N} \lambda_n = 1$$

$$Z_0 \geq 0; \lambda_n \geq 0; \quad n = 1, 2, ..., N$$

5.2.4 Customer Service Quality

Multivariate statistical analysis (MSA) (or multivariate statistics) describes a collection of procedures which involve observation and analysis of more than one

statistical variable at a time. Sometimes a distinction is made between univariate (e.g., ANOVA, t-tests) and multivariate statistics, where univariate statistics only have one dependent variable, whereas multivariate statistics have two or more dependent variables. In this research, for the purpose of aggregate performance analysis in customer service, MSA helps the researcher to summarise data and reduce the number of variables necessary to describe it. The characteristics of the specific MSA techniques adopted in this research can be found in Appendix C. Here, the attention focuses on how these techniques are used to address the research questions.

The design of the NBS provides a broad range of attributes covering various service dimensions. In the first stage, factor analysis is used to extract factors from the 24 NBS attributes, which allowed the author to condense a large number of variables to a fewer number of factors representing the main service quality dimensions of public sport facilities. Another aggregation technique, cluster analysis, is then used to group homogeneous facilities with similar service quality levels so as to compare performance differences between clusters. Also, the results of the factor analysis and cluster analysis are cross-tabulated with the four NBS families. Simultaneously, *t*-tests, analysis of variance (ANOVA), multivariate analysis of variance (MANOVA) and chi-squared tests are applied to test whether significant differences could be found between different family categories. In the second stage two prediction techniques, stepwise discriminant analysis and stepwise regression analysis, are performed to identify the most important service dimension(s) in classifying different levels of overall service quality and shaping customers' perception of overall value for money.

In terms of the service quality measurement models, four commonly used models have been presented in Chapter 4. To provide an insight into the industry's performance in customer service quality as well as examine whether alternative models yield consistent results, three models are adopted in this part of analysis (i.e. gap analysis, performance-only analysis and weighted gap analysis). Importance-performance analysis is not included because, as a relative measurement approach, it cannot yield a single value representing a centre's performance and therefore cannot be used for statistical tests.

The mean values of performance and importance scored by the respondents from one centre are used to represent the performance of that centre, and the performances of 72 centres are compared on the basis of four NBS families. The service quality scores of each model are calculated using the following three formulas, where Q, S and I denote service quality, satisfaction and importance respectively:

- Gap analysis: $Q= S-I$
- Performance-only analysis: $Q= S$
- Weighted gap analysis: $Q= (S-I) \times I$

5.2.5 Customer Segmentation

According to Dolnicar (2002), a predominant data-driven segmentation approach is the combinational use of factor analysis for data reduction, cluster analysis for classification, analysis of variance and discriminant analysis for cross-method validations of the classification results, descriptive profiling of each cluster, and other subsequent analyses depending on the study objectives. The

analysis of this research, basically following this analytic procedure, consists of the following four steps.

The 24 importance statements are firstly grouped using factor analysis to find the underlying constructs associated with the perceived importance. The rationale to run a factor analysis is to reduce 'noise' in the data and therefore enhance the ability to classify cases into distinct groups (Johnson and Wichern, 1998). The factor-mean scores from the factor analysis are then used to group the respondents using cluster analysis. The goal of cluster analysis is to arrive at clusters of homogeneous people which differ in meaningful ways and display small within-cluster variation, but large between-cluster variation. Given the sample size, more than twenty thousand, a non-hierarchical (K-means) clustering method is performed because it can be efficiently applied to larger data sets (n> 200) compared to the hierarchical technique (Johnson and Wichern, 1998). It is worth mentioning that the use of factor-cluster technique was questioned by some researchers (e.g. Green and Krieger, 1995) in that preliminary spatial reduction of the original variables can discard relevant information and distort the true cluster structure of the data. However, this approach is still widely adopted in marketing research because it allows data reduction and eases interpretation. To provide a global view of the underlying constructs of various attributes, this study follows the factor-cluster approach.

After emergence of segments, the generality and dissimilarities of the segments could be further accentuated by using statistical analyses. Descriptive analyses are used to portray segments' characteristics. Using mean analysis to rank the importance of segmentation attributes allows assigning a name for each segment that could be used for further segment analyses. Moreover, inferential statistics could be

employed to identify the demographic and participation variations. Chi-squared tests are employed to investigate if there are any significant differences between/among the segments in relation to the attributes under investigation. By consolidating the results from descriptive and inferential analyses, it is able to furnish relevant marketing and managerial implications (Chen, 2003). Finally, discriminant analysis is performed on the four clusters in an effort to identify which service dimensions best discriminate among the four clusters as well as assess the classification rate of the derived segments.

5.3 Qualitative Methods

5.3.1 Research Strategy

Action research is selected as the research strategy to evaluate the practicability of aggregate performance analysis at the individual facility level. The underlying argument is that the role of the researcher is to conduct the transfer of innovative performance measurement techniques, insofar as the technique was not previously installed in some of the organisations studied. This thus requires that the researcher plays a part in the implementation process. To achieve this research objective without being involved would be impossible. Action research allows the researcher to actively participate in some form of change in a system. In this way a change can be triggered by the researcher and then the outcome of that change examined (Greenwood and Levin, 1998). Also, adopting action research, the researcher can not only provide assistance to the organisation in improving benchmarking activities

but also contribute to knowledge in the practicability of aggregate performance analysis.

• ***Definition of action research***

There are numerous definitions of action research. One of the most widely cited is that of Rapoport (1970, p. 499) who defines action research as: "Action research aims to contribute both to the practical concerns of people in an immediate problematic situation and to the goals of social science by joint collaboration within a mutually acceptable ethical framework". This definition draws attention to the collaborative aspect of action research. The dual aim of being both a mechanism for practical problem solving and for generating and testing theory provides a win-win scenario for both researcher and participants, which is the major characteristic of an action research study (Elden and Chisholm, 1993).

Action research can be seen as a variant of case study research, but whereas a case study researcher is an independent observer, an action researcher is a participant in the implementation of a system, but simultaneously wants to evaluate a certain intervention technique. The action researcher is not an independent observer, but becomes a participant, and the process of change becomes the subject of research (Baskerville and Wood-Harper, 1996). Stringer (1996) also defines the role of the action researcher as a catalyst to stimulate people to change.

Action research can be contrasted with positivist science. The aim of positivist science is the creation of universal knowledge or covering law, while action research focuses on knowledge in action. Accordingly, the knowledge created in positivist science is universal while that created through action research is particular,

situational and out of praxis (Susman and Evered, 1978). Even though action research is committed to the production of new knowledge through the seeking of solutions or improvements to real-life practical problem situations, it is more than just another approach to problem solving, for the action researcher is working from within a conceptual framework (Stringer, 1996). According to Checkland, 1991, 'knowledge' created by action research is seen to be embedded within cycles of action and reflection.

• ***Action research cycles***

In action research, action and research are combined into a structured process usually referred to as the action research cycle (Carr and Kemmis, 1983). The cyclic process starts with the recognition of the problem, then plans the action, proceeds to carrying this out and finally evaluates the results obtained (Suojanen, 2001). According to Susman and Evered (1978), this process comprises the following five stages:

- '*Diagnose*' involves the identification by the researcher of an improvement opportunity at a prospective client organisation that is likely to lead to the development of relevant knowledge. In this study, literature review is the first stage of action research, which aims to figure out the deficiencies of traditional performance measurement approach.

- '*Action planning*' involves the joint development and consideration of alternative courses of action to attain the improvement identified and knowledge development. In this study, the purpose of this stage is to analyse the performance data and design an individual report for each facility. The information about overall

performance, a centre's strengths and weaknesses and suitable benchmarking partners were provided.

- '*Action taking*' involves the selection and implementation of one of the courses of action considered in the previous stage. In this research, this stage aims to communicate the key findings to management is stressed. It is done by having a workshop to explain how the report should be interpreted.

- '*Evaluation*' involves the study of the outcomes of the selected course of action. In this research, focus group discussions was held after the workshop, which allow the researcher to investigate the suitability and acceptability of proposed model.

- '*Reflection*' (also known as specifying learning) involves assessing the outcomes of the evaluating stage and, based on this assessment, knowledge generation in the form of a conceptual or theoretical model describing the situation under study. Reflection leads to a revised plan followed by a new action research circle. In this research, this stage focused on the practicability of each element in the individual facility report so as to generate new knowledge of aggregate performance analysis.

5.3.2 Research Design

• ***Research orientation***

Action research focuses on addressing a situation where problems exist. Sometimes the action researcher may discover the problems, but in other situations the problems discover the action researcher (Root-Bernstein, 1989). The former case is 'research-driven' initiation, in that the action researcher might be in possession of a general theoretical approach to addressing problem situations and looking for settings that are characterised by such problems. In this situation, the practitioners

may be somewhat dubious or indifferent, particularly if they are unaware that they are in fact confronting serious problems. The latter case is 'problem-driven' initiation, in that practitioners might be confronted by a seemingly insurmountable problem and seeking help from theoretical specialists. In this situation, the researchers may have to undertake a series of research projects that have a broad theoretical span.

Focusing on the practical application of scientific theories, this research is characterised as 'research-driven' action research, since the researcher has particular objectives and research questions which he wishes to pursue, i.e. investigating to what extent the client organisations can adopt or adapt to the proposed performance measurement model. The key issues and theoretical frameworks are identified by referring to the relevant literature in the first step. From there, the researcher plans and designs a research project with the express purpose of finding answers to research questions.

• ***Action research model***

The action research cycle can be passed through once in an action research study and repeated in the same context until satisfactory outcomes have been achieved or a similar process can be applied at a number of different sites. Baskerville and Wood-Harper (1998) define these different action forms as the following three process models.

- ***Iterative process model***: Iterative action research uses iteration as its primary organising principle, and the entire set of research activities is repeated until the practical problem is resolved.

- ***Linear process model***: This model is characterised by an overall process that the activities are not programmed to be repeated until a result becomes settled, but a single sequence of activities.

- ***Reflective process model***: This model is necessarily iterative, but focuses less on problem diagnosis than on reflective analysis of theory-in-use versus espoused-theory. An espoused theory is one that an individual claims to be following. A theory-in-use is one that can be inferred from action (Argyris *et al.*, 1985). Diagnosis and iteration are implied by this kind of research for an understanding of the distinction between theory-in-use and espoused-theory, but iteration is no longer an end in itself.

Considering the action research is 'research-driven', the model adopted is characterised as 'reflective process model', which is based on a reflective analysis of the relationship between 'espoused-theory' (e.g. aggregate performance analysis) and 'theory-in-use' (e.g. partial measures).

• ***Multiple-case design***

Based on the reflective process model, a multiple-case design is also adopted in order to expand research scope and enhance the rigour of the research. The reasons are twofold. First, this research aims at developing a general aggregate performance measurement model. A degree of generality can be achieved by having several applications and the risk of misjudging of a single event can be overcome (Voss *et al.*, 2002). Second, cross-case analysis can provide the opportunity to search for cross-case patterns (Eisenhardt, 1989; West and Oldfather, 1995), such as the 'typical' barriers centres encounter when they use the proposed model, and thus

increase the robustness of analysis. Rather than being based on a replication logic, the main reason for choosing this design is to enable learning as much as possible about the phenomena under study.

In the context of this research, perhaps the most critical aspect is the fact that it provides a limited basis for the traditional 'direct scientific generalisation' (Yin, 1994; Remenyi *et al.*, 1998). However, while case study data cannot be generalised to populations or universes (i.e. statistically generalised) it can be generalised to theoretical propositions (i.e. an analytical generalisation). Thus, the aim of the case studies was not to infer global findings from a sample to a population, but rather to understand and articulate patterns and linkages of theoretical importance (Remenyi *et al.*, 1998). Furthermore, another obvious defect of multiple-case action research design is it inevitably suffers from the number of the variables that change from case to case, and the difficulty of interpretation that this presents.

5.3.3 Methodological Approach

It was decided to adopt a qualitative approach rather than use a formal questionnaire. There are three reasons. First, in general, qualitative researchers are more interpretive and subjective in their approach (Denzin and Lincoln, 1994). Based on an interpretivist stance, the use of qualitative methods can help the researcher to obtain in-depth understanding of the perceptions of managers towards the innovative benchmarking technique. Second, the complexity of the context studied here, such as different benchmarking motivations among the client organisations, would yield a number of variables that would make use of inferential statistics unmanageable. Third, the limited number of participants in the action

research projects also represents a sampling space that is relatively small for use of quantitative analysis.

• ***Sampling***

Sampling (i.e. the selection of cases) inevitably involves discretion and judgement. The aim of this research is to investigate the general acceptability of proposed benchmarking technique in typical sport facilities. Centres with different management types (i.e. facilities operated directly by local authority, contracted out to private sector and managed by trust) were deliberately chosen. The author states no claim that the centres selected are necessarily a representative sample, but it enables the introduction of diversity into the sample. This selection can be seen as a 'typical case' sampling strategy (Sekaran, 1992). The second criterion is to select centres that have recently received a conventional NBS report so that the managers are still familiar with the benchmarking results and hence provide a reasonable comparison between the two performance measurement approaches. This is consistent with 'judgement sampling' (Sekaran, 1992).

In terms of the number of cases, the decision is restricted mainly by the time and resources available. Eisenhardt (1989) suggests that between 3-10 cases in a study are sufficient for generating data that can be handled by the researcher without getting into problems of information overload. Given the requirement of action research to implement and revise the model developed, the action research is conducted in three centres, which we refer to as centres A (trust), B (in-house) and C (commercial contracted).

• ***Data collection***

A combination of focus group discussion and semi-structured interviews was conducted in this research. In the beginning of intervention, focus group discussion was held after the workshop which introduced the model in the three centres. At the end of the process, semi-structured interviews were conducted with the management teams to evaluate the model's practicability. The rationale of this choice is demonstrated below.

- ***Focus group discussion*** is an informal assembly of targeted individuals who are willing to provide data of a qualitative nature in a focused discussion. The goal of a focus group is to obtain perceptions, feelings, attitudes, and ideas about a selected topic (Krueger, 1998; Vaughn *et al.*, 1996). The focus group method offers several advantages which help to justify the selection. First, it is compatible with the basic assumptions of this research, i.e. interpretivist paradigm (Vaughn *et al.*, 1996). Second, it has been found helpful in assessing needs, developing plans, testing new ideas, or improving existing programs (Morgan and Krueger, 1998). Third, it can enable the researcher not only to learn or confirm the facts, but also the meaning behind the facts (Fern, 2001). Fourth, it capitalises on group interaction and communication to generate data. The synergistic effects of focused discussions can help researchers explore and clarify views in ways that are not possible or accessible in one to one interviews (Kitzinger, 1998). Finally, it can yield a great deal of specific information on a selected topic in a relatively short period of time (Vaughn *et al.*, 1996). However, the focus group method may not be a reliable technique for determining an individual's authentic point of view. The noisy social environment of

focus groups makes it an inappropriate setting in which to assess an individual's knowledge of content (Krueger and Casey, 2000).

- ***Semi-structured interviews*** were implemented, where the investigator was guided by an interview schedule but had the freedom to pursue interesting avenues that emerge and follow the respondent's interests or concerns. The advantages of the semi-structured interview are that it facilitates rapport, allows a greater flexibility of coverage and the exploration of novel areas and, as a consequence of all these factors, has the potential to produce rich data (Smith, 1995). The interview questions (as shown in Appendix D) were tailored to elicit the main reasons for continuing to implement the model and the main reasons for not continuing. Successful centres were to be asked to identify problems they encountered but overcame. In the interviews undertaken with the centres that did not proceed to implement their performance measures, detailed probing of the responses lead to greater insights into why the initiative fails.

Focus group discussions and interviews were conducted with directors/ managers who are in charge of making the strategic and operational decisions to gain a senior management perspective on how model should be improved and was used. Tape recording were used to enhance the accuracy of data collected and to improve the efficiency of data analysis (Creswell, 1994; Patton, 1990).

5.3.3 Research Quality

This section discusses the question of which criteria that should be used for evaluating the quality of this research. The standard criteria used for judging the quality of research are validity and reliability. Action research is a fairly close

relative of case research, and raises similar questions of methodology. A criticism of the case study method is that it suffers from a lack of rigour and an excess of bias. According to Westbrook (1995), action research case study raises an additional methodological problem due to the intervention of action researcher which results in increased researcher bias. To maximise the rigour of the research method employed, a number of safeguards were used and discussed, with the aim of increasing the validity and reliability.

• ***Criticisms of action research***

In spite of the advantages that this combination of action and research can bring about, the use of action research in organisational research has some inherent weaknesses. According to Kock (2003), there are three main criticisms of action research by positivists: 'uncontrollability', 'contingency', and 'subjectivity'. These criticisms seem to be associated with the 'emergent' nature of most action research investigations, where a theoretical model emerges from the research data rather than being defined a priori and tested against that data.

- ***Uncontrollability***: This lack of control is one of the main reasons for action research being seen as inappropriate to test or produce strong theories, or build up research models based on solid evidence. The influence of a particular variable might take too long to be isolated in action research studies testing or refining a causal model (Jonsonn, 1991).

- ***Contingency***: To say that research findings are highly contingent is the same as saying that the findings have a low external validity. The threat to external validity in action research is because most action research projects involve a small

number of client organisations in in-depth and often longitudinal studies, and very seldom generalisability assessments across a number of organisations.

- ***Subjectivity***: This criticism hinges on the fact that, in action research, the personal involvement of the researcher is likely to push him or her into interpreting the research data in particular and potentially subjective ways, and, as a result, some of these interpretations may end up being completely wrong (Francis, 1991).

• ***Quality in this research***

The aim here is to demonstrate some methods used in this research to meet the criteria of quality interpretive research and overcome the threats of action research. The related tactics suggested here are highly related to the research design and can be seen as constituting a basis for the data collection in this study. According to Kock (2003), one of the reasons why action research is seen as preferably carried out in cycles is the opportunity that it allows for strengthening research findings by building on evidence gathered from previous iterations in the action research cycle. One way to make action research more rigorous is to expand the research scope and building up the generality of the results through the identification of invariable patterns. According to Kock (2003), multiple cases and successively reflective process model can help to enhance the validity and reliability of this research by the following three ways.

- *Uncontrollability*: The first contribution of multiple iterations research design is to counteract the uncontrollability threat of single site action research, where problems can occur that are outside the sphere of control of the researcher and that can undermine the data collection and analysis.

- *Contingency*: Multiple iterations can counter the negative effects of the contingency threat by allowing for the observation of invariable patterns in different contexts, for which the research scope can be expanded and the degree of transferability can be claimed.

- *Subjectivity*: The other benefit of multiple-case design is that disconfirmatory evidence in further iterations may help correct distortions in the findings of previous iterations caused by personal over-involvement.

CHAPTER 6

Results of Efficiency and Equity Measurement

Road map of the thesis

Introduction	Theoretical context	Methodology	Results	Conclusion
Ch1 ▶	Ch2 Performance measurement overall Ch3 Efficiency & Equity Ch4 Customer service	▶ Ch5 ▶	Ch6 Efficiency & Equity Ch7 Customer service Ch8 Action research	▶ Ch9

6.1 Operational Efficiency

The aim of this section is to apply DEA to measure the operational efficiency of English public sport facilities at the industrial level. The first step in applying the DEA model is to identify the set of input and output measures to be included in the analysis. The NBS provides multiple input/output measures. Based on the data used to compute twelve NBS efficiency performance indicators (2 for utilisation and 10 for financial) and a further consultation with the main developer of the NBS, this study was conducted by adopting operating cost, opening hours and facility area as the input variables, and income and visits as the output variables. Research data were derived

from the performance data and benchmarking results of 105 sports halls and swimming pools in England during the year 2001.

6.1.1 Descriptive Statistics

The descriptive statistics of the research variables are shown in Table 6.1. The next step involves a determination of the relationship between inputs and outputs. According to Golany and Roll (1989), for the validation of the developed DEA model, there should be an 'isotonicity' relationship between the input and output factors, i.e. an increase in any input should not result in a decrease in any output. They suggest that a correlation analysis on the selected input and output factors is a useful procedure for the isotonicity test. The correlations between all input and output variables for the efficiency model are demonstrated in the second part of Table 6.1. The result shows that the relationships between inputs and outputs are all positive and statistically significant at the 0.01 level. It provides an assurance that the model using three inputs and two outputs is a good representation of the efficiency measurement.

Furthermore, since one of the research questions is to investigate whether it is the different features of facility that differentiate the efficient from the inefficient ones and whether the strengths and weaknesses differ between facilities with different features, in the following analysis, the whole sample was divided into different subgroups according to the four NBS families, i.e. facility type, location, size and management type. For definitions of the NBS families, please refer to Chapter 5.

Table 6.1 Descriptive statistics and correlation analysis of research variables

	Operating cost (£)	Opening hours	Facility area (m^2)	Operating income (£)	Visits
Mean	657,852	4,864	2,919	493,983	236,566
S.D.	438,900	712	2,139	407,648	159,466
Max	2,183,639	6,407	14,000	2,323,884	817,639
Min	57,322	3,000	561	9,634	17,968
Correlation analysis					
Operating cost	1.00				
Opening hours	0.43 **	1.00			
Facility area	0.60 **	0.33 **	1.00		
Operating income	0.87 **	0.52 **	0.53 **	1.00	
Visits	0.80 **	0.55 **	0.58 **	0.83 **	1.00

Notes: * <0.05; ** <0.01

6.1.2 Results of DEA

The analysis in this section aims to answer the following three sets of questions:

- How many facilities are deemed to be efficient? What is the main source of inefficiency? Does economies of scale exist in the industry?
- Is it the different features of facility that differentiate the efficient from the inefficient ones? Do strengths and weaknesses differ between facilities with different features?
- Is the information yielded from DEA consistent with the NBS results? To what extent can the two approaches complement one another?

We first examine the operational efficiency of the whole sample at the aggregate level. The results of the analysis of technical, pure technical and scale efficiency of 105 facilities are presented in Table 6.2. A facility is defined as efficient if its pure technical efficiency and scale efficiency are both 1, resulting in technical efficiency

being 1. On the basis of observable best-practice, the average technical efficiency score is 0.64. It implies that, compared with the efficient facilities, the inefficient facilities only produced 64% of outputs by using the same level of inputs. Further, in the case of overall technical efficiency there are 9 (or 8.6%) facilities defining the frontier, 17 centres (or 16.2%) are purely technical efficient with a mean value of 0.71 and 10 centres (or 9.5%) are scale efficient with a mean value of 0.91.

Table 6.2 DEA scores: technical, pure technical and scale efficiency

	Technical efficiency	Pure technical efficiency	Scale efficiency
No. of efficient facilities (%)	9 (8.6%)	17 (16.2%)	10 (9.5%)
Mean	0.64	0.71	0.91
S.D.	0.21	0.20	0.14

In order to investigate the source of inefficiency, the frequency distributions of the three efficiency indices are shown in Table 6.3. Similar distributions can be found between technical efficiency and pure technical efficiency, that is, 0.50 ~ <0.75 is the most frequent interval. However, for scale efficiency, 0.75 ~ 1.00 is the most frequent interval and only 9 facilities have an efficiency score less than 0.75. This evidence explains that differences in technical efficiency are mainly driven by differences in pure technical efficiency. More precisely, the main problem of inefficient facilities in the research sample is the inability to use the resources in the technically most efficient way, rather than affected by the scale effect.

Table 6.3 Frequency distribution: technical, pure technical and scale efficiency scores

Interval (DEA scores)	Number of facilities		
	Technical efficiency	Pure technical efficiency	Scale efficiency
0.00 ~ <0.25	2	1	0
0.25 ~ <0.50	28	17	2
0.50 ~ <0.75	41	44	7
0.75 ~ 1.00	34	43	96

The next set of results that were calculated reveals the nature of returns to scale. The analysis shows that a high proportion of facilities (82 facilities, or 78%) exhibit increasing returns to scale (output increases more than in proportion to inputs), whereas only 14 facilities (or 13.3%) exhibit decreasing returns to scale (output increases less than in proportion to inputs) and the remaining 9 facilities (8.6%) exhibit constant returns to scale (output increases just in proportion to inputs). It implies that 78% facilities in the sample benefit from the economies of scale.

As the efficiency score itself does not give much information about what characterises the efficient and inefficient facilities, we try to identify whether it is different features of facility that distinguish the efficient from the inefficient facilities. Based on four NBS families, Table 6.4 first reveals that the average efficiency scores of mixed, DE<15%, 3000+ sq.m. and commercial contracted facilities are higher than the remaining categories in the same families. When segmenting by management type, the differences of average efficiency scores between in-house and other three subgroups are especially evident. Furthermore, the quartiles of each subgroup's efficiency scores are also presented in Table 6.4. DE <15%, 3000+ sq.m. and commercial contracted facilities outperform other subgroups in the same family

categories for all levels of the percentiles. However, the above descriptive statistics reveal only performance and comparison of average values. A more effective conclusion on the determinants of efficiency requires further statistical tests.

Table 6.4 DEA scores of four NBS families and nonparametric test

	n	DEA scores					Kruskal-Wallis (*p*-value)
		Mean	(S.D.)	Percentiles			
				25%	50%	75%	
Type							
Dry	21	0.62	(0.24)	0.41	0.65	0.81	χ^2=3.05
Mixed	47	0.69	(0.19)	0.55	0.63	0.86	(p = 0.22)
Wet	37	0.61	(0.21)	0.44	0.57	0.78	
Location							
DE <15%	15	0.73	(0.21)	0.56	0.73	0.93	χ^2=5.90
DE 15% to 20%	39	0.67	(0.19)	0.55	0.67	0.84	(p = 0.053)
DE 20%+	51	0.60	(0.21)	0.46	0.61	0.77	
Size							
<1500 sq.m.	26	0.62	(0.24)	0.43	0.66	0.80	χ^2=4.18
1500 to <3000 sq.m.	35	0.60	(0.18)	0.49	0.56	0.66	(p = 0.12)
3000+ sq.m.	44	0.69	(0.20)	0.50	0.70	0.86	
Management							
In-house	56	0.56	(0.19)	0.44	0.55	0.67	χ^2=17.73
Commercial contracted	15	0.77	(0.20)	0.63	0.78	0.94	(p = 0.00)
Trust	24	0.73	(0.19)	0.60	0.75	0.86	
Other	10	0.74	(0.14)	0.62	0.76	0.82	

In order to further investigate the distribution of efficiency, the non-parametric Kruskal-Wallis test was used to test the null hypothesis that the facilities with different type, location, size and management type follow the same efficiency

distribution. Non-parametric tests were used since DEA is a non-parametric method assuming the distributions of efficiency scores yielded are not normal (Sueyoshi and Aoki, 2001). The result, shown in the final column of Table 6.4, reveals significant differences in technical efficiencies at the 0.05 level due to management type, where the chi-squaredd test statistic is 17.73 (p=0.00). There are no statistically significant differences due to facility type, location or size at the 0.05 level, although differences by location are very close to being significant at the 0.05 level (p=0.053).

Since management type has a significant effect on operational efficiency and the p value of facility location is close to 0.05, we then conducted further statistical tests to investigate the sources of performance differences. As the sample sizes of certain subgroups are relatively small, we combined commercial contracted, trust and other management type into one subgroup- 'non in-house' as well as DE <15% and DE 15% to 20% into one subgroup- 'DE <20%'. The statistics of in-house and non in-house and DE <20% and DE 20%+ are provided in Table 6.5. Non in-house outperforms in-house and DE <20% outperforms DE 20%+ regardless of percentage of efficient facilities or mean efficiency score. In addition, as proposed by Brockett and Golany (1996), the non-parametric Mann-Whitney U test was adopted to investigate whether two subgroups follow the same efficiency distribution. The Mann-Whitney U test reveals that there are significant differences between two management types and two facility locations at the 0.05 level (the actual p values are 0.00 and 0.02 respectively).

Table 6.5 Statistics of two management types and two facility locations

	n	Efficient facilities		Mean efficiency scores		
		Number	(%)	Mean	(S.D.)	Mann-Whitney U
Management						
Non in-house	49	6	(12.2%)	0.74	(0.18)	Z=-4.56
In-house	56	3	(5.4%)	0.56	(0.19)	(p=0.00)
Location						
DE <20%	54	6	(11.1%)	0.69	(0.20)	Z=-2.31
DE 20%+	51	3	(5.9%)	0.60	(0.21)	(p=0.02)

In an attempt to have an insight into the product and resource structure of facilities with various management types and facility locations, the virtual weights of inputs and outputs were examined in the next step. In terms of the two management types, as shown in Table 6.6, in-house facilities place more emphasis on visits (virtual weight=0.59), whereas non in-house facilities focus more on income (virtual weight=0.71). In-house and non in-house facilities appear to have a similar resource structure, i.e. the importance of cost is higher than opening hours and facility area. However, compared with in-house facilities, a higher proportion of weight was assigned to cost for non in-house facilities. Mann-Whitney U test further implies significant differences in three inputs and two outputs, especially the p values of cost, income and visits are all equal to 0.00. It confirms the evidence provided by the comparison of mean virtual weights, that is, the strength of in-house facilities is visits whilst the strengths of non in-house facilities are income and cost. Also, the differences in input/output structures help to explain the performance gaps between in-house and non in-house facilities.

Table 6.6 Virtual weight analysis: in-house and non in-house

	Mean virtual weights		Mann-Whitney U	
	In-house	Non in-house	(p-value)	
Inputs				
Cost	0.52	0.75	Z= -4.63	(p=0.00)
Opening hours	0.31	0.15	Z= -2.84	(p=0.01)
Facility area	0.17	0.10	Z= -2.52	(p=0.01)
Outputs				
Income	0.41	0.71	Z= -4.20	(p=0.00)
Visits	0.59	0.29	Z= -4.20	(p=0.00)

As to the category of facility location, Table 6.7 shows that both facilities in DE <20% and DE 20%+ have higher mean virtual weights in cost and income. However, the Mann-Whitney U test shows that there are no significant differences in the virtual weights of five input/output variables due to facility location (the actual p values range between 0.16 and 0.52). It implies that even though there are significant differences in efficiency distribution due to facility location, the differences in input/output configuration are not evident enough to explain the source of performance gaps between two subgroups.

Table 6.7 Virtual weight analysis: DE <20% and DE 20%+

	Mean virtual weights		Mann-Whitney U	
	DE <20%	DE 20%+	(p-value)	
Inputs				
Cost	0.66	0.60	Z= -1.24	(p=0.22)
Opening hours	0.19	0.28	Z= -1.41	(p=0.16)
Facility area	0.15	0.13	Z= -1.41	(p=0.16)
Outputs				
Income	0.58	0.52	Z= -0.65	(p=0.52)
Visits	0.42	0.48	Z= -0.65	(p=0.52)

Finally, in order to investigate whether the information yielded from DEA is consistent with the NBS results and whether the two approaches can complement one another, Table 6.8 presents a comparison between different management types and facility locations in terms of the mean values of twelve existent NBS efficiency performance indicators linking different input and output variables. T-tests were also conducted to test the null hypothesis that the performance values of two sub-samples follow the same distribution.

Table 6.8 Mean values and *t*-test of twelve NBS efficiency indicators

	Management Type			***Facility Location***		
	Mean value			Mean value		
	Non in-house	In-house	*t*-value	DE <20%	DE 20%+	*t*-value
Utilisation						
Visits per sq. m.	87	97	0.72	100	85	1.13
Visits per hour	44	49	0.93	46	48	-0.26
Financial						
Income per visit	2.51	1.79	-3.40 **	2.20	2.05	-0.74
Income per sq. m.	199	165	-1.40	203	157	1.93
Income per hour	104	92	-0.84	99	96	0.23
Cost recovery	92	60	-7.63 **	80	69	2.02 *
Cost per visit	2.82	3.24	1.73	2.86	3.24	-1.55
Cost per sq. m.	218	278	2.28 *	259	240	0.72
Cost per hour	111	151	2.52 *	124	141	-1.05
Subsidy per visit	0.31	1.45	6.34 **	0.65	1.19	-2.55 *
Subsidy per sq. m.	19	111	7.32 **	54	83	-1.89
Subsidy per hour	7.5	59.1	7.32 **	25.2	45.4	-2.33 *

Notes: * <0.05; ** <0.01

As shown in the left part of Table 6.8, even though, on first impression, in-house facilities outperform non in-house facilities in two utilisation indicators, there is no significant difference between the two management types. As to ten financial indictors, non in-house facilities all outperform in-house facilities. In addition, t-tests shows that there are significant differences in seven financial indicators, including 'income per visit', 'cost per sq.m.', 'cost per hour', 'cost recovery' and three subsidy related indicators. This finding demonstrates a high consistency between DEA and the NBS results, that is, the strengths of non in-house and in-house facilities are finance and utilisation respectively. The NBS results, i.e. significant differences in specific performance indicators, further demonstrate specific aspects resulting in the differences of overall efficiency between two management types. In a word, in-house facilities outperform non in-house facilities in utilisation, but the difference is not statistically significant; however, the success of non in-house facilities relative to in-house facilities in financial efficiency is more significant.

In terms of facility location, as shown in the right part of Table 6.8, the information provided by ratio analysis is mixed, for instance facilities in DE <20% have a better performance in 'visits per sq.m.' but facilities in DE 20%+ have a better performance in 'visits per hour'. Also, contradictory information is yielded in three cost related indicators. In this case, DEA can complement ratio analysis by demonstrating that facilities in DE<20% have a better overall efficiency relative to those in DE 20%+. The phenomenon that various performance indicators indicate different directions can also be explained by DEA, that is, the differences in input and output virtual weights are not significant. As no specific strength or weakness can be

identified, a sub-sample appearing to perform well in one indicator may perform less successfully while considering another.

6.1.3 Discussion

The virtual weights analysis in Tables 6.6 and 6.7 suggest that the strengths of the industry in efficiency terms lie in cost control on the input side, and both income and visits on the output side, with a slightly greater emphasis on income. This is not surprising, because the other two inputs in the efficiency model presented, opening hours and facility area, are both relatively fixed in the short run.

From an empirical perspective, one of our major findings is evidence that non in-house facilities outperform in-house facilities both in the percentage of efficient facilities and in the average efficiency scores. Furthermore, the success of non in-house facilities relative to in-house facilities in financial efficiency is statistically significant; whilst the apparent stronger performance in utilisation efficiency by in-house facilities compared with non in-house facilities is not statistically significant. This finding can be further complemented by other evidence in order to generate management implications.

In the UK, numerous services previously provided by public agencies at the central and local government level have been competitively tendered and contracted out to the private sector over the past decade. Trusts or Non-profit Distributing Organisations (NPDO) have also emerged as a principal management alternative for local authorities (Worthington and Dollery, 2000; Mintel, 2006). In 2006, there were 4,215 public sport facilities in the UK, with the majority managed in-house (42%), followed by education management (24%), trust management (14%), commercial

contracted management (9%) and other (11%) (Mintel, 2006). Although in-house management accounts for the largest proportion of the whole market, unlike some relatively fixed factors (e.g. facility type, location or size), management type is under the control of local authorities, at least in the medium term. Therefore, an interesting issue worth further discussion is: which management type is more appropriate in the future? The results above, on efficiency considerations alone, suggest that non in-house management will provide better results.

However, the 'which management type?' question requires consideration of not just the operational efficiency but also other important aspects of facility performance. From an investigation of top-quartile performing centres from the NBS data set, Taylor (2004) demonstrated that all management types have the potential to achieve top-quartile performance and there is a considerable crossover between top-quartile access, utilisation and financial performance. More than half the centres with top-quartile financial performance also had top-quartile access performance, and vice versa. There does not appear to be a trade-off between these dimensions to performance. This gives further support to a conclusion based only on efficiency performance.

From a theoretical perspective, the value of DEA in measuring efficiency lies in three areas. First, DEA allows distinguishing overall efficiency and inefficiency, which is unavailable from ratio analysis, especially when conflicting signals exist among various indicators. Second, virtual weight analysis allows clarifying strengths and weaknesses of each centre. In turn, the performance indices provided by the NBS provide an insight into the gaps in specific performance dimensions. Third, the benchmarking partners suggested by the NBS are based on those who perform well in

specific performance indicators; however, DEA can complement the NBS by selecting benchmarking partners which have better all-round performance. Also, the NBS type of benchmarking partners is on the basis of similar operating context; however, DEA can enhance the NBS by providing benchmarking partners with similar strengths and weaknesses.

Finally, the above analysis is based on conventional DEA models; however, it is of course possible that the weighting method and the performance targets emerging from DEA do not correspond to the preferences of decision makers. Some methodological extensions of DEA have been made to solve these problems.

In terms of the weighting method, the conventional DEA model allows weights to vary freely so that each organisation is evaluated in the best possible light. Indeed we frequently find in unconstrained DEA that the highest efficiency score for an organisation can be secured simply by assigning unreasonably low or excessively high values to the multipliers in an attempt to drive the efficiency rating for a particular DMU as high as possible (Allen *et al.*, 1997). An unbalanced set of weights is a signal that, although efficient, the organisation may be achieving that efficiency by employing an unusual mix of inputs and outputs, and may be far from efficient in the allocative sense. It is also problematic because the efficiency score ignores the performance of the unit on the remaining inputs and outputs, perhaps to the detriment of performance in other dimensions. The problem of weighting flexibility has led to the development of methods for using weights restrictions in DEA assessments to ensure that the DEA efficiencies estimated reflect the performance of units on all, and not just a subset of the input and output variables (Thanassoulis *et al.*, 1996).

In terms of performance targets, one of the main purposes of a DEA study is to project the inefficient DMUs on to the production frontier. In addition to input-oriented and output-oriented models, there is a third choice, represented by the Additive Model and Slack-Based Model that deal with the input excesses and output shortfalls simultaneously in a way that maximizes both (Cooper *et al.*, 2000). The Additive model, developed by Charnes *et al.* (1985), can deal with the input excesses and output shortfalls directly and can discriminate between efficient and inefficient DMUs, but it has no means to gauge the depth of inefficiency by a scale measure. To make up this deficiency, the Slack-Based Model can provide a scale measure which reflects non-zero slack in inputs and outputs when they are present (Cooper *et al.*, 2000). In addition, the conventional DEA model can be called 'radial' efficiency measures, that is, the performance targets suggested by DEA is to improve all outputs or inputs simultaneously and proportionally. This assumption becomes invalid when a preference structure over the improvement of different inputs (outputs) is present in evaluating DMUs. Zhu (2002) provides a Non-Radial Model which allows non-proportional inputs reductions or outputs augmentations.

6.2 Sport Equity

Based on the same database (i.e. the 2001 NBS), this section benchmarks the performance of 100 English public sport facilities in sport equity by adopting statistical analysis and DEA. Then, the appropriateness of DEA in benchmarking is critically evaluated. As mentioned in Chapter 5, the five CPA equity PIs (i.e. social

class DE, 11-19 years, ethnic minority, 60+ years and disabled <60 years) are chosen as the research variables

6.2.1 Descriptive Statistics

The descriptive statistics of the 2001 NBS results are shown in Table 6.9. 75%, 50% and 25% percentile scores represent three levels of benchmarks. For the four 'representativeness' PIs, 'ethnic minority' has the highest proportion of well-represented (with a PI ≥1) facilities (63%), followed by '11-19 years' (35%), '60+ years' (2%) and 'social class DE' (0%). 'Ethnic minority' is also the only one PI with a 50% benchmark (i.e. median) higher than 1.0 and a relatively high 75% benchmark (2.5). The whole industry has a mediocre performance in '11-19 years' with a 75% benchmark equal to 1.1. By contrast, 'social class DE' and '60+ years' are significantly under-represented.

Table 6.9 2001 NBS results: descriptive statistics

(*n*=100)	Social class DE	11-19 years	Ethnic minority	60+ years	Disabled < 60 years
Benchmarks					
75%	0.4	1.1	2.5	0.6	7.7
50%	0.3	0.8	1.2	0.4	6.2
25%	0.2	0.5	0.8	0.3	5.0
% of well-rep. [a]	0%	35%	63%	2%	n.a.
Max.	0.81	3.78	9.33	1.10	11.79
Min.	0.00	0.20	0.00	0.00	0.42
Mean	0.33	0.96	1.93	0.46	6.43
CV	0.54	0.67	0.95	0.51	0.32

Notes: [a] 'well-rep.' are defined as those facilities with a PI ≥1 for the representativeness ratios.

To confirm the evidence shown by the NBS benchmarks (i.e. the quartile analysis), the sample variance is examined. Since the 'representativeness' ratios (the first four PIs) and 'share' ratio (the last PI) have different units, the coefficient of variance (CV) is chosen instead of standard deviation. Among the four representativeness ratios, '60+ years' has the lowest CV (0.51), followed by 'social class DE' (0.54). Also, the ranges and the inter-quartile ranges of these two PIs are much smaller than the other two representativeness ratios (i.e. '11-19 years' and 'ethnic minority'). These findings therefore confirm the consistent under-representation of '60+ years' and 'social class DE'. In terms of the share ratio 'disabled <60 years', the 2001 NBS median score of 6.15% compares with 15% of the population with a disability, according to the 2002 General Household Survey. The lowest CV (0.32) among the five PIs further demonstrates the consistent under-representation of 'disabled <60 years'.

The next set of results tries to identify whether performance differences exist between facilities with different profiles. The analysis is based on the four NBS families. As the sample sizes of certain subgroups are relatively small, we combine DE <15% and DE 15% to 20% into one subgroup-'DE <20%' as well as commercial contracted, trust and other management types into one subgroup- 'non in-house' in order to yield more balanced sub-sample comparisons. Both the mean values of PIs and the number of facilities in the top-quartile (≥75% benchmark) are examined through ANOVA/*t*-test and chi-squared test. The results are shown in Table 6.10.

Table 6.10 NBS families: mean values of PIs and % of facilities in the top-quartile

NBS families	*n*	Mean values of PIs [% of facilities in the top-quartile [a]]				
		Social class DE	11-19 years	Ethnic minority	60+ years	Disabled < 60 years
Type						
Dry	21	0.32	1.54 **	1.92	0.28 **	5.72
		[24%]	[62%] **	[24%]	[5%] **	[19%]
Mixed	46	0.32	0.88 **	1.89	0.49 **	6.41
		[28%]	[20%] **	[26%]	[33%] **	[26%]
Wet	33	0.34	0.71 **	1.98	0.53 **	6.91
		[33%]	[9%] **	[30%]	[36%] **	[27%]
Location						
DE <20%	51	0.37 *	0.97	1.80	0.47	5.70 *
		[37%]	[29%]	[26%]	[31%]	[18%] **
DE 20%+	49	0.28 *	0.95	2.06	0.44	7.19 *
		[20%]	[20%]	[29%]	[25%]	[33%] **
Size [b]						
Small	25	0.37	1.40 **	2.37	0.43	6.37
		[36%]	[48%] *	[32%]	[28%]	[20%]
Medium	33	0.33	0.85 **	1.96	0.44	6.40
		[27%]	[18%] *	[33%]	[27%]	[27%]
Large	42	0.30	0.79 **	1.64	0.49	6.49
		[26%]	[17%] *	[19%]	[29%]	[26%]
Management						
In-house	54	0.35	0.97	1.73	0.49	6.71
		[32%]	[24%]	[20%]	[32%]	[26%]
Non in-house	46	0.30	0.95	2.16	0.43	6.10
		[26%]	[26%]	[35%]	[24%]	[24%]

Notes: [a] Number within the [ ] bracket means % of facilities in the top-quartile.

[b] Small = <1,500 sq.m.; Medium = 1,500 to <3,000 sq.m.; Large = 3,000+ sq.m..

* <0.05; ** <0.01. (ANOVA/*t*-test for mean values comparison and Chi-squared test for the number of facilities in the top-quartile).

First, if we only compare the mean values and the percentage of facilities in top-quartile, it seems that no specific kind of facility outweighs other kinds of facility in all PIs. For instance, dry facilities under-perform wet and mixed facilities in 'social class DE', '60+ years' and 'disabled <60 years', but attract more '11-19 years'. Further statistical tests reveal significant performance differences due to facility type in '11-19 years' and '60+ years'; due to facility location in 'disabled <60 years' and in 'social class DE' (only in mean value comparison); and due to facility size in '11-19 years'.

6.1.2 Results of DEA

Table 6.11 shows the DEA effectiveness scores representing the overall performance of the facilities; and also the virtual weights demonstrating the strengths and weaknesses of each sub-sample. Non-parametric tests are conducted to test whether significant differences exist between different facility profiles (Kruskal-Wallis test for 'type' and 'size' and Mann-Whitney U test for 'location' and 'management type'). Even though, on average, wet, DE <20%, small and in-house facilities have higher DEA scores, no statistically significant differences can be found in the DEA scores of each family. This finding can be further complemented by the virtual weights analysis shown in the same table. The DEA assigns higher weights on each centre's strength, in order to maximise its performance score. The net result, examining the weight distributions, is that no family sub-sample has weights higher than those of the other sub-sample(s) in all five PIs. Instead, different family sub-samples have strengths in different PIs.

Table 6.11 DEA results: mean values of DEA scores and virtual weights

NBS families	DEA scores	DEA virtual weights				
		Social class DE	11-19 years	Ethnic minority	60+ years	Disabled < 60 years
Type						
Dry	0.73	25%	28% **	6%	10% *	31%
Mixed	0.74	20%	11% **	8%	27% *	34%
Wet	0.76	19%	7% **	8%	25% *	42%
Location						
DE <20%	0.75	27% **	14%	7%	26%	26% **
DE 20%+	0.73	14% **	13%	8%	19%	46% **
Size						
Small	0.81	24%	22% **	9%	22%	23% *
Medium	0.73	21%	11% **	9%	19%	40% *
Large	0.71	17%	10% **	6%	26%	41% *
Management						
In-house	0.76	21%	14%	6%	25%	35%
Non in-house	0.72	19%	13%	10%	20%	37%

Notes: * <0.05; ** <0.01 (Non-parametric tests: Kruskal-Wallis and Mann-Whitney U).

The statistical tests in Table 6.10 and Table 6.11 generally reflect the same patterns. In Table 6.11, significantly higher weights (i.e. strengths) are assigned to DE <20% in 'social class DE'; dry and small in '11-19 years'; wet/mixed in '60+ years'; and DE 20%+ and medium/large in 'disabled <60 years'. These aspects (except for medium/large in 'disabled <60 years') also have higher mean values and/or more facilities in the top-quartile as shown in Table 6.10. The coincidence between the findings of simple statistical analysis (Table 6.10) and DEA (Table 6.11) raises an important issue. It appears that simply comparing the mean values or examining the percentage of facilities in the top-quartile can clarify the strengths and weaknesses,

and the results are highly consistent with those yielded by DEA. If this is the case, it poses the question 'why should analysts resort to a more complicated model?'

The other function of DEA is to select the best facilities from the sample which are deemed 100% DEA effective (i.e. with a DEA score equal to 1.0). In this research, 11 centres have a performance score with 1.0; they are therefore regarded as the benchmarking partners for other ineffective centres. Their performance values, quartiles related to the 2001 NBS benchmarks, virtual weights and family categorisation are shown in Table 6.12. The number and percentage of effective facilities in each family is shown in the bottom of the tables, which indicates that all types, locations, sizes and management types have potential to become DEA effective.

Even though all of these 11 centres have the same DEA score, their performance in each PI varies a lot. The second column of Table 6.12 summarises the number of PIs in the top-quartile (4^{th}) and the bottom-quartile (1^{st}) of each centre. Some centres have a more balanced performance. For example, the first six centres (from A-01 to A-06) have at least three PIs in the top-quartile and no more than one PI in the bottom-quartile. However, there also exist some extreme examples. For instance, A-09 and A-10 both have two PIs in the bottom-quartile and A-11 only has one PI in the top-quartile, but they are still deemed as the best practices by DEA. Taking A-09 as an example, it has two PIs in the 1^{st} quartile and one in the 2^{nd} quartile, but it has an excellent performance in '60+ years' (1.08), which is quite close to the maximum of the sample (1.10). As such, the three weaknesses are assigned very low weights (0%) and a great proportion of weight is assigned to its strength- '60+ years' (83%).

Table 6.12 DEA results: benchmarking partners

Centre code	No. of PIs in the 4th / 1st quartile	PIs, quartiles and virtual weights					Families			
		Social class DE	11-19 years	Ethnic minority	60+ years	Disabled <60 years	Type	Location	Size	Management
A-01	4 / 0	0.44 : 4th (12 %)	0.62 : 2nd (3 %)	3.66 : 4th (1 %)	0.89 : 4th (22 %)	11.79 : 4th (62 %)	Wet	DE 20%+	Large	In-house
A-02	4 / 1	0.43 : 4th (0 %)	0.44 : 1st (0 %)	6.09 : 4th (47 %)	0.83 : 4th (53 %)	8.43 : 4th (0 %)	Mixed	DE <20%	Large	Non in-house
A-03	3 / 0	0.81 : 4th (66 %)	1.33 : 4th (11 %)	3.78 : 4th (23 %)	0.48 : 3rd (0 %)	6.09 : 2nd (0 %)	Mixed	DE <20%	Medium	Non in-house
A-04	3 / 0	0.80 : 4th (50 %)	0.91 : 3rd (0 %)	4.19 : 4th (17 %)	0.80 : 4th (33 %)	5.70 : 2nd (0 %)	Wet	DE <20%	Small	Non in-house
A-05	3 / 1	0.14 : 1st (0 %)	3.78 : 4th (64 %)	2.74 : 4th (9 %)	0.59 : 3rd (27 %)	8.23 : 4th (0 %)	Dry	DE 20%+	Small	Non in-house
A-06	3 / 1	0.30 : 3rd (0 %)	2.01 : 4th (13 %)	9.33 : 4th (44 %)	0.12 : 1st (0 %)	11.14 : 4th (43 %)	Dry	DE 20%+	Small	Non in-house
A-07	2 / 0	0.60 : 4th (20 %)	0.90 : 3rd (0 %)	0.80 : 2nd (0 %)	1.10 : 4th (74 %)	6.30 : 3rd (6 %)	Wet	DE <20%	Small	In-house
A-08	2 / 1	0.80 : 4th (56 %)	2.30 : 4th (44 %)	2.20 : 3rd (0 %)	0.00 : 1st (0 %)	7.50 : 3rd (0 %)	Dry	DE 20%+	Small	In-house
A-09	2 / 2	0.27 : 2nd (0 %)	1.21 : 4th (17 %)	0.10 : 1st (0 %)	1.08 : 4th (83 %)	4.53 : 1st (0 %)	Mixed	DE <20%	Large	In-house
A-10	2 / 2	0.75 : 4th (60 %)	0.38 : 1st (0 %)	1.17 : 3rd (0 %)	0.24 : 1st (3 %)	8.05 : 4th (37 %)	Mixed	DE <20%	Medium	In-house
A-11	1 / 0	0.39 : 3rd (18 %)	0.93 : 3rd (0 %)	8.38 : 4th (72 %)	0.32 : 2nd (10 %)	5.99 : 2nd (0 %)	Wet	DE <20%	Medium	Non in-house

Number [%] of DEA effective facilities in each family

Type	Dry: n= 3 [14%] Mixed: n= 4 [9%] Wet: n= 4 [12%]	***Location***	DE <20%: n= 7 [14%] DE 20%+: n= 4 [8%]	***Size***	Small: n= 5 [20%] Medium: n= 3 [9%] Large: n= 3 [7%]	***Management***	In-house: n= 5 [9%] Non in-house: n= 6 [13%]

Notes: Definitions of quartiles: 4th ≥ 75%; 75% > 3rd ≥ 50%; 50% > 2nd ≥ 25%; 1st < 25%. Number within the ( ) bracket is the virtual weight of each PI.

Table 6.13 Alternative way to select benchmarking partners

Centre code	No. of PIs in the 4th / 1st quartile	DEA Score	PIs and quartiles					Families			
			Social class DE	11-19 years	Ethnic minority	60+ years	Disabled <60 years	Type	Location	Size	Management
A-01	4 / 0	1.00	0.44 4th	0.62 2nd	3.66 4th	0.89 4th	11.79 4th	Wet	DE 20%+	Large	In-house
A-02	4 / 1	1.00	0.43 4th	0.44 1st	6.09 4th	0.83 4th	8.43 4th	Mixed	DE <20%	Large	Non in-house
A-03	3 / 0	1.00	0.81 4th	1.33 4th	3.78 4th	0.48 3rd	6.09 2nd	Mixed	DE <20%	Medium	Non in-house
A-04	3 / 0	1.00	0.80 4th	0.91 3rd	4.19 4th	0.80 4th	5.70 2nd	Wet	DE <20%	Small	Non in-house
A-05	3 / 1	1.00	0.14 1st	3.78 4th	2.74 4th	0.59 3rd	8.23 4th	Dry	DE 20%+	Small	Non in-house
A-06	3 / 1	1.00	0.30 3rd	2.01 4th	9.33 4th	0.12 1st	11.14 4th	Dry	DE 20%+	Small	Non in-house
B-01	4 / 0	0.99	0.57 4th	1.45 4th	4.88 4th	0.45 3rd	8.88 4th	Wet	DE <20%	Small	In-house
B-02	4 / 0	0.95	0.42 4th	1.70 4th	4.25 4th	0.33 2nd	9.83 4th	Mixed	DE <20%	Large	In-house
B-03	4 / 0	0.88	0.40 4th	1.10 3rd	4.60 4th	0.60 4th	8.40 4th	Mixed	DE 20%+	Medium	In-house
B-04	4 / 0	0.84	0.43 4th	0.62 2nd	2.65 4th	0.77 4th	8.04 4th	Wet	DE 20%+	Small	Non in-house
B-05	3 / 0	0.84	0.40 4th	0.54 2nd	0.84 2nd	0.72 4th	9.39 4th	Wet	DE <20%	Large	Non in-house
B-06	3 / 0	0.83	0.38 3rd	0.61 2nd	2.53 4th	0.78 4th	8.30 4th	Wet	DE 20%+	Medium	In-house
B-07	3 / 0	0.83	0.37 3rd	1.14 4th	2.89 4th	0.70 4th	6.57 3rd	Wet	DE <20%	Medium	Non in-house
B-08	3 / 0	0.81	0.28 2nd	0.55 2nd	3.01 4th	0.64 4th	9.57 4th	Mixed	DE <20%	Medium	Non in-house
B-09	3 / 1	0.82	0.32 3rd	1.83 4th	2.53 4th	0.08 1st	8.23 4th	Dry	DE <20%	Medium	Non in-house

Number of facilities in each family

Type	Dry: n= 3 Mixed: n= 5 Wet: n= 7	***Location***	DE <20%: n= 9 DE 20%+: n= 6	***Size***	Small: n= 5 Medium: n= 6 Large: n= 4	***Management***	In-house: n= 5 Non in-house: n= 10

Notes: The dot line separates effective/ineffective centres deemed by DEA.

In fact, A-09, A-10 and A-11 are three typical examples reflecting an issue with DEA's weighting flexibility. For these three centres, full marks are achieved simply by placing a lot of weight on one dimension in which they perform well and assigning a zero weight to the dimensions in which they perform poorly. Whether this kind of unbalanced performance can be deemed as the best practice in achieving sport equity is doubtful. This result raises a series of questions: Is this kind of weighting method appropriate for practitioners? Which criterion, highlighting the strengths or focusing on the improvement of weaknesses, should be taken? Can DEA-selected benchmarking partners be regarded as beacons for other centres? Do they really achieve sport equity?

The above questions stimulate the researcher to reflect whether there is another alternative way to select benchmarking partners. Table 6.13 provides an easier and probably more convincing approach to select suitable benchmarking partners. The proposed criterion is to select those centres that have at least three PIs in the top-quartile and no more than one PI in the bottom-quartile. It must be admitted that this criterion is somewhat subjective, but it offers the possibility of developing other rules to reflect different priorities if they are considered necessary in the future. As shown in Table 6.13, 15 centres meet this criterion, where the first six (from A-01 to A-06) are consistent with the DEA results. Even though the last nine centres (from B-01 to B-09) are not deemed as DEA effective, they still have excellent and balanced performance. For example, four of them (from B-01 to B-04) have four PIs in the top-quartile and no PI in the bottom-quartile. They can not achieve 100% DEA effectiveness simply because they have a more balanced performance and no specific PI is high enough to be assigned an extremely high weight. Obviously, this alternative

reflects a critical issue: using DEA scores to establish a league table for ranking the centres' overall performance is questionable because of the weighting procedure adopted by standard DEA. Finally, the statistics in the bottom of Table 6.13 show again that all types, locations, sizes and management types can reach best practice performance, in principle.

6.2.3 Discussion

The value of DEA is limited while measuring equity. A series of questions raised in the above analysis indicate that several technical problems need to be addressed so that the potential benefits of DEA can be fully realised.

First, the problem of weighting flexibility is especially highlighted while measuring equity. From the research findings, certain facilities achieve effectiveness simply by having an unusual mix of outputs and are far from achieving equity because the access of certain target groups is ignored. As such, the application of DEA without weights restrictions may provide a misleading measurement of overall performance. As argued by Smith and Street (2005), the lack of a single set of weights implies that it is inappropriate to rank DEA scores in a conventional 'league table' format.

Second, the function of DEA in identifying strengths and weaknesses is more important in measuring efficiency rather than equity. The reason is that the results of DEA are basically consistent with the NBS results so there is no reason for the practitioners to adopt a more complicated model. Both the analyses at the industrial and individual facility levels confirm this point. While measuring efficiency, contradictory information usually exists between different efficiency ratios, as they link the interaction between different sets of inputs and outputs. Efficiency can be

achieved through either output maximisation or input minimisation. As for effectiveness, the measurement is more straightforward since output maximisation is the only target. Consequently, the identification of strengths and weaknesses in effectiveness PIs can simply examine their values relevant to the national benchmarks.

From an empirical perspective, the statistical evidence confirms the consistent under-representation of social class DE, 60+ years and disabled <60 years. In addition, the type, location and size of facility are found to be the main sources of performance gaps in certain PIs. Major concerns are the implications of the consistent under-representation performance of some PIs and structural performance differences. An important question is: how far can this be overcome by innovative management (e.g. pricing, programming, transport provision) and how far is this due to other reasons which are out of the control of facility managers (e.g. consumer preferences, wider social pressures)?

Sport inequities occur as a consequence of a complex interaction of cultural, social, geographical and economic factors. Coalter (1998; 2000) pointed out that many of those who do not use publicly provided facilities are not 'constrained' or 'excluded', they simply do not wish to use them. In his opinion, non-participation does not equal exclusion. Exclusion or inequity can be said to occur only when people want to take part but cannot. Such considerations raise important issues about the extent to which low usage indicates inequity. Gratton and Taylor (2000) also emphasised that it is not necessarily realistic to expect socially disadvantaged people to be well represented in the usage profiles of local authority sports facilities. Many constraints other than price prevent them from participating, some of which may be

beyond the reach of public policy. Nevertheless, the enduring nature of their under-representation is of concern if public subsidy continues to be used for equity reasons.

By contrast, if under-representation and performance difference are due to misguided policy or implementation failures, the experiences of good practice will become valuable for benchmarking purposes. In research investigating the centres that participated in the NBS and achieved top-quartile scores in access performance, Taylor (2004) noticed that common successful factors include: appropriate location near to large numbers of the relevant target groups, low prices (e.g. extensive use of leisure cards and possibly some free usage) and specific programming for target groups (especially young people, 60+, and health and disability targets). Veal (2002) also mentioned that often the facilities are available, but the challenge for management is to provide programmes at the right price, at the right time and at the right place to attract specific market segments. Table 6.13 is a starting point for developing a benchmarking platform. From this table, centres can choose those with similar performance levels and/or similar family categorisation. Then qualitative research can be further conducted to investigate the reasons for performance difference and in turn facilitate process benchmarking.

On the other hand, local authorities' sport services are facing serious funding problems. While social exclusion is on the increase, financial pressures are diverting councils from their aims and objectives (Connolly, 2002). Furthermore, according to Taylor (2004), it is the mainstream market (e.g. younger, middle classes) instead of social inclusion targets that can make the achievement of Government's target for sport and physical activity more realistic. As such, the challenge for facility managers

is to increase market penetration and frequency of participation by mainstream markets as well as to design specific, targeted access for the socially excluded. Undoubtedly, creative thinking is necessary to find the correct balance between social, financial and political objectives, and benchmarking is a helpful vehicle to achieve this.

The other implication of the findings is to demonstrate the complexities of what is meant by sport equity. Different sports and different settings seem to have their own (different) inequities. In this case, 'ethnic minority' and '11-19 years' are relatively well-represented while comparing with other targeted groups. It implies that future interventions to tackle inequity in sport will need to be far more subtle and targeted.

CHAPTER 7

Results of Customer Service Measurement

Road map of the thesis

Introduction	Theoretical context	Methodology	Results	Conclusion
Ch1 ▶	Ch2 Performance measurement overall Ch3 Efficiency & Equity Ch4 Customer service	▶ Ch5 ▶	Ch6 Efficiency & Equity Ch7 Customer service Ch8 Action research	▶ Ch9

7.1 Customer Service Quality

The first objective of this section is to provide an insight into the industry's performance in customer service quality. The second objective is to examine whether alternative service quality measurement models yield consistent results and discuss the implications. This is done by aggregating a series of performance attributes that allow the strengths and weaknesses of the industry to be clarified. Also, statistical tests are conducted to examine whether significant performance differences exist between facilities with different profiles.

7.1.1 Descriptive Statistics

Research data are derived from the NBS database, where user surveys of 23,329 respondents were surveyed at 72 public sport facilities in England during 2001. The descriptive statistics illustrating the profile of respondents and facilities are shown in Table 7.1.

Table 7.1 Respondent and facility profiles

Respondent profiles (n = 23,329)				Facility profiles (n = 72)	
Gender		Join type		Facility type	
Male	40%	Individual user	59%	Dry	26%
Female	60%	Organised class	27%	Mixed	46%
		Club/team member	11%	Wet	28%
Ethnicity		Other	3%		
White	94%			Facility location [a]	
Ethnic minority	6%	Discounted card user		DE <15%	13%
		Yes	15%	DE 15% to 20%	39%
Age		No	85%	DE 20%+	48%
<16 yrs	6%				
16-24 yrs	12%	First visit		Facility size	
25-44 yrs	55%	Yes	5%	<1,500 sq.m.	30%
45-64 yrs	20%	No	95%	1,500 to <3,000 sq.m	35%
65+ yrs	7%			3,000+ sq.m.	35%
		Frequency (last 1 week)			
Occupation		1 time	29%	Management type	
Working full-time	44%	2 times	33%	In-house	50%
Working part-time	19%	3+ times	38%	Commercial	15%
Student	13%			Trust	17%
Retired	10%			Other	18%
Housewife/husband	7%				
Unemployed	4%				
Other	3%				

Notes: [a] DE = the lowest social-economic grouping in the UK

Generally speaking, over 50% of respondents are female, or aged from 25 to 44 years old, or working full-time or part-time, and a high proportion of users are white. In terms of visit profiles, the majority is individual, not first visit, non-member, non-discount-card user, and visited at least three times in the previous week. The distribution of these families in the study is shown in the right hand side of Table 7.1. As the sample sizes of certain subgroups are relatively small, we combined DE <15% and DE 15% to 20% were again combined into one subgroup - 'DE <20%' as well as commercial contracted, trust and other management types into one subgroup- 'non in-house' in order to yield more balanced sub-sample comparisons.

As to the survey administration, it is conducted over a standard survey period of nine consecutive days, including two weekends. Two-part 25-item questions are included, which seek to identify customers' satisfaction and importance levels on a five-point Likert scale extending from 1 (very dissatisfied/very unimportant) to 5 (very satisfied/very important). Then, for each statement the mean importance (I) and satisfaction (S) values, along with a service quality value (Q) from the formula are presented as: Q= S-I. If the S-I is negative, it refers to the presence of a service quality gap. These 25 items are further grouped into six service dimensions: accessibility, availability, quality of facility, cleanliness, staff and value for money. The descriptive statistics illustrating the average importance, performance and gap scores are shown in Table 7.2.

Table 7.2 NBS attributes: satisfaction, importance and gap means and *t*-tests

Attribute	Satisfaction	Importance	Gap	*t*-value	
Accessibility					
Activity at convenient time	4.22	4.48	-0.25	14.55	**
Ease of booking in advance	4.30	4.14	+0.16	6.40	**
The activity charges/fees	4.09	4.35	-0.26	13.67	**
Availability					
Availability of car parking	4.20	4.41	-0.21	4.92	**
Availability of crèche facilities	3.42	3.59	-0.18	4.06	**
Availability of food/drink	3.74	3.67	+0.07	2.07	*
Range of activities available	4.03	4.26	-0.23	12.34	**
Availability of equipment	3.90	4.30	-0.40	13.17	**
Quality of facility					
Quality of flooring in sports hall	3.97	4.36	-0.39	11.77	**
Quality of lighting in sports hall	4.00	4.36	-0.36	11.22	**
Water quality in pool	3.97	4.57	-0.60	17.12	**
Water temperature of pool	3.83	4.47	-0.64	14.98	**
Number of people in pool	3.83	4.36	-0.53	15.01	**
Quality of food/drink	3.68	3.93	-0.25	8.24	**
Cleanliness					
Cleanliness of reception area	4.22	4.36	-0.14	4.57	**
Cleanliness of changing area	3.75	4.61	-0.86	15.11	**
Cleanliness of café/bar	3.98	4.42	-0.45	11.77	**
Staff					
Helpfulness of reception staff	4.41	4.59	-0.17	8.71	**
Helpfulness of other staff	4.43	4.59	-0.17	9.55	**
Friendliness of staff	4.43	4.59	-0.16	9.20	**
Availability of coaching	4.22	4.49	-0.27	14.21	**
Standard of coaching	4.29	4.56	-0.27	14.73	**
Value for money					
Value for money of activities	4.18	4.58	-0.39	22.69	**
Value for money of food/drink	3.76	4.15	-0.39	14.47	**
Value for money overall	4.08	4.57	-0.48	24.18	**
Max	4.43	4.61	-0.86		
Min	3.42	3.59	+0.16		
Average	4.04	4.35	-0.31		

Notes: * <0.05; ** <0.01.

The average satisfaction and importance scores, ranged from 3.42 to 4.43 and from 3.59 to 4.61 respectively, are all above the neutral point 3, signifying relatively high evaluations. However, the findings showed shortfalls on 23 of the 25 attributes measured - only 'ease of booking in advance' and 'availability of food/drink' have positive gap scores. Paired *t*-tests demonstrate significant differences between satisfaction and importance in 24 attributes at the 0.01 level and 1 attribute at the 0.05 level, identifying statistically significant service quality gaps. The five staff-related attributes generally had higher satisfaction scores as well as higher importance scores. Even though availability of crèche facilities and availability and quality of food/drink had lower satisfaction scores, their service gaps are not large because the importance assigned to these attributes is relatively low. The largest service gaps are found in 'cleanliness of changing area' and the three attributes related to the quality of swimming pool- water temperature and quality and number of people in the pool.

7.1.2 Factor Analysis

The analysis at the first stage aimed to determine the major factors underlining the NBS framework. To achieve this objective, the original NBS service quality attributes (both satisfaction and importance means) are factor-analysed by principal component analysis. In factor analysis, a rotation procedure is commonly applied, which maximises the correlations of items within a factor. The NBS construct comprises many interrelated items and, therefore, oblique rotation is applied as the rotation procedure. Advocates of oblique rotation assert that, in the real world,

important factors are likely to be correlated, thus searching for unrelated factors is unrealistic (Hair *et al.*, 1995).

First, six factors are extracted in the analysis using a standard eigenvalue of 1.0 (Hair *et al.*, 1995). The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy is used to measure the adequacy of the sample for extraction of the six factors. The KMO value 0.88 is indicative of a data set considered to be highly satisfactory for factor analysis. The Bartlett's test of sphericity is used to test the multivariate normality of the set of distributions. This procedure also tested whether the correlation matrix is an identity matrix (factor analysis would be meaningless with an identity matrix). A significance value at the 0.05 level confirms that the data are approximately multivariate normal and acceptable for factor analysis. The details of the factor analysis are shown in Table 7.3.

To compare the performance of facilities with different type, location, size and management type and investigate whether performance differences exist between these NBS families, the mean values of six factors yielded from the factor analysis and from three different service quality measurement models are cross-tabulated with the four NBS families. The ANOVA/*t*-test is then conducted to test the null hypothesis that the means are equal between sub-samples. In terms of the whole sample, the overall means and rankings of six service dimensions yielded from the three measurement models are shown in the bottom of Table 7.4.

Table 7.3 Result of factor analysis: satisfaction and importance

Factors	*Satisfaction*				*Importance*				NBS dimensions
	Factor loading	Eigen-Value	% of Variance	Cronbach α	Factor loading	Eigen-Value	% of Variance	Cronbach α	
Factor 1: Catering		7.67	31.97	0.85		1.91	7.97	0.84	
Availability of food/drink	0.84				0.84				Availability
Quality of food/drink	0.83				0.78				Quality
VFM of food/drink	0.73				0.69				VFM
Factor 2: Staff		1.80	7.49	0.89		8.17	34.04	0.91	
Standard of coaching	0.86				0.88				Staff
Availability of coaching	0.85				0.87				Staff
Helpfulness of other staff	0.81				0.84				Staff
Friendliness of staff generally	0.67				0.73				Staff
Factor 3: Quality of wet facility		1.43	5.95	0.78		1.61	6.72	0.87	
Water quality in pool	0.85				0.89				Quality
Water temperature in pool	0.85				0.90				Quality
Number of people in pool	0.75				0.83				Quality
Factor 4: Accessibility		1.30	5.42	0.79		1.44	6.00	0.71	
Activity charges/fees	0.79				0.76				Accessibility
Activities at convenient times	0.73				0.78				Accessibility
Ease of booking in advance	0.65				0.65				Accessibility
Factor 5: Quality of dry facility		1.26	5.26	0.79		1.25	5.21	0.93	
Flooring in sports hall	0.85				0.94				Quality
Lighting in sports hall	0.85				0.94				Quality
Factor 6: Cleanliness		1.13	4.72	0.85		1.10	4.57	0.83	
Cleanliness of reception area	0.77				0.77				Cleanliness
Cleanliness of changing area	0.74				0.71				Cleanliness
Cleanliness of café/bar area	0.70				0.69				Cleanliness
Overall			60.80	0.93			64.49	0.93	

Table 7.4 Cross-tabulation analysis for service quality measurement: factors with NBS families

		Mean values																	
	n	***F1-Catering***			***F2-Staff***			***F3-Wet facility***			***F4-Accessibility***			***F5-Dry facility***			***F6-Cleanliness***		
		I	II	III	I	II	III	I	II	III	I	II	III	I	II	III	I	II	III
Type																			
Dry	19	3.77	-0.13	-0.53	4.36	-0.14**	-0.63**	n.a.	n.a.	n.a.	4.28*	-0.15*	-0.65*	4.11	-0.36	-1.64	4.18**	-0.19**	-0.84**
Mixed	33	3.72	-0.24	-0.97	4.31	-0.27**	-1.25**	3.89	-0.66	-3.00	4.16*	-0.23*	-0.99*	3.94	-0.43	-1.87	3.88**	-0.62**	-2.77**
Wet	20	3.68	-0.17	-0.63	4.36	-0.21**	-0.99**	4.00	-0.58	-2.66	4.19*	-0.15*	-0.66*	n.a.	n.a.	n.a.	3.96**	-0.54**	-2.40**
Location																			
DE <20%	37	3.71	-0.17	-0.65	4.34	-0.19	-0.89	3.95	-0.61	-2.77	4.20	-0.18	-0.79	4.04	-0.34	-1.51	4.04	-0.39*	-1.73*
DE 20%+	35	3.74	-0.22	-0.87	4.35	-0.25	-1.14	3.91	-0.65	-2.97	4.19	-0.19	-0.83	3.97	-0.47	-2.06	3.92	-0.58*	-2.62*
Size																			
Small	22	3.78	-0.07*	-0.26**	4.39	-0.15**	-0.66**	4.09	-0.51	-2.34	4.26	-0.12**	-0.52**	4.10	-0.38	-1.71	4.18**	-0.23**	-1.05**
Medium	25	3.70	-0.23*	-0.92**	4.32	-0.25**	-1.13**	3.89	-0.65	-2.93	4.18	-0.19**	-0.82**	4.02	-0.34	-1.46	3.93**	-0.54**	-2.41**
Large	25	3.70	-0.26*	-1.03**	4.32	-0.26**	-1.20**	3.90	-0.66	-3.02	4.16	-0.24**	-1.05**	3.91	-0.48	-2.12	3.87**	-0.64**	-2.89**
Management																			
In-house	36	3.71	-0.22	-0.89	4.33	-0.24	-1.10	3.90	-0.64	-2.91	4.20	-0.18	-0.80	3.99	-0.41	-1.83	3.92	-0.56*	-2.52*
Non in-house	36	3.73	-0.16	-0.63	4.35	-0.20	-0.93	3.96	-0.62	-2.83	4.19	-0.19	-0.82	4.00	-0.40	-1.75	4.05	-0.40*	-1.80*
Overall mean		3.72	-0.19	-0.76	4.34	-0.22	-1.01	3.92	-0.63	-2.87	4.20	-0.18	-0.81	4.00	-0.40	-1.79	3.98	-0.48	-2.16
Ranking		6th	2nd	1st	1st	3rd	3rd	5th	6th	6th	2nd	1st	2nd	3rd	4th	4th	4th	5th	5th

Notes: I = performance-only analysis; II = gap analysis; III = Weighted gap analysis; n.a. = not applicable

* <0.05; ** <0.01 (ANOVA/*t*-test)

Generally speaking, wet facility and cleanliness have relatively lower performance and accessibility has relatively higher performance no matter which model is considered. In terms of sub-sample comparisons (as shown in the upper part of Table 7.4), on first impression, the performance of DE <20%, small, non in-house and dry facilities are generally better than the remaining ones in the same family categories, i.e. higher scores for model I and smaller gaps for models II and III. The exceptions could be found in: location (catering and staff of DE <20% in the performance-only model are worse than DE 20%+), size (medium outperformed others in dry facilities when using two gap-based models) and management type (in-house outperformed non in-house in accessibility no matter which model is adopted). However, the ANOVA/*t*-test reveals there are only statistically significant differences in:

- catering, due to size,
- staff, due to facility type and size,
- accessibility, due to facility type and size, and
- cleanliness due to facility type, size and location, and management type.

The implications of this analysis are threefold. First, the two gap-based models have similar results either in overall ranking of six factors or in the ANOVA/*t*-test. However, highly inconsistent results are shown between performance-only analysis and two gap-based analyses for catering. Even though catering has the lowest ranking in terms of satisfaction, the service quality gaps of catering are relatively small because lower importance is assigned to it. Second, facility type and size seem to be the main sources of performance differences in three dimensions - cleanliness, accessibility and staff - this phenomenon being especially evident in the two

gap-based models. That is, for these three service dimensions, facilities without pools outperform those with pools (wet and mixed) and smaller facilities outperform medium and large facilities. Third, significant service gaps are found in cleanliness no matter which NBS family is considered.

7.1.3 Cluster Analysis

The second stage of data analysis involved a cluster analysis to identify possible segments according to the average performance gap scores and satisfaction scores of the whole sample based on the six factor groupings. Facilities with similar performance gap scores and satisfaction scores are grouped into the same cluster. K-means cluster analysis is performed on different cluster solutions. An examination of these solutions revealed that three clusters is the most appropriate solution in terms of the separation of the clusters and homogeneity in the size of the groups. For solutions with less than three clusters, the *F*-values are smaller, and for solutions with more than three clusters, very small clusters appeared from the splitting of larger ones, resulting in clusters of very different sizes. It is worth mentioning that there are no strict rules for assessing the adequacy of a cluster solution. It is essentially a tool used to summarise and simplify a data set in a way that makes it manageable and provide insights into grouping patterns. The results of cluster analysis, including the means of each cluster for each of the quality dimensions yielded from factor analysis as well as ANOVA and MANOVA - performed to evaluate the cluster separation - are shown in Table 7.5.

Table 7.5 Cluster analysis for service quality measurement: cluster means, ANOVA and MANOVA

	Performance-only analysis				***Gap analysis***				***Weighted gap analysis***			
	Cluster 1 (n = 28)	Cluster 2 (n = 21)	Cluster 3 (n = 23)	F value	Cluster 1 (n = 36)	Cluster 2 (n = 24)	Cluster 3 (n = 12)	F value	Cluster 1 (n = 30)	Cluster (n = 30)	Cluster 3 (n = 12)	F value
F1- Catering	3.95 ↑	3.51 ↓	3.65 –	26.10**	-0.05 ↑	-0.34 ↓	-0.33 –	18.87**	-0.17 ↑	-1.15 –	-1.24 ↓	14.14**
F2- Staff	4.47 ↑	4.32 –	4.21 ↓	26.63**	-0.14 ↑	-0.28 –	-0.35 ↓	26.44**	-0.53 ↑	-1.22 –	-1.71 ↓	37.64**
F3- Wet facility	4.13 ↑	3.93 –	3.78 ↓	22.32**	-0.50 ↑	-0.68 –	-0.78 ↓	14.07**	-2.07 ↑	-3.02 –	-3.67 ↓	21.50**
F4- Accessibility	4.34 ↑	4.15 –	4.08 ↓	36.17**	-0.11 ↑	-0.25 –	-0.27 ↓	15.39**	-0.44 ↑	-1.01 –	-1.23 ↓	17.51**
F5- Dry facility	4.15 ↑	4.02 –	3.76 ↓	16.57**	-0.29 ↑	-0.42 –	-0.79 ↓	22.58**	-1.15 ↑	-1.89 –	-3.48 ↓	24.46**
F6- Cleanliness	4.27 ↑	3.95 –	3.66 ↓	55.27**	-0.24 ↑	-0.63 –	-0.93 ↓	59.71**	-0.82 ↑	-2.72 –	-4.12 ↓	74.96**
MANOVA	Wilk's $\lambda = 0.16^{**}$				Wilk's $\lambda = 0.19^{**}$				Wilk's $\lambda = 0.19^{**}$			

Notes: ↑ = higher performance or smaller gap; – = medium performance or gap; ↓ = lower performance or larger gap.

* <0.05; ** <0.01.

With regard to mean comparisons among the three clusters, except for catering in performance-only analysis and gap analysis, cluster 1 is the group with the highest satisfaction and smallest performance gap, followed by cluster 2 and cluster 3. In terms of separation accuracy, the ANOVA indicates significant differences between clusters for each of the six factors at the 0.01 level, where cleanliness has the highest *F*-value of all of the three measurement models, followed by accessibility in the performance-only model and staff in two gap-based models. When all six quality dimensions are considered simultaneously, the MANOVA shows significant differences between clusters at the 0.01 level no matter which measurement model is applied.

In order to investigate whether different cluster distributions existed between different NBS families, cross-tabulation analysis and chi-squared tests are then conducted and the results are shown in Table 7.6. First, solely considering the cluster distributions, the majority of dry, DE <20%, small and non in-house facilities are in cluster 1 (the highest performance cluster), ranging from 41% to 82% in all of the three measurement models. While comparing their performance with the remaining ones in the same family categories, relatively few facilities are located in cluster 3 (the lowest performance cluster), ranging from 5% to 27%. Chi-squared tests further demonstrate that significant differences are found in facility type no matter which measurement model is considered and in facility size when the two gap-based models are used. The differences are especially significant ($p<0.01$) when weighted gap analysis is adopted.

Table 7.6 Cross-tabulation analysis for service quality measurement: clusters with NBS families

		Performance-only analysis						***Gap analysis***						***Weighted gap analysis***					
	n	Cluster 1		Cluster 2		Cluster 3		Cluster 1		Cluster 2		Cluster 3		Cluster 1		Cluster 2		Cluster 3	
		n	(%)	*n*	(%)	*n*	(%)	*n*	(%)	*n*	(%)	*n*	(%)	*n*	(%)	*n*	(%)	*n*	(%)
Type				$\chi^2 = 10.79^*$						$\chi^2 = 11.82^*$						$\chi^2 = 13.60^{**}$			
Dry	19	12	(63%)	6	(32%)	1	(5%)	15	(79%)	3	(16%)	1	(5%)	14	(74%)	4	(21%)	1	(5%)
Mixed	33	9	(27%)	11	(33%)	13	(40%)	11	(33%)	16	(49%)	6	(18%)	8	(24%)	19	(58%)	6	(18%)
Wet	20	7	(35%)	4	(20%)	9	(45%)	10	(50%)	5	(25%)	5	(25%)	8	(40%)	7	(35%)	5	(25%)
Location				$\chi^2 = 0.90$						$\chi^2 = 5.39$						$\chi^2 = 7.22$			
DE <20%	37	15	(41%)	12	(32%)	10	(27%)	23	(62%)	8	(22%)	6	(16%)	21	(57%)	11	(30%)	5	(13%)
DE 20%+	35	13	(37%)	9	(26%)	13	(37%)	13	(37%)	16	(46%)	6	(17%)	9	(26%)	19	(54%)	7	(20%)
Size				$\chi^2 = 7.06$						$\chi^2 = 13.09^*$						$\chi^2 = 18.63^{**}$			
Small	22	13	(59%)	6	(27%)	3	(14%)	18	(82%)	2	(9%)	2	(9%)	17	(77%)	2	(9%)	3	(14%)
Medium	25	7	(28%)	7	(28%)	11	(44%)	9	(36%)	11	(44%)	5	(20%)	8	(32%)	13	(52%)	4	(16%)
Large	25	8	(32%)	8	(32%)	9	(36%)	9	(36%)	11	(44%)	5	(20%)	5	(20%)	15	(60%)	5	(20%)
Management				$\chi^2 = 5.28$						$\chi^2 = 1.11$						$\chi^2 = 2.07$			
In-house	36	12	(33%)	8	(22%)	16	(45%)	16	(44%)	14	(39%)	6	(17%)	12	(33%)	17	(47%)	7	(20%)
Non in-house	36	16	(45%)	13	(36%)	7	(19%)	20	(55%)	10	(28%)	6	(17%)	18	(50%)	13	(36%)	5	(14%)

Notes: * <0.05; ** <0.01.

This finding could be cross-referred to Table 7.4; that is, statistically significant differences are found in cluster distributions (Table 7.6) and mean value comparisons of certain factors (Table 7.4) due to facility type and size, especially in the two gap-based measurement models. It could therefore be concluded that dry and small facilities had smaller service quality gaps, and there are structural differences in service quality due to facility type and size.

7.1.4 Discriminant Analysis

Discriminant analysis is performed on the above three clusters in order to identify the classification accuracy of cluster analysis and to investigate which quality dimensions best discriminate the three clusters. The results of three independent stepwise discriminant analyses are shown in Table 7.7.

Two discriminant functions are yielded from the analysis of three quality measurement models, but within the two gap-based models, only function 1 is significant at the 0.05 level and therefore shown in the table. To test the significance of discriminant functions, the canonical correlation coefficients and the Wilks' λ statistics are examined. The values of Wilks' λ transformed as χ^2 indicate the significance of all models at the 0.01 level. Also, the overall classification rates indicate that three discriminant analysis models could correctly classify 84.7%, 81.9% and 86.1% of the individuals into clusters respectively, suggesting high classification accuracies and the validity of the three-cluster solution. In terms of the relative contributions of six service dimensions in discriminating the three clusters, the standardised discriminant coefficients and structure coefficients (only selected variables are shown in the table) indicate that cleanliness has the highest

discriminating power and in the positive direction in all of the three quality measurement models. It implies that cleanliness is the best predictor of classifying different levels of overall service quality.

Table 7.7 Stepwise discriminant analysis for service quality measurement

	Performance-only analysis				***Gap analysis***		***Weighted gap analysis***	
	Function 1		Function 2		Function 1		Function 1	
	a	b	a	b	a	b	a	b
F1- Catering	-0.37	0.05	1.06	0.99	-	-	-	-
F2- Staff	0.58	0.77	0.09	0.17	0.58	0.77	-	-
F3- Wet facility	-	-	-	-	-	-	-	-
F4- Accessibility	-	-	-	-	-	-	-	-
F5- Dry facility	-	-	-	-	-	-	0.50	0.68
F6- Cleanliness	0.70	0.82	-0.16	0.38	0.66	0.83	0.75	0.87
Eigenvalue	2.35		0.61		2.80		2.92	
% of variance	79.39%		20.61%		99.99%		97.64%	
Canonical correlation	0.84		0.62		0.86		0.86	
Wilks' λ	0.19		0.62		0.26		0.24	
χ^2	48.88**		13.81**		39.43**		42.34**	
% of classification accuracy	84.7%				81.9%		86.1%	

Notes: a= standardised discriminant coefficients; b= structure coefficients.
* <0.05; ** <0.01.

7.1.5 Regression Analysis

The final analysis examined which service performance dimensions contributed the most to overall value for money (VFM) perceived by customers. It is ideal that perceptions relating to VFM need to be measured, particularly in the existing public sector context, given the Best Value regime, and this has been highlighted in previous

researches (e.g. Brysland and Curry, 2001). However, it should be noticed that one limitation of this study is to take overall VFM as a surrogate of overall service quality measurement because the other more suitable summary indicator- 'overall satisfaction' was not available from the 2001 NBS database. To test the relationship between VFM and service quality dimensions, the performance-only scores, gap scores and weighted gap scores of six dimensions anchored by factor analysis are used as independent variables in three stepwise regression analyses against a separate measure of overall VFM. The results are shown in Table 7.8.

Table 7.8 Stepwise regression analysis for service quality measurement

	Performance -only analysis			***Gap analysis***			***Weighted gap analysis***		
	β	t	VIF	β	t	VIF	β	t	VIF
F1- Catering	0.14	2.23*	1.64	-	-	-	-	-	-
F2- Staff	-	-	-	-	-	-	-	-	-
F3- Wet facility	-	-	-	-	-	-	-	-	-
F4- Accessibility	0.66	9.67**	1.90	0.57	5.85**	1.42	0.56	5.72**	1.40
F5- Dry facility	-	-		-	-	-	-	-	-
F6- Cleanliness	0.22	2.93**	2.27	0.25	2.62*	1.42	0.25	2.59*	1.40
R	0.91			0.74			0.73		
R^2	0.83			0.54			0.53		
Adjusted R^2	0.83			0.53			0.52		
F	112.99**			41.05**			38.77**		
Condition index	1.00 ~ 91.44			1.00 ~ 4.27			1.00 ~ 4.22		

Notes: VIF = variance inflation factor. Criteria: VIF <10, condition index <30 (Hair *et al.*, 1995). * <0.05; ** <0.01

First, the standardised regression coefficients (β) show that the most important factor in predicting customers' perceived VFM is accessibility, followed by

cleanliness. Within the three models, the *t*-values of accessibility and cleanliness are all significant at the 0.05 level, and the positive regression coefficients indicate that the selected variables positively affect the respondents' overall VFM. The dimensions staff, quality of wet facility and quality of dry facility are excluded in the regression models, signifying the relatively less important weight of these three dimensions in shaping overall VFM. Second, the coefficient of determination (R^2) of the three models indicates that 83%, 54% and 53% of the variances in overall VFM are explained by the selected service quality factors. Also, the *F*-values indicate that the three regression models are all significant at the 0.01 level. Finally, the multicollinearity of the variables in the regression equations is also examined, and variance inflation factor (VIF) values as well as condition index are performed. All VIF values of the selected independent variables are less than 10. However, high condition indexes are found in the performance-only model (highest value= 91.44), indicating a significant multicollinearity problem in this model. For the other two models, all condition indexes are less than 30, indicating no violation occurred (Hair *et al.*, 1995). The lack of significant multicollinearity between the factors makes the two gap-based models more reliable.

In sum, it can be concluded that overall VFM is strongly related to accessibility and cleanliness, no matter which service measurement model is considered. This finding could be further complemented by the descriptive statistics shown in Table 7.4, i.e. in terms of satisfaction level and performance gap, accessibility outperformed other factors with rankings of 2nd, 1st and 2nd and cleanliness had the worse performance with rankings of 4th, 5th and 5th in the three quality measurement models. The evidence jointly provided by regression analysis and descriptive statistics implies

that accessibility is the main driver of customers' perceived VFM and the strength of public sport facilities. On the other hand, as a major weakness of public sport facilities, cleanliness also has a significant impact on shaping the perception of overall VFM.

7.1.6 Discussion

From a theoretical perspective, the above analysis reviews alternative models of measuring customer service quality. A comparison between three models shows that two gap-based models basically yield similar results in the rankings of six service quality dimensions. Whether it is necessary to consider simultaneously two models is therefore debatable. In this research, it appears that traditional gap analysis is simpler and can be a good proxy of weighted gap analysis in clarifying strengths and weaknesses of the industry. On the other hand, inconsistent results are shown between performance-only analysis and the two gap-based analyses in the dimension of catering. In this case, the performance-only measure can be misleading for management, i.e. a low satisfaction attribute which is also low in importance to customers.

Although in line with some previous research (e.g. Burns *et al.*, 2003; Dabholkar *et al.*, 2000; Parasuraman *et al.*, 1994), the re-examination of alternative measurement models in a sport context, especially focusing on public facility provision, does contribute to justify the appropriateness of gap analysis, which might be the most popular approach in the industry. It means that gap analysis is not only highly acceptable by practitioners due to its simplicity, but its validity has also been demonstrated through this study.

From a managerial perspective, according to the typologies of service dimensions proposed by Grönroos (1984) and Booms and Bitner (1981), the performance of English public sport facilities in technical quality/physical evidence (i.e. cleanliness, wet facility and dry facility) are generally worse than functional quality/process and participants (i.e. accessibility and staff). These findings tend to support the importance of physical evidence suggested by some sport facility studies (e.g. Afthinos *et al.*, 2005; Alexandris *et al.*, 2004; Greenwell *et al.*, 2002; Lentell, 2000 etc.) For example, Lentell (2000) investigated customer satisfaction of seven local authority-owned indoor leisure facilities in the UK during the period 1996 to 1997 and found that the physical evidence dimension is considered to be the most important one to customers. Lentell (2000) also recommended that in order to secure better customer satisfaction with public leisure services, improving the tangibles may be the most effective. Another argument to complement our finding is the 'zone of tolerance' theory, i.e. customers are willing to absorb some positive or negative disconfrimation of expectations before expressing satisfaction or dissatisfaction (Johnston, 1995). It is generally agreed that if the service fails to satisfy customers, there will be a wider 'zone of tolerance' for process elements of the service (functional quality) but narrower for the outcome elements (technical quality).

In addition, according to Zeithaml *et al.* (1993), 'hygiene factors' (i.e. cleanliness in this case) are more important than others in shaping customers' assessment of quality and have the potential to lead to higher dissatisfaction when they are not present at the required level. Also, Bartikowski and Llosa (2004) and Bodet (2006) propose that some attributes always have an impact on satisfaction (the concept of fixed attribute weights), whereas other attributes are performance related (the concept

of variable attribute weights). The cleanliness of a facility is a typical example of the variable attribute weight because, generally, dirty facilities strongly negatively impact on customer satisfaction, although when the facilities are clean, it does not strongly affect customer satisfaction because it is considered to be a basic requirement or norm. In sum, it is suggested that more managerial concerns should be paid to the improvement of cleanliness and facility quality, not only because they are shown as the weaknesses of the industry but also because they are more important to customers and customers' 'zone of tolerance' for these dimensions is narrow.

7.2 Customer Segmentation

To ensure high standard customer service, the facility managers need not only to improve the service quality from the supply side but also need to identify customer needs from the demand side. This section therefore aims to identify customer segments based on customer's perceived importance of service quality. Research data are the same as the last section, i.e. user surveys of 23,329 respondents are collected from 72 public sport facilities in England during 2001. The analysis of this section basically follows the analytic procedure proposed by (Dolnicar, 2002), i.e. the combinational use of factor analysis for data reduction, cluster analysis for classification, analysis of variance and discriminant analysis for cross-method validations of the classification results and descriptive profiling of each cluster. The descriptive statistics and factor analysis of this section are the same as those in Section 8.1, and the results can refer to Table 7.1, 7.2 and 7.3.

7.2.1 Cluster Analysis

After conducting factor analysis, the second stage of data analysis involves a cluster analysis to identify possible customer segments based on a similar priority structure. The six service dimensions extracted from the factor analysis are used as clustering variables. Similar to the analysis in 7.1.3, K-means cluster analysis is performed and cluster solutions ranging from two to five clusters are examined. An examination of these solutions reveals that four clusters, rather than other number, is the most appropriate solution in terms of the separation of the clusters and producing the most interpretable results. The results of cluster analysis, including the means of each cluster for each of the quality dimensions yielded from factor analysis as well as means difference tests (ANOVA) performed to evaluate the cluster separation are shown in Table 7.9.

Table 7.9 Cluster analysis for customer segmentation analysis

Factors	Overall mean		Cluster mean								*F*-value
			Cluster 1 41.6%		Cluster 2 22.5%		Cluster 3 29.0%		Cluster 4 6.9%		
Physical evidence											
Wet facility	4.55	(1^{st})	4.75	(1^{st})	4.85	(3^{rd})	4.26	(2^{nd})	3.09	(5^{th})	4,832
Cleanliness	4.50	(2^{nd})	4.65	(2^{nd})	4.86	(2^{nd})	4.13	(3^{rd})	3.18	(4^{th})	5,181
Dry facility	4.35	(4^{th})	4.63	(3^{rd})	4.88	(1^{st})	3.78	(6^{th})	2.78	(6^{th})	6,246
Non-physical evidence											
Catering	4.36	(3^{rd})	3.86	(5^{th})	4.80	(4^{th})	4.32	(1^{st})	3.52	(1^{st})	2,568
Accessibility	4.32	(5^{th})	4.29	(4^{th})	4.75	(5^{th})	4.02	(4^{th})	3.26	(2^{nd})	2,945
Staff	3.76	(6^{th})	2.67	(6^{th})	4.43	(6^{th})	3.85	(5^{th})	3.19	(3^{rd})	5,865

Notes: Number with bracket is the mean ranking across factors. All reported *F*-values are significant at the 0.001 level.

In terms of separation accuracy, the ANOVA indicates significant differences (p<0.001) between clusters for each of the six factors, where 'dry facility' has the highest F-value, followed by 'staff' and 'cleanliness'. However, these F-values should be used only for descriptive purposes because the clusters have been chosen to maximise the differences among cases in different clusters. Furthermore, to delineate the clusters and to label them, the mean importance scores for each service dimensions for the members of each cluster are calculated and then form a priority structure of the four clusters. Based on the mean score characteristics with respect to the factors, these clusters are named as follows. In order to facilitate the denomination, the six service dimensions are grouped into two major kinds of service- physical evidence and non-physical evidence. The former consists of the quality of dry and wet facility as well as the cleanliness of facility. The latter consists of catering, accessibility and staff.

• Cluster 1: Physical evidence-wet focused (41.6%). This cluster represents the largest sample of the respondents. It has relatively higher mean scores on the physical evidence; especially the quality of wet facility which is the most important one. This customer segment, however, pays less attention to the non-physical evidence. It has the lowest mean score on staff, followed by catering and accessibility.

• Cluster 2: Physical evidence-dry focused (22.5%). This cluster is found to have the highest mean scores across all the factors. It has a similar priority structure to cluster 1, i.e. both segments place higher importance on the physical evidence, but lower importance on the rest of the factors. Also, the least important factor for both groups is staff. However, cluster 2 differs from cluster 1 in that its major concern is

the dry rather than the wet facility. In addition, there is a slight difference in the ranking of two non-physical evidence: catering and accessibility.

• Cluster 3: Physical evidence-wet & catering focused (29.0%). This cluster represents the second largest sample of the respondents. Compared to the above two clusters, the most special feature of cluster 3 is the highest importance level on catering and the lowest importance level on dry facility. Except for these two dimensions, physical evidence (i.e. wet facility and cleanliness) are still more important than non-physical evidence (i.e. accessibility and staff) for this customer segment. This cluster also resembles cluster 1 in terms of the importance of wet facility and cleanliness.

• Cluster 4: Non-physical evidence focused (6.9%). This cluster is characterised by the smallest segment of the market, the lowest mean scores across all the factors and higher importance ratings on the non-physical evidence. The customers score the highest on catering, followed by accessibility and staff. The quality of facility (wet and dry) is the least important dimension to them. In particular, both clusters 3 and 4 assign the lowest importance to the quality of dry facility. Basically, this segment has a contrasting priority structure compared with the other three clusters.

7.2.2 Cross-Tabulation Analysis

To further examine the differences among segments and provide practical information to formulate marketing strategy, we next turn to explore how these four customer segments differ. The way is to cross-tabulate each cluster with external variables, including customers' demographics and participation profiles. The differences are checked for statistical significance using chi-squared tests. As shown

in Table 7.10, the chi-squared tests indicate significant differences ($p<0.05$) between clusters in most of the profiles, except for 'with discounted card' and 'first visit'. To avoid problems with different sample sizes across clusters and among profiles, the data in Table 7.10 are normalised by using ratios of the percentage of customers with a characteristics in one cluster divided by the share of that characteristics in the whole sample. Take 'male' in cluster 1 as an example, the value 0.88 is yielded from the percentage of male in cluster 1 (34.85%) divided by the percentage of male in the whole sample (40.00%). A ratio over 1.0 indicates well-represented and less than 1.0 indicates under-represented. For example, '<16 yrs' group is highly represented in cluster 4, with a ratio of 1.84, meaning that the percentage of '<16 yrs' in cluster 4 (11.87%) is 84% higher than the percentage of '<16 yrs' in the whole sample (6.46%), i.e. 11.87% ÷ 6.46%= 1.84. In addition, to ease data interpretation, the highest ratio across clusters is underlined for each profile.

The four customer segments are delineated as follows and only the profiles with statistically significant differences between clusters are taken into account.

• Cluster 1: Physical evidence-wet focused (41.6%). This segment is strongly represented by older people aged 45-64 yrs (1.19), the 64+ yrs group (1.46) and retired people (1.49). It also contains a proportionally higher discounted card users (1.18), housewife/husband (1.11), unemployed (1.09) and females (1.08). By contrast, the representativeness of the age group <16 yrs is relatively low in this cluster (0.62). It appears that older people highlight more the physical evidence, especially the quality of the wet facility and cleanliness.

Table 7.10 Demographic and participation profiles of clusters

Profiles (n = 23,329)	Cluster 1 Physical -wet (41.6%)	Cluster 2 Physical -dry (22.5%)	Cluster 3 Physical -wet & catering (29.0%)	Cluster 4 Non-physical (6.9%)	χ^2	df	Sig.
Gender					204	3	0.000
Male	0.88	1.01	1.09	1.33			
Female	1.08	0.99	0.94	0.75			
Ethnicity					56	12	0.000
White	1.01	0.99	1.01	0.97			
Ethnic minority	0.87	1.21	0.89	1.56			
Age					682	12	0.000
<16 yrs	0.62	0.66	1.05	1.84			
16-24 yrs	0.83	0.85	1.18	1.78			
25-44 yrs	0.92	1.12	1.05	0.90			
45-64 yrs	1.19	0.97	0.83	0.66			
64+ yrs	1.46	0.67	0.70	0.51			
Occupation					742	18	0.000
Work full-time	0.88	1.09	1.08	1.11			
Work part-time	1.03	1.14	0.94	0.60			
Student	0.89	0.74	1.17	1.78			
Retired	1.49	0.66	0.67	0.53			
Housewife/husband	1.11	0.99	0.94	0.61			
Unemployed	1.09	0.97	0.94	0.79			
Join type					88	9	0.000
Individual	0.98	0.98	1.03	1.03			
Organised class	1.08	1.00	0.95	0.75			
Club/team member	0.92	1.06	0.95	1.47			
Discounted card user?					99	6	0.000
Yes	1.18	1.11	0.88	0.80			
No	0.97	0.98	1.02	1.04			
First visit?					4	3	0.227
Yes	1.03	0.90	1.06	0.93			
No	1.00	1.01	1.00	1.00			
Frequency (week)					9	9	0.419
1 time	0.99	1.03	0.99	0.97			
2 times	0.99	1.04	1.00	0.93			
3+ times	1.02	0.97	1.00	0.98			

Note: Ratios underlined are the highest one across the four clusters; Chi-squared tests are based on original percentages.

• Cluster 2: Physical-dry focused (22.5%). This segment tends to attract higher proportions of ethnic minority (1.21), part-time workers (1.14), people aged 25-44 yrs (1.12) and discounted card users (1.11). They are particularly concerned about the quality of dry facility and cleanliness. By contrast, older people (0.66 for retired and 0.67 for 64+ yrs) and the age group <16 yrs (0.66) are relatively under-represented in this cluster.

• Cluster 3: Physical evidence-wet & catering focused (29.0%). The age group 16-24 yrs (1.18), student (1.17), male (1.09) and full-time workers (1.08) are well-represented in this cluster. The catering service is particularly important to them, followed by the quality of wet facility and cleanliness. By contrast, this cluster constitutes relatively few older people (0.67 for retired and 0.70 for 64+ yrs).

• Cluster 4: Non-physical evidence focused (6.9%). A significant share of the customers in this cluster is younger people (1.84 for <16 yrs and 1.78 for 16-24 yrs), student (1.78), and other well represented groups are ethnic minority (1.56), club/team members (1.47), male (1.33) and full-time workers (1.11). This segment highlights particularly the non-physical evidence, but only contains a small proportion of customers. By contrast, it shows clear under-representation of older people (0.51 for 64+ yrs and 0.53 for retired) in this segment.

Generally speaking, the majority of older people are more concerned about the physical evidence, especially the quality of the wet facility, whilst younger people are more concerned about the non-physical evidence. By contrast, older people are less

concerned with non-physical evidence and the dry facility. On the other hand, the physical evidence is basically less important to younger people. This characteristic is also reflected by the occupation status, where retired people are dominant in cluster 1 and students are dominant in cluster 4, respectively. Disadvantaged groups, including retired/64+ yrs, female/housewife, unemployed and discounted card user, seem to be prevalent in cluster 1 and cluster 2, where much focus is put on the physical evidence. However, it appears that non-physical evidence and the dry facility are especially important to ethnic minorities.

7.2.3 Discriminant Analysis

Finally, discriminant analysis is performed to investigate which quality dimensions are driving the differences among the four clusters and to identify the classification accuracy of cluster analysis. The dependent variable in the discriminant analysis is the four clusters and the independent variables are factor scores on the six service dimensions. As shown in Table 7.11, three discriminant functions are yielded from the analysis. To test the significance of discriminant functions, the canonical correlation coefficients and the Wilk's λ statistics are examined. The values of Wilk's λ transformed as χ^2 indicate the significance of three functions (p<0.001). The inspection of the eigenvalues and canonical correlation coefficients demonstrates that the first two discriminate functions are relatively more powerful in differentiating the four clusters. Function 1 and 2 explain, respectively, 74.5% and 24.5% of the variance (eigenvalue= 3.49 and 1.14), as opposed to 1% by function 3 (eigenvalue= 0.05).

Table 7.11 Discriminant analysis for customer segmentation analysis

Panel A: Factors (standardised discriminant coefficients)

	Wet facility	Cleanliness	Dry facility	Catering	Accessibility	Staff
Function 1	0.47	0.40	0.58	0.26	0.30	0.22
Function 2	-0.19	-0.20	-0.30	0.29	0.01	0.87
Function 3	0.54	0.26	-0.69	0.17	-0.08	-0.09

Panel B: Functions

	Eigenvalue	% of variance	Canonical correlation	Wilks' λ	χ^2	*df*	Sig.
Function 1	3.49	74.50	0.88	0.10	8,443	18	0.000
Function 2	1.15	24.51	0.73	0.45	2,958	10	0.000
Function 3	0.05	1.00	0.21	0.96	166	4	0.000

In terms of the relative contributions of six service dimensions in discriminating the four clusters, the standardised discriminant coefficients are used to interpret the function. They represent the relative contribution of the associated factors to the discriminant functions. The discriminant coefficients for the first functions show that 'dry facility' best differentiates the clusters and the primary discriminator is 'staff' on the second function. It signifies that these two factors vary the most across clusters and are the best predictors of classifying the four customer segments. Finally, the classification matrix revealed that 82.0% of the cases are classified correctly (the value for cross-validated group cases amounts to 79.9%). The high accuracy rate also confirms the validity of four-cluster solution.

7.2.4 Discussion

Whilst the level of routine market research activity is increasing in the UK public leisure sector, market segmentation studies are very rare to date. Any attempt to better understand customers relies on generating high quality market research data and making good use of them. From an empirical perspective, the implications of this study are twofold.

First, it demonstrates that over 60% of customers are more concerned about the physical evidence and only about 7% of customers place relatively higher importance on the non-physical evidence. These findings tend to support the importance of physical evidence suggested by some sport facility studies (e.g. Afthinos *et al.*, 2005; Alexandris *et al.*, 2004; Greenwell *et al.*, 2002; Lentell, 2000). In the UK, Lentell (2000) investigated customer satisfaction at seven local authority-owned indoor leisure facilities during the period 1996 to 1997 and found that the physical evidence is considered to be the most important one to customers. He also recommended that in order to secure better customer satisfaction with public leisure services, improving the tangibles may be the most effective. In other words, according to the typologies proposed by Grönroos (1984) and Booms and Bitner (1981), the technical quality/physical evidence is more important than functional quality/process and participants to the majority of customers.

In the context of social inclusion policy advocated by the UK Government, the access of disadvantaged groups to public sport facilities has, for a long time, been a major concern of local authorities. The second empirical implication of this study is to show that it is inappropriate to treat disadvantaged groups as homogeneous, since different groups place different emphasis on specified service dimensions. Ideally,

specific and differentiated marketing strategies should be developed. Facilities aim to increase the access of older, female or unemployed customers need to realise that they may be more critical of the physical evidence. By contrast, facilities hoping to attract more adolescents or ethnic minorities should emphasise the non-physical evidence, since they tend to be less demanding of the physical evidence.

Although market segmentation is a valuable tool for marketing and understanding customers' needs, it is important to recognise that, in the public sector, "segmentation can not be used as a method for strategically ignoring any group, but instead as a method to allow more targeted communications with the diversity of the population" (Borrie *et al.*, 2002, p.55). Fountain (2001) also emphasised that public service delivery agencies need to serve a variety of target populations. Differentiation of service levels according to customer segments may place agencies on a slippery slope, leading easily to inequality. Some evidence has shown that public agencies tend to serve those clients who are easiest to serve and therefore sacrifice the needs of others. As such, to some extent, the role of market segmentation in the public sector slightly differs from which in the private sector.

CHAPTER 8

Results of Action Research

Road map of the thesis

Introduction	Theoretical context	Methodology	Results	Conclusion
Ch1 ▶	Ch2 Performance measurement overall Ch3 Efficiency & Equity Ch4 Customer service	▶ Ch5 ▶	Ch6 Efficiency & Equity Ch7 Customer service Ch8 Action research	▶ Ch9

Modelling is usually a means of securing a more satisfactory aggregate model; however, managerial interest is in the estimate of performance for individual organisations (Smith and Street, 2005). As such, careful attention should be paid to how the model's results can be properly interpreted and implemented. This chapter aims to discuss the practicability of aggregate performance analysis at the individual facility level. The action research begins with developing, from theory, an aggregate performance measurement model and applies this model in three public sport/leisure sport centres over a two year period. The findings presented here are from three sources: (1) the focus group discussions in the workshops while introducing the model to the three centres; (2) the post process semi-structured interviews with the management teams of the three centres; and (3) the author's own reflection during the research process.

The following sub-sections start by describing the research phases before going on to discuss the process and outcomes of action research. In the first place, it is necessary to explain the strategy used to analyse the research findings. Traditionally, either 'within-case' and/or 'cross-case' analysis are used to analyse the results of case study research (Yin, 1994). Within-case analysis entails becoming familiar with each case individually and documenting it thoroughly. In cross-case analysis, similarities and differences across cases are explored. Both approaches will be adopted in the following analyses.

In Section 8.2, the process of introducing the aggregate analysis model is first presented by adopting within-case analysis. The analysis is constructed on the basis of the five-stage action research cyclic process (as mentioned in Section 5.3). The aim is to provide an in-depth analysis of the managers' attitudes toward the proposed model and the problems encountered in each centre. Once the data displays for each case were completed, the search for cross-case patterns could begin.

In Section 8.3 and 8.4, the focus turns to the usefulness and applicability of the proposed model. Research questions were used as categories to look for cross-case similarities and differences. The model's usefulness is evaluated by examining the constituent elements of each model. Since the model was evolving and parts of the model's elements were changing throughout the process, it is inappropriate to compare the model's overall usefulness across cases. The only possibility is to assess the usefulness of each constituent element. In addition, the aim of the model's evaluation stage is to reflect on which elements of the model are generally applicable and therefore suitable to be included in the final model. It is argued that the usefulness

of individual elements is more appropriate than the overall usefulness of a model which is under development.

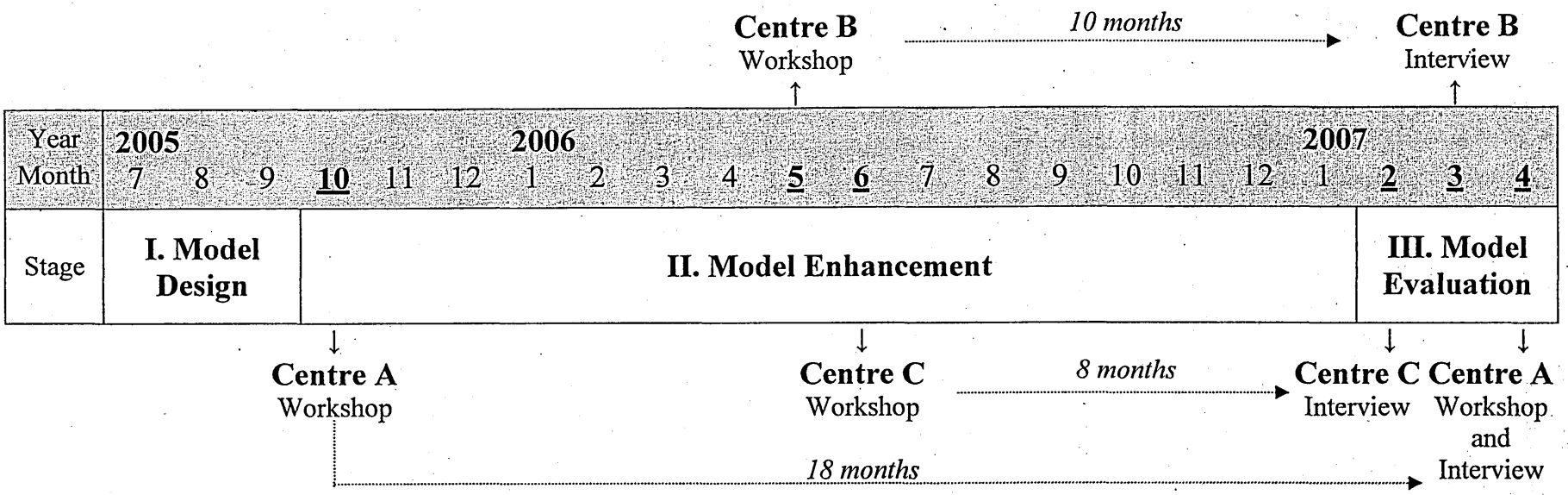
In order to evaluate the model's overall applicability, the managers were also asked to discuss the model as a whole during the post process interviews. The determinants of success of the model's applicability can be grouped into two different categories. The first set concerns the model design itself, while the second set concerns the implementation process, such as the actual use of the model. However, it should be admitted that it is difficult to generalise the findings because the model was evolving and the feedback from each centre was not based on identical models.

Finally, action research should be judged by both its practical implications and its ability to add to the stock of knowledge of the research community. As such, the implications of action research for academic theories will be discussed alongside the analysis of research findings.

8.1 Research Phases

Following the cyclic principle of action research, this research is divided into three phases in order to design, enhance and evaluate the model. Figure 8.1 illustrates the research time scale. The progress of each stage is summarised as follows.

Figure 8.1 Time scale of model's evolution



• ***Model design***, lasting 3 months (from July to September 2005), aimed to establish an applicable performance measurement model. A model was designed on the basis of academic theories, which could subsequently be tested through live application. The process began by reviewing the literature and theories related to performance measurement and benchmarking, specifically focusing on the issue of DEA, aggregate performance analysis and process benchmarking. The information gleaned from this review, together with the DEA analysis of NBS data for each centre, is structured in the form of an individual facility report designed to implement the developed model in selected public sport centres.

• ***Model enhancement***, lasting 16 months (from October 2005 to January 2007), aimed to test the developed model in three centres with different management types. Given the fact that the initial model is based on academic theory but had yet to be applied, the validity of the model was an open question. The model was therefore first tested in a pilot centre (Centre A). During this application, many shortcomings of the model were identified, both by the members of centre visited and by the researcher. Then, the revised model was introduced into the other two centres (Centre B and Centre C) in order to test whether the model can be applied in different settings. It then took another 10 months (for Centre B) and 8 months (for Centre C) before the model reached the stage of being reviewed. The characteristics of the three centres and the respondents involved in the research are shown in Table 8.1. In this phase, the elements of the model continually evolved and improved according to the feedback from each centre. The main changes are summarised in Table 8.2, which will be discussed in detail in the next section.

Table 8.1 Centres and members involved in the research

Centre	Management type	Management team involved
Centre A	Trust	Chief Executive and General Manager
Centre B	In-house	Project Officer and Deputy Manager
Centre C	Commercial contract	Contract Manager

• ***Model evaluation***, lasting 3 months (from February to April 2007), consists of two sub-stages. First, semi-structured interviews were planned to investigate the practicability of model in Centre B and Centre C. Second, a revised model, on the basis of the feedback from the three centres, was constructed and Centre A was revisited. The aims of revisiting Centre A were threefold: (1) examine their actual use of the pilot model; (2) investigate whether the revised model was more acceptable than the pilot model; and (3) test the acceptability of the revised model. Upon completion of the model evaluation phase a final model was developed, which can be applied by other centres. In addition, the experiences of three centres in using the model and the barriers encountered can provide useful reference for other centres who may adopt the model, and for development of theory of decision-making in public sport facilities.

In all three centres, the fact that the model is under development was made explicit to the management team at the outset of visit and members of the teams were continually encouraged to comment on model's practicability. For Centre B and Centre C, one workshop was held at the beginning of the intervention and

semi-structured interviews at the end of the process. For Centre A, two workshops were conducted- in the beginning of the 'model enhancement' stage and at the end of 'model evaluation' stage. While revisiting Centre A the second time, an interview was also conducted immediately after the workshop to investigate managers' first impressions of the revised model. Evaluation of the actual use of the model is unlikely at this stage of the research, but the feedback from the three centres allows the author to formalise the model and critically evaluate the practicability of aggregate performance analysis. Each workshop presented the model and each constituent part of the model was critically appraised. After each workshop, the author reflected on the acceptability of model and in turn triggered the modifications of the model's constituent elements. In the following section, concentration is given to how the model was modified as well as the reasons for and the implications of these modifications.

8.2 Model's Evolution

It is argued that the practicability of aggregate performance analysis is not only determined by the model's design, but also the interactions and communications between the researcher and the managers. As such, the process of introducing the proposed model into the three centres will be first presented in this section.

As shown in Table 8.2, the model consists of three independent dimensions (i.e. efficiency, equity and customer service measurement) and eight constituent elements. Five of the eight elements are suggested to be included in the final model and an illustrative example can be found in Appendix B. The rationale of this decision will be

justified in Section 8.3. Here, the focus is on the action research process in each centre and how the model was evolved. The evidence reported emerged from the focus group discussions and the post process interviews.

Table 8.2 Elements included in the report for each centre

Dimensions and elements	Centre A (Pilot)	Centre B	Centre C	Centre A (Revised)	Final model
Efficiency and Equity					
1. DEA- performance score and target	▲	▲	▲		
2. DEA- strengths and weaknesses	▲	▲	▲	▲	▲
3. DEA- benchmarking partners	▲	▲	▲	▲	▲
4. AHP- Subjective weighting method			▲		▲
Customer Service					
5. Applying DEA to measure service quality	▲				
6. Grid and gap analysis diagram		▲	▲	▲	▲
7. Applying factor analysis to measure service quality				▲	
8. Customer segmentation analysis				▲	▲

Notes: ▲ signifies the elements included in the report for each centre.

8.2.1 Centre A (Pilot)

Centre A (managed by a trust) was chosen as the pilot centre as it had just received the NBS report and had a feedback session presented by Professor Peter Taylor at the outset of this research. During the feedback session held in the middle of July 2005, the Chief Executive criticised several aspects of NBS and presented his willingness to see how the current model can be improved. It therefore stimulated the author to seek a way to improve the current NBS.

According to the literature, the author constructed a standard DEA report which consisted of four parts of analysis: overall performance score, performance targets, strength and weakness clarification and benchmarking partners' identification. Before presenting the analyses, the basic concept of DEA was also concisely explained associated with a comparison of DEA and ratio analysis. The pilot report was sent to the manager two weeks before the visit. The focus group discussion was attended by the Chief Executive and General Manager.

During the focus group discussion, held on 21 October 2005 (as shown in Figure 8.1), some merits of the proposed model were appreciated by the managers, such as the function of DEA in identifying strengths and weaknesses. The General Manager stated that: "The headline information of centre's strength and weakness forces us to look at the issues that are important to the centre". The Chief Executive also criticised the design of the NBS report and suggested providing a summary table illustrating the centre's strengths and weaknesses such as the proposed model did. Another useful part of the report is the identification of benchmarking partners. The Chief Executive stated that: "benchmarking partners with similar background are better than figures (i.e. benchmarks)".

However, several deficiencies of the proposed model were also identified during the discussion. The major criticisms on DEA are its complexity and inappropriateness in measuring service quality. In terms of the complexity of DEA, the General Manager remarked: "if you are here presenting it, we have opportunities to ask questions. If you just give me the report, I will struggle." The General Manager also doubted the appropriateness of applying DEA to measure service quality. He emphasised that customers' comments on individual attributes are more important to the managers, and implied that "the result is concise but not necessarily to be the right information". The Chief Executive further suggested that "the report should be readable and digestible". He said: "it is not necessary to show all the survey results ... a concise report showing the most important information is more likely to be remembered". From his point of view, in the report, some places are too complicated (e.g. the concept of DEA) but some places are not detailed enough (e.g. service quality measurement).

In line with the feedback from Centre A, it was decided that the content of the report should be simplified to make the model more accessible to the managers and the adoption of DEA to measure service quality should be substituted by other techniques. In terms of data collection, when the focus group discussion was initially planned, it was expected that the discussion would focus exclusively on the DEA model. However, the managers were more interested in discussing the current NBS report. Around 50% of the discussion time was focused on the topics irrelevant to this research, such as the definition of catchment area. Consequently, it was decided that the discussion should adopt a semi-structured approach so that it encourages the management team to focus on answering specific research questions.

8.2.2 Centre B

The second action research cycle was started by reviewing the literature to seek out a better approach to present the customer survey data. Grid analysis and visualised gap analysis were found to be the most common ways used to measure service quality. These two approaches were therefore used to develop the report for Centre B. Some of the text in the report was also replaced by figures or tables to ease reading.

The workshop and focus group discussions in Centre B were held on 3 May 2006 and attended by the Project Officer from the Council and the Deputy Manager of the centre. The attendants appreciated especially the grid analysis which provides a snapshot of the centre's overall service quality performance. The Project Officer highlighted that: "it is a good example where joint use of visual representation and brief text allows managers to know good and bad areas immediately". The identification of benchmarking partners was another benefit perceived by the manager. The Deputy Manager said: "similarity is important... we can mimic what they are doing and identify why we are underperforming." The managers also expressed their willingness to contact other better performers provided in the report.

However, the concept of DEA itself was again criticised as too complicated. The Project Officer stated that the presentation made the model clearer, but it was complicated when she went back to the report alone. From her point of view, "the customer part can be understood, but DEA is difficult... The outcome of DEA is understandable, but how to get that is too complicated". To facilitate understanding, the Project Officer recommended a meeting with the analysts. She said: "a meeting with you allows us to ask questions and clarify suspicious points."

Reporting too much information was also identified as an issue. This was particularly the case when the managers hand the report to other staff. The Deputy Manager took the report of QUEST as an example. He remarked: "the report is more than one hundred pages, but we only used the first four pages... Too many reports, we don't have enough time to analyse and interpret the report". He suggested that "the content of the report should be easily accessible to different level of staff and departments ... it is better to provide a summary of findings in the first few pages and reduce the quantity of text".

The post-process evaluation revealed that Centre B provided an example of using the proposed model. The information within the report was used to produce the action plan in July 2006. According to the Deputy Manager, it was done by "listing out the areas capable to do and taking the NBS and the new model simultaneously to address the weaknesses". He said: "we know our strengths and weaknesses. The report reinforced our thoughts and a discussion allowed to double check the evidence... The report shows me an overall picture. Together with the details of NBS, we can have a clearer view of where we are". The Project Officer further remarked: "the new model provided a holistic overview of the organisation's performance and the NBS report provides more detailed information. A joint use of two models has complementary benefits".

The Deputy Manager also indicated that the cooperation between council and facility manager was the determinant of success in using the model. He remarked: "effective use of the measurement system is due to her (the Project Officer) promotion of the importance of this report". Although Centre B provides an example of using the proposed model to produce the action plan, no action was taken regarding contact

with the benchmarking partners. This fact was an unexpected result since the manager expressed his intention to conduct process benchmarking during the workshop. The Project Officer confessed that "we intended to contact them, but time passed by... Other more important issues occupied our time".

8.2.3 Centre C

Following the second application, the model was again revised. The major change was to adopt Analytic Hierarchy Process (AHP) to complement the deficiency of DEA. This initiative was not derived from the feedback from previous centres, but from a further literature review conducted by the author. As illustrated in Figure 8.3, the Contract Manager of Centre C was asked to make a pair wise comparison between each variable to generate subjective weights reflecting the priority of facility manager. The weights were then used to run the DEA model so as to overcome the problem of weighting flexibility. Other parts of the report remained basically unchanged as a result of previous application. Only minor changes were made to simplify some wordings.

The workshop in Centre C was held on 14 June 2006 and was attended by the Contract Manager and other five Assistant Managers from different departments. The headline information of the centre's strengths and weaknesses provided by DEA is once again regarded as an important function. The Contract Manager stated: "It gives me a snapshot view of our performance ... A sharp line separating the strength and weakness results in a more focused organisation". The post-process interview with the Contract Manager also showed that the joint use of AHP and DEA provided a medium that could be used to communicate with the local authority. He remarked:

"The weighting system (i.e. AHP) is the most useful part in the model ...We are strong in financial performance, but cash is not relevant to the council ... A balance can be struck between us when I show them the results"

However, differing from Centre A and Centre B, the Contract Manager was sceptical of the performance targets suggested by DEA, which were significantly higher than the CPA higher threshold and the 75% NBS benchmarks. He was confident of the centre's financial performance and expressed his doubt about the DEA results and implied that the DEA targets were unrealistic to him.

The Contract Manager also criticised that there was too much jargon in the report (e.g. virtual weights, radial movement etc.). He said: "It is easy for people to stop reading because they will find that the document is not relevant to them ... Ensuring that the report reflects the issues that are important to us is important if measurement is to be useful and help management". He further suggested to simplify the technical concepts and make the report a more friendly and accessible document. He said: "it is important that the report can be skimmed by a manager within a short period of time". A meeting with the researcher to explain the report was suggested as the manager can "question the findings" (Contract Manager).

In terms of the actual use of the proposed model, the manager hasn't either used the report in a formal way or contacted with the benchmarking partners. The Contract Manager proposed that: "the report gives us some points... However, this kind of analysis is probably more important to people in a higher level, such as councillor or director". He also explained the reason why no attempt has been made to contact with the benchmarking partners: "it is not because of unimportance, but because of priority".

8.2.4 Centre A (Revisit)

After testing the model in three centres, it was believed that the model's general acceptability had increased significantly so it was worth revisiting Centre A to investigate whether the revised model was more acceptable than the pilot model. The revisit to the Centre was on 26 April 2007 and the workshop and interview were attended by the same management group as the first visit.

As shown in Table 8.2, one significant modification was made, i.e. applying MSA to the analyses of customer service quality and customer segmentation. This initiative derived both from the analysis at the industrial level and Centre A's previous criticism on the inappropriateness of using DEA to measure service quality. The motivation was to investigate whether MSA can further enhance the proposed model. The author also intended to apply AHP to generate subjective weights to be included in the DEA model. However, the major barrier encountered while revisiting Centre A was the reluctance of the Chief Executive to respond to the emails and to set a date for revisiting. Consequently, it was necessary to exclude AHP from the report.

During interviews after the workshop, the burden of considering simultaneously two models and the availability of management time were identified by the Chief Executive as the main reason for slow response. He commented: "There were two real problems with implementing your model. The first was getting the management to feel happy with it. It is easy to feel threatened by them and they needed to be persuaded that it was good for the business and not a threatening initiative... (Secondly), we are bombarded by too many reports, we do not have enough time to put everything into action". The management team therefore failed to use the pilot

model, in favour of the traditional NBS report. By contrast, the acceptability of the revised model is much higher than the pilot model from the Chief Executive's point of view. He commented: "compared with the previous one, the new one is more readable and digestible... There are too many figures and a lot of cross-references are required when reading the previous report".

With respect to the constituent elements of the revised model, similar to Centre B, grid analysis was found to be the most useful part in the report. However, the suitability of applying DEA was again criticised by the manager: "even though the outcome of DEA is succinct, the process is not transparent, which makes the results less convincing" (General Manager). He further recommended that "such an aggregate analysis may be more useful at the strategic level but a facility manager needs more detailed information". The acceptability of using MSA to measure service quality was also low. The General Manager criticised that: "the result may be distorted by aggregating various attributes ... certain attributes are very individual".

Although based on the same approach, customer segmentation analysis was appreciated by the managers. The General Manager expressed that segmentation helps to break down the customers, especially those target groups government are prioritising. However, the manager didn't illustrate how he will action this segmentation evidence (e.g. promotion targeting). A post-process reflection of the author found that the outcomes of this segmentation analysis are somewhat difficult to be operationalised. The manager did not challenge this part of analysis probably because customer segmentation does not exist in the original NBS report, and therefore it is hard to judge the validity of analysis.

8.2.5 Concluding Remarks

After displaying the findings in each case, this section will search for cross-case patterns by discussing the similarities and differences across cases. Table 8.3 summarises a cross-case comparison in eleven issues which have been raised in the aforementioned within-case analysis.

Table 8.3 Cross-case comparison

	Centre A	Centre B	Centre C
Pros			
- Strengths and weaknesses clarification	▲	▲	▲
- Grid analysis for service quality measurement	▲	▲	
- Benchmarking partners identification	▲	▲	
- Adopting AHP to enhance DEA			▲
- Customer segmentation analysis	▲		
Cons			
- Complexity of DEA	▲	▲	▲
- DEA performance targets			▲
- Aggregate service quality measurement	▲		
Recommendations			
- Simplify the model	▲	▲	▲
- Meeting with the analyst	▲	▲	▲
- More suitable at the strategic level	▲		▲

Notes: ▲ signifies the issues raised by the managers.

These eleven issues can be grouped into the following three dimensions:

• ***The pros of aggregate analysis***

The summary information of strengths and weaknesses was found, across three centres, as the most useful part of the model. For Centre A and Centre C, the benefit of separating sharply the strengths and weaknesses allows managers to become more focused. For Centre B, the headline information of centre's strengths and weaknesses has the biggest effect in complementing the NBS. Applying grid analysis to measure service quality is the other initiative appreciated by the managers. Similar to the function of the clarification of strengths and weaknesses, applying grid analysis to measure service quality allows the managers of Centre A and Centre B to know good and bad areas immediately. The third function appreciated by the managers is the identification of benchmarking partners. For Centre A and Centre B, providing models with similar backgrounds to follow is a benefit.

• ***The cons of aggregate analysis***

Although aggregate analysis has several merits, it still has three major shortcomings according to the managers. First, and probably the most important barrier to apply DEA is its complexity. Both for Centre A and Centre B, the outcome of DEA is acceptable, but the process is too complicated and therefore undermines the potential value of DEA. The managers of Centre A and Centre C also felt threatened by the technical concepts or jargons of DEA. Except for the process of DEA, its outcome was also questioned by Centre C in terms of target setting. Finally, for Centre A, either DEA or MSA is unsuitable for assessing service quality.

• ***Managers' recommendations***

Three recommendations were made by the managers to facilitate the use of the aggregate model. First, due to the complexity of aggregate analysis, the technical concepts should simplified (Centre C), the report content should be condensed (Centre A and Centre B) and it is useful to provide an executive summary at the front of the report (Centre B). The second suggestion, made across the three centres, is to have a meeting with the analysts to question the findings and clarify suspicious points. In this research, the author played a proactive role in introducing the new model; however, it is not feasible to have a meeting with each centre in any future service. Consequently, if it is difficult for managers to interpret the report individually, the feasibility of aggregate performance analysis will be seriously constrained. Finally, both Centre A and Centre C proposed that such an aggregate model may be more useful at a higher strategic level of decision makers. This suggestion basically coincides with the argument of Smith and Street (2005). From a strategic regulator's point of view, global measures may be useful. For instance, it might identify beacons of good practice. However, at the individual facility level, it is more important to examine specific organisational functions rather than be given a summary measure of overall performance, i.e. much more detailed benchmarking data is needed for local managerial purposes

8.3 Model's Usefulness

After illustrating in detail the action research process in each case, the focus of this section is to evaluate the usefulness of specific elements in the model. The eight elements shown in Table 8.2 are discussed in turn.

8.3.1 Efficiency and Equity Measurement

• ***DEA: Performance Score and Target***

One characteristic of DEA is to integrate multiple inputs and outputs to yield a single indicator presenting the overall performance of a centre. The score, ranging from 0 to 1, represents the degree of inefficiency of the centre under evaluation. The performance targets are then given based on the performance of the best practice(s) within the sample, i.e. those on the DEA frontier (see Chapter 3 for details). An example is given in Table 8.4, which was included in Centre A's pilot report but also appeared in the report for the other two centres. These results reveal three problems according to the feedback of the managers and the author's reflection.

Table 8.4 DEA performance score and target

	Inputs			Outputs	
	Cost	Open Hour	Area	Income	Visits
Original value	620,023	5,147	2,307	633,410	277,909
Performance targets to 100% efficiency	620,023	5,147	2,307	796,324	349,388
% of improvement required	0%	0%	0%	+26%	+26%

First, a single score representing the centre's overall performance provides no operational meaning for the facility managers. This reflects the argument of Smith and Street (2005)- that is, for a local manager, it may be more important to examine specific organisational functions rather than be given a summary measure of overall organisational performance.

A second issue raised is the inappropriateness of DEA performance targets. This issue, relating to the principle of 'radial movement', has been discussed in previous literature and could be regarded as a theoretical deficiency of DEA (Bell and Morey, 1994)(refer to p.40 for a detailed description). The '% of improvement required' in Table 8.4 provides a good example.

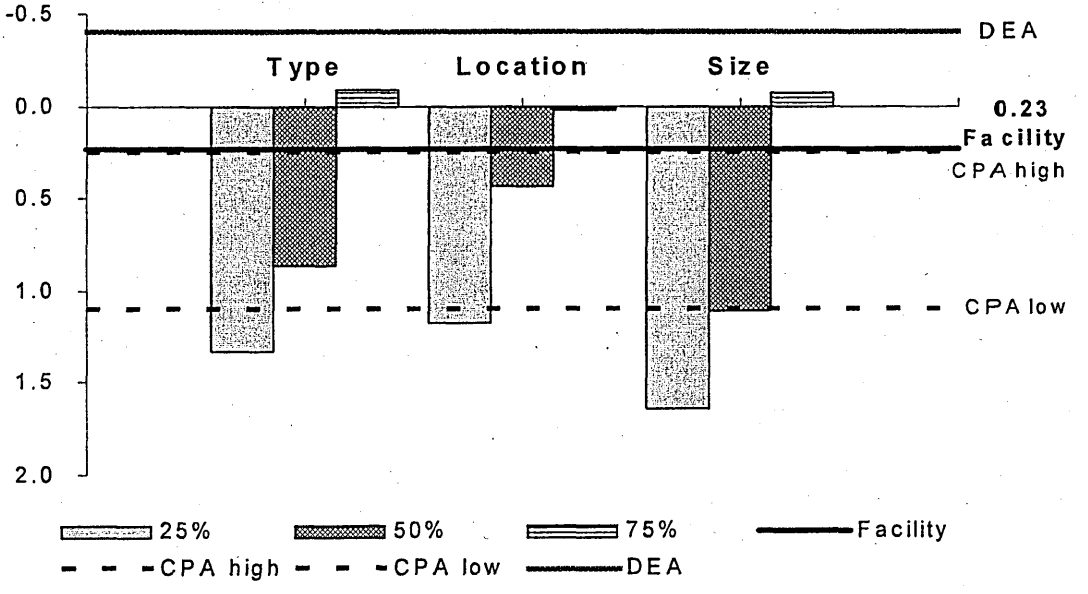
Keeping the level of cost, opening hours and facility area constant, DEA suggests to increase, simultaneously, Centre A's income and visits by 26% to become efficient. Although this point has not been mentioned by the managers, the author noticed this pitfall while designing the facility report. The users of DEA may find this kind of suggestion for improvement is inflexible or unrealistic.

The third problem also relates to the DEA targets but is more relevant to the nature of public service. The target suggested by DEA is to become the best, but it may not be the aim of most public service providers. According to Ball *et al.* (2000), private sector benchmarking has best practice as its primary focus. In the private sector, benchmarking is often undertaken in order to be the best. In contrast public sector organisations may strive, through benchmarking, to be 'good enough' rather than the best. In the absence of market competition, they might want to use benchmarking to target services that are below standard aiming to bring performance up to an average level. This assumption is reflected in the case of Centre C. In order to

examine the practicability of performance targets suggested by DEA, a diagram comparing the NBS benchmarks, the CPA thresholds and the DEA targets in terms of 'subsidy per visit' was included in Centre C's report (as shown in Figure 8.2).

The performance of Centre C lies in between the 50% and 75% NBS benchmarks and is slightly higher than the CPA higher threshold, but is much lower than the DEA targets. The Contract Manager was confident with the centre's financial performance and the NBS benchmarking results also supported his argument. He thus expressed his doubt about the DEA results and implied that the DEA targets are unrealistic.

Figure 8.2 Performance targets: DEA, NBS and CPA



• ***DEA: Identifying Strength and Weakness***

As mentioned in Chapter 3, another function of DEA is to identify the strengths and weaknesses of a centre. It is done by examining the 'virtual weight' of each

input/output variable (Despic, 2004). Throughout the research process, this function was found to be the most useful part for the managers, and was the most frequently cited benefit of DEA. Taking Table 8.5 as an example, the summary information of centre's strength and weakness forces the managers to "look at the issues that are important to the centre" (General Manager of Centre A) and results in "a more focused organisation" (Contract Manager of Centre C).

Table 8.5 Strengths and weaknesses identification

	Strengths	Weaknesses
1. Efficiency	- Income	- Cost - Visits
2. Equity	- 11-19 years - 60+ years - Disabled	- Social class DE - Ethnic minorities

Another issue that was not raised by the managers but was reflected by the author while designing the model is the usefulness of DEA in equity measurement. The major function of DEA is for efficiency measurement. This research tried to consider DEA in a broader context by developing a DEA effectiveness model for the purpose of equity measurement. However, the results of DEA virtual weight analysis are basically consistent with the NBS benchmarking results, so there is no reason for the managers to adopt a more complicated model. This problem has been raised by Chen and Ali (2002) who argued that if the consistency between ratio analysis and DEA is high, the computation effort of running DEA is less necessary. In fact, the function of DEA in identifying strengths and weaknesses is more important in an efficiency model rather than in effectiveness model. This is because contradictory

information usually exists between different efficiency ratios, as they link the interaction between different sets of inputs and outputs. Efficiency can be achieved through either output maximisation or input minimisation. As for effectiveness, the measurement is more straightforward since output maximisation is the only target. Consequently, the identification of strengths and weaknesses in effectiveness PIs can simply examine their values relevant to the national benchmarks.

• ***DEA: Selecting Benchmarking Partners***

Apart from identifying strengths and weaknesses, the second function of virtual weights is to demonstrate the input/output orientation of the inefficient centre; and then select benchmarking partners which have similar input/output orientation as the inefficient centre but appear efficient (Boussofiance, 1991). It is also argued that the comparison of virtual weights need not refer to the DEA concept and so they may be suitable for the lay-person (Staat and Maik, 2000). An example of benchmarking partner selection is shown in Table 8.6. The eight benchmarking partners identified all have similar strengths and weaknesses as Centre A but appear effective. The facility profiles of these centres are also shown in the right hand side of the table to facilitate the selection of the most ideal benchmarking partners.

The model design in this part aims to facilitate the shift from 'data benchmarking' to 'process benchmarking' (as discussed in Chapter 2). However, although there is a general consensus among the managers interviewed of the importance of process benchmarking, none of them has taken steps to contact the benchmarking partners provided by the model. Although some managers expressed their willingness to contact other centres, all of the interviewees admitted that the

priority of process benchmarking is relatively low. It appears that there are substantial barriers to move beyond data benchmarking to process benchmarking. This result also echoes the study of Ogden and Wilson (2000) who found that it was rare to find cases where a full range of benchmarking activities has taken place and data benchmarking is still prevailing in the UK public sector.

- ***AHP: Subjective Weighting Method***

As mentioned in Chapter 3, weighting flexibility of DEA has been criticised by many researchers and could be regarded as a theoretical deficiency of DEA. The conventional DEA allows weights to vary freely so that each organisation is evaluated in the best possible light, but the weights assigned by DEA do not necessarily reflect the priorities of management. Following the proposal of Zhu (2002), the author attempted to introduce another tool - Analytic Hierarchy Process (AHP), which can be used to obtain subjective weights to be included in the DEA model. The advantage is to let facility managers make decisions about how weights should be assigned to each variable while benchmarking with other centres. An example of AHP's application is illustrated in Figure 8.3, where the manager of Centre C was asked to make a pair wise comparison between each variable, and then the subjective weights were calculated.

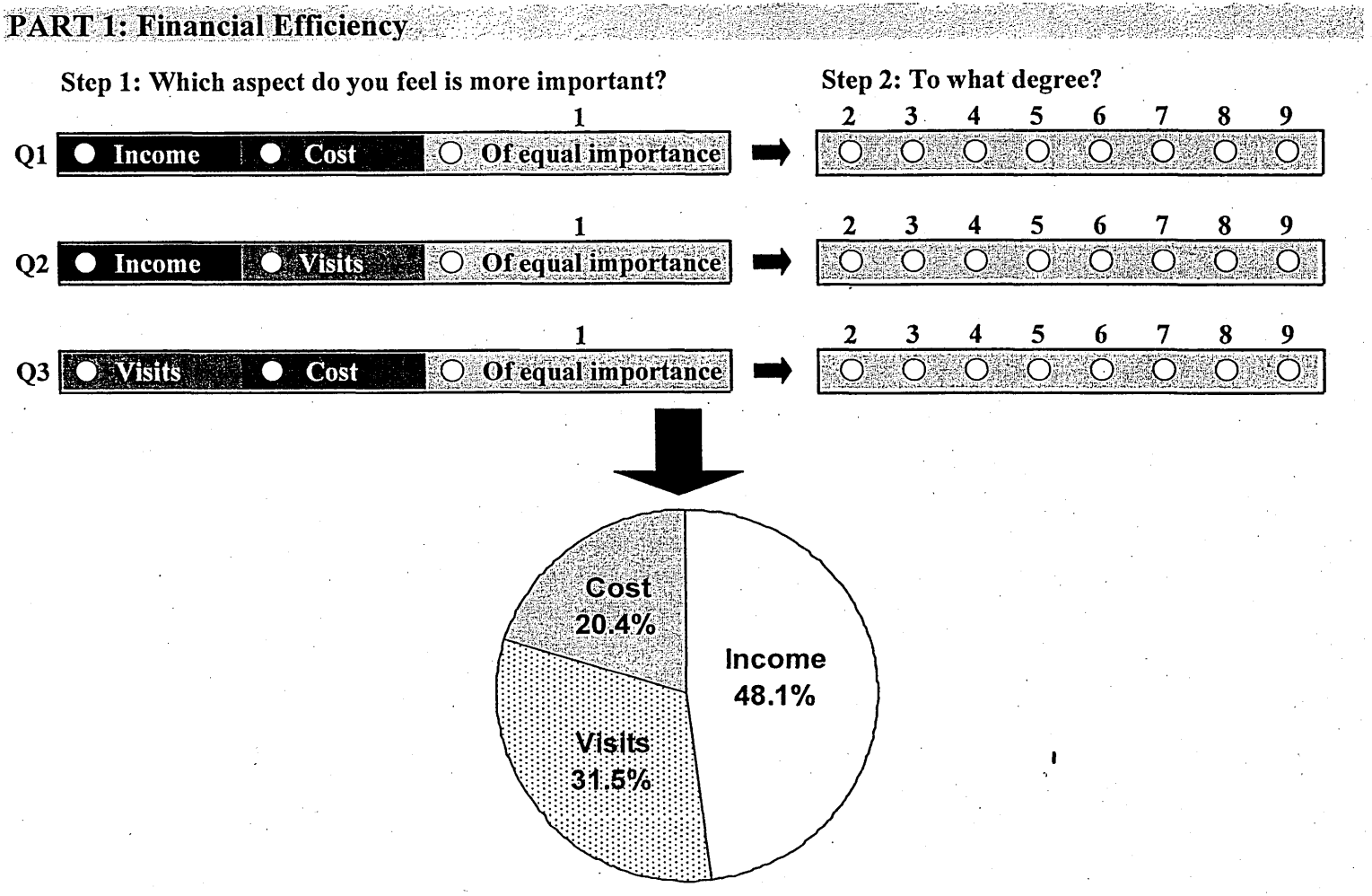
The post process interview with the Contract Manager of Centre C showed that the weighting system (i.e. AHP) is the most useful part in the model. It is therefore argued that if the management's priorities and DEA results differ, the joint use of AHP and DEA may compensate the theoretical deficiency of DEA in weighting flexibility.

Table 8.6 Selection of benchmarking partners

Centre	PIs					Family			
	Social class DE	11-19 years	Ethnic minority	60+ years	Disabled <60 years	Type	Location (DE)	Size	Management
Centre A	**0.20 (2^{nd})**	**0.60 (2^{nd})**	**2.60 (3^{rd})**	**0.50 (3^{rd})**	**3.30 (1^{st})**	**Dry**	**15-20%**	**Medium**	**Trust**
Centre 1	0.37	1.14	2.89	0.70	6.57	Wet	**15-20%**	**Medium**	**Trust**
Centre 2	0.32	0.96	4.77	0.76	4.72	Mixed	**15-20%**	**Medium**	Other
Centre 3	0.28	0.55	3.01	0.64	9.57	Mixed	**15-20%**	**Medium**	Commercial
Centre 4	0.39	0.93	8.38	~~0.32~~	5.99	Wet	**15-20%**	**Medium**	Commercial
Centre 5	0.30	1.54	~~0.21~~	0.63	3.69	Mixed	**15-20%**	**Medium**	In-house
Centre 6	0.32	1.83	~~2.53~~	~~0.08~~	8.23	**Dry**	**15-20%**	**Medium**	Other
Centre 7	0.40	~~0.54~~	~~0.84~~	0.72	9.39	Wet	**15-20%**	Large	**Trust**
Centre 8	0.47	~~0.45~~	~~2.53~~	0.54	7.51	Wet	20%+	**Medium**	**Trust**

Notes: Number with delete line means that Centre A has a better performance than its benchmarking partners;
Number with bracket is quartile adopted by the NBS (4^{th} is the highest and 1^{st} is the lowest);
Family in bold is the same as Centre A.

Figure 8.3 AHP survey and result



8.3.2 Customer Service Measurement

• *Applying DEA to Measure Service Quality*

Similar to the measurement of equity, a DEA effectiveness model was included in the pilot report for Centre A to measure customer service quality. Average attributes scores of each NBS dimension (i.e. accessibility, availability, facility quality, cleanliness, staff and value for money) were calculated to represent the performance of that dimension. As presented in Chapter 2, the limitation on the number of variables to be included in DEA is the reason for aggregating attributes to dimensions. If the original 28 NBS attributes are used, DEA may report lots of centres as efficient and lead to little discrimination (Thanassoulis *et al.*, 1996; Boussofiance, 1991). This aggregating process makes the DEA result in service quality less convincing both to the managers and to the author.

• *Grid Analysis and Gap Analysis Diagram*

Considering the inappropriateness of DEA in measuring service quality, it was replaced by grid analysis and gap analysis from the second action research cycle (the report for Centre B). Grid analysis and gap analysis diagrams (as shown in Figure 8.4 and 8.5) are alternative ways to represent the NBS results rather than aggregate analysis approaches. This modification was also stimulated by the focus group discussion in Centre A where the managers recommended visualising the text and figures so as to facilitate data interpretation.

Figure 8.4 Grid analysis diagram

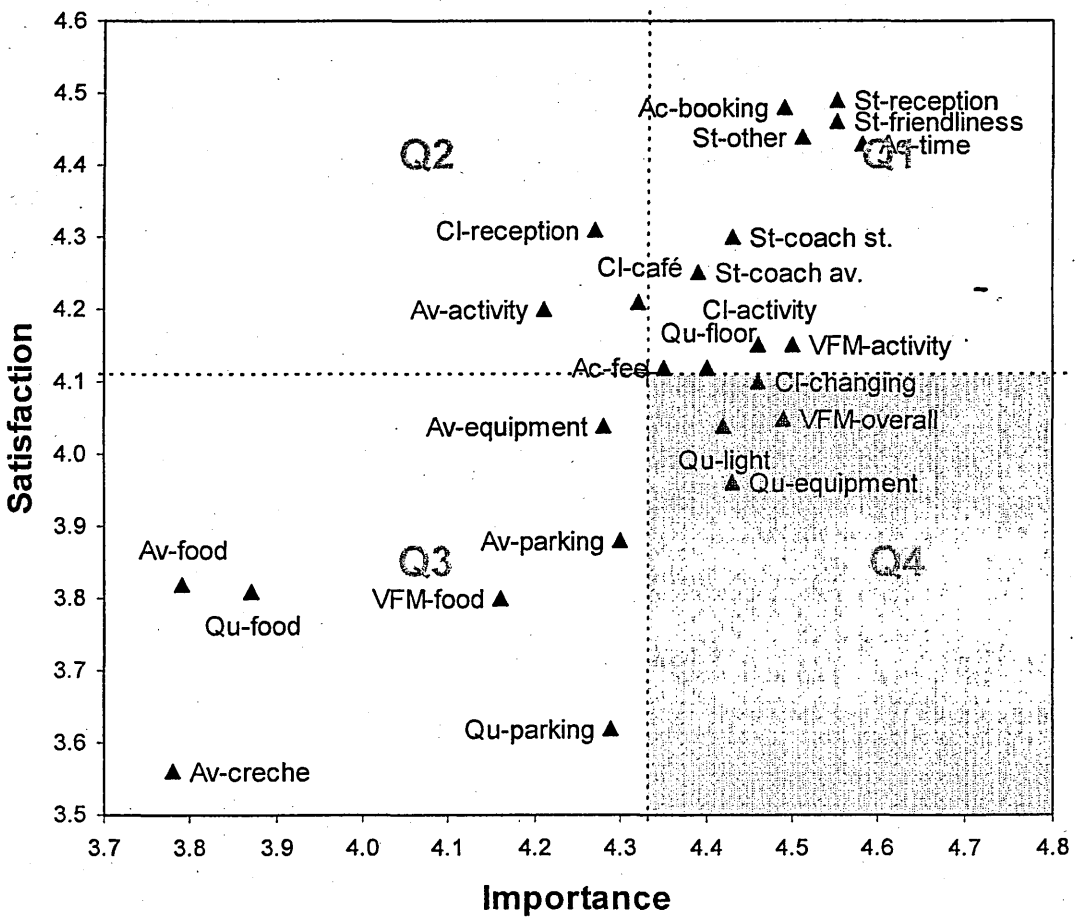
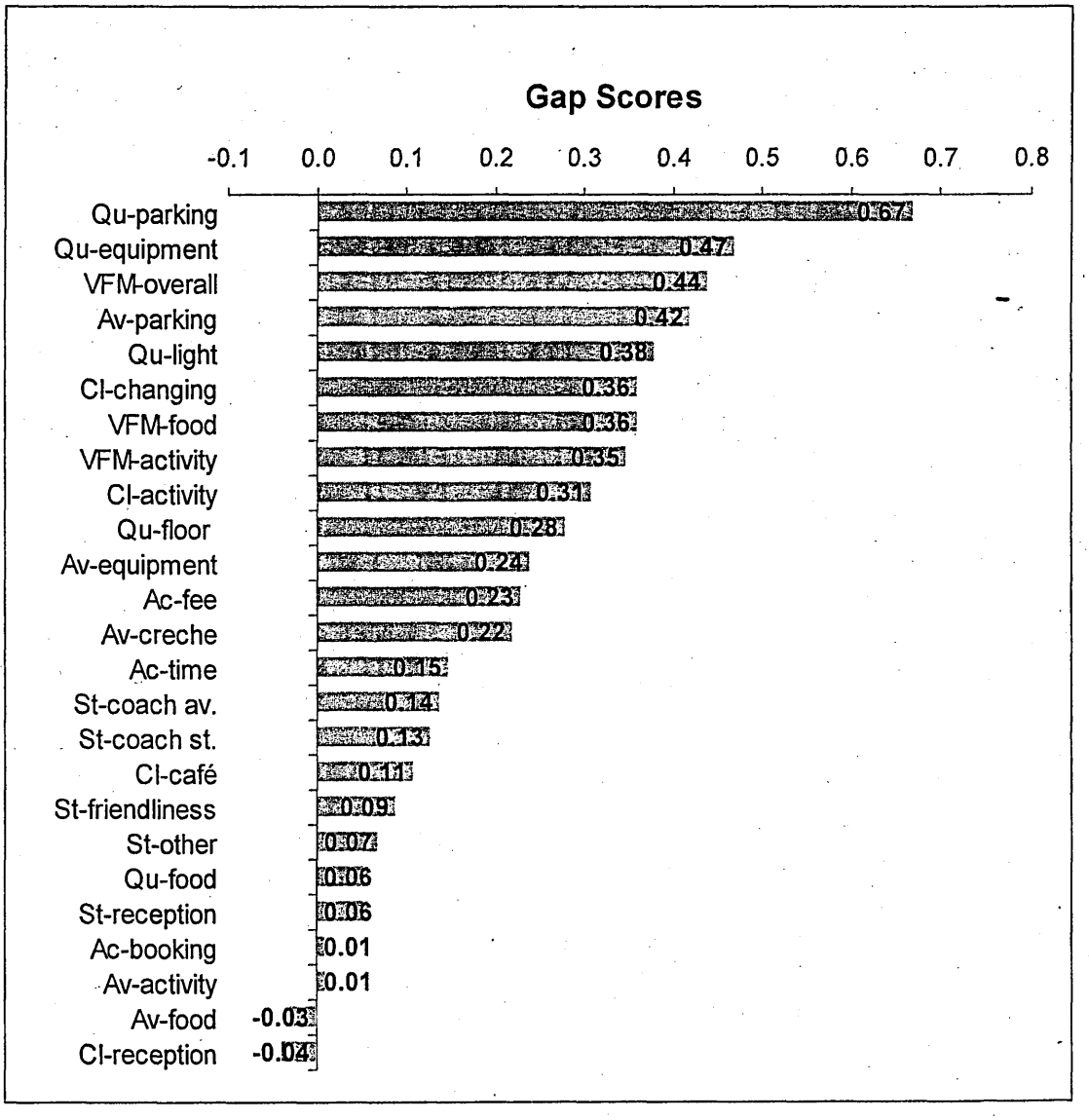


Figure 8.5 Gap analysis diagram



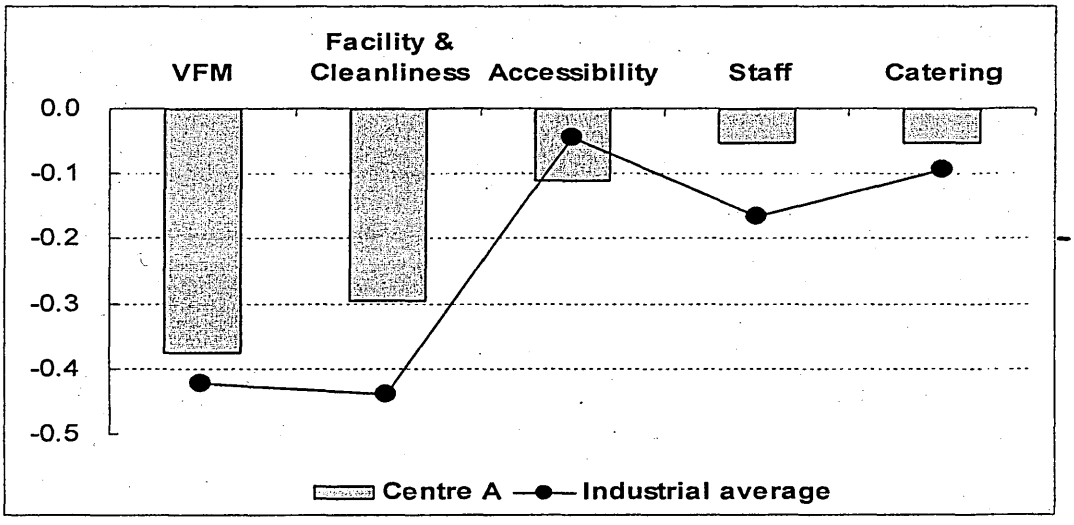
More details of grid analysis and gap analysis are provided in Chapter 4. Roughly speaking, gap analysis is conducted to identify the absolute gap of each attribute. Alternatively, grid analysis presents the original value of satisfaction and importance ratings in a two-dimensional grid rather than calculating the gap score. Whilst gap analysis yields an 'absolute' value representing centre's performance on each attribute, grid analysis provides a 'relative' measurement of performance showing the relative position of attribute in the grid.

Amongst the three centres investigated, the grid analysis diagram was found to be the most useful element in the model. Indeed, grid analysis (also known as importance-performance analysis) has been widely used since it was developed by Martilla and James in 1977. Generally speaking, two major benefits of grid analysis were pointed out by the managers during the interviews: a snapshot of overall performance and a clear-cut line separating the strength and weakness.

• ***Applying Factor Analysis to Measure Service Quality***

Except for DEA, factor analysis is another common aggregate analysis that has been widely used in the research of service quality or customer satisfaction. The author therefore attempted to test its practicability in the revised report for Centre A (as shown in Figure 8.6). Five service dimensions were extracted from the original 25 NBS attributes, and the industrial averages of these dimensions were also plotted in the figure. However, the acceptability of this aggregate approach was again very low. This result basically coincides with the comments on using DEA in service quality measurement; that is, more detailed information is required for the facility managers to understand customers' comments on each single service attribute.

Figure 8.6 Factor analysis and industrial norm



• ***Customer Segmentation Analysis***

Another attempt to apply factor analysis in the model is to provide customer segmentation analysis. This part, based on customer's perceived importance of service quality, was also included in the revised report for Centre A and the results were demonstrated by two tables (Table 8.7 and 8.8). As shown in Table 8.7, three segments were yielded with different priorities on the five service dimensions. These three customer segments were then cross-tabulated against customers' demographic profiles (Table 8.8). A detailed explanation can be found in the Appendix B.

In sum, 'staff' is the major concern of older people whereas 'accessibility' is the major concern of younger and unemployed people. There are no significant differences between segments for other targeted groups, such as female and ethnic minority. Differing from the previous section, customer segmentation analysis was appreciated by the managers in Centre A. However, the pitfall of factor analysis in

measuring service quality has not been noticed by the managers in this part of analysis, i.e. detailed information was lost while aggregating the service attributes.

Table 8.7 Segment priorities

Dimensions	Mean importance scores							
	Overall		Segment 1 (48%)		Segment 2 (18%)		Segment 3 (34%)	
Accessibility	4.55	(1st)	4.83	(2nd)	3.73	(3rd)	4.64	(1st)
Staff	4.51	(2nd)	4.92	(1st)	4.06	(1st)	4.20	(3rd)
Facility & Cleanliness	4.43	(3rd)	4.82	(3rd)	3.66	(4th)	4.31	(2nd)
Value for money	4.38	(4th)	4.81	(4th)	3.77	(2nd)	4.17	(4th)
Catering	3.96	(5th)	4.59	(5th)	2.95	(5th)	3.68	(5th)

Notes: The number with bracket is the mean ranking across factors.

Table 8.8 Segment profiles

	Profiles		Segment 1 (48%)	Segment 2 (18%)	Segment 3 (34%)
Gender	Male	51%	0.91	1.09	1.09
	Female	49%	1.11	0.90	0.90
Ethnicity	White	88%	0.97	1.04	1.03
	Ethnic minority	12%	1.10	0.89	0.79
Age	<19 years	7%	1.00	0.29	1.39
	19-59 years	83%	0.98	1.04	1.00
	60+ years	10%	1.20	1.17	0.43
Occupation	Student	7%	0.99	0.32	1.38
	Work FT	62%	1.01	1.02	0.97
	Work PT	13%	0.74	1.33	1.19
	Retired	8%	1.40	1.21	0.33
	Unemployed	2%	0.53	1.19	2.18
	Housewife	8%	1.05	0.91	0.98

Indeed, in the academic community, using factor analysis for data reduction has been criticised by many researchers, e.g. Woo (1998) and Arabie and Hubert (1994). First, factor mean score will replace the original attribute score to run factor analysis so the result may not truly reflect customers' judgements. Second, certain attributes that do not load on any factor need to be excluded from the analysis. Even so, the joint use of factor analysis and cluster analysis is still the most common approach in segmentation analysis, which facilitates the classification of customers into distinct groups and the following cross-tabulation analysis.

8.3.3 Concluding Remarks

From the above findings, some implications could be drawn, which will help to justify why the five elements in Table 8.2 were suggested to be included in the final model.

- Identifying of strengths and weaknesses is the most useful part of DEA, followed by selecting benchmarking partners. The function of DEA is limited while measuring equity and service quality for two reasons. First, at the operational level, detailed information on service quality is more important to managers than a summary score representing the centre's overall performance. Second, the results of DEA for equity are basically consistent with the NBS results. By contrast, since contradictory information is often found while examining a series of efficiency ratios, virtual weights provided by DEA allow the drawing of a sharp line separating the strength and weakness.

- Although the usefulness of DEA in identifying benchmarking partners is still appreciated by the managers, no real action was taken to contact the benchmarking partners. It implies that the practicability of DEA is influenced not only by its complexity but also by the motivation of conducting process benchmarking.

- If the manager is not convinced by the weighting system provided by DEA, the joint use of AHP and DEA can be an alternative way to run the model.

- Grid analysis was found to be the best tool for measuring customer service quality. It is clear-cut (identifying clearly the weaknesses), inclusive (providing detailed information), understandable (easy analytical approach) and concise (condensing the result into one diagram).

- Although factor-cluster analysis has its limitations for segmentation analysis, it is the simplest way to provide managers with a holistic view about the priorities of different target groups.

8.4 Model's Applicability

The results presented above are of specific interest for the usefulness of each element within the model. Here, the concentration turns to evaluate the determinants of success of model's applicability as a whole.

8.4.1 Model's Design

The quality of the performance measurement model is critical in establishing the credibility of the measurement processes, and therefore critical to the confidence managers would have in using the model to assess and evaluate the programs

(Bernstein, 2001). Two quality characteristics were cited most frequently by the interviewees. Managers valued a model that is:

• ***Understandable.*** The lack of appropriate skills to interpret the DEA results was identified as a common barrier. It emerged that there was considerable confusion in the minds of those interviewed regarding their understanding of the terms related to DEA. For instance, there is a lot of jargon (Contract Manager of Centre C) and there is a difficulty to understand the report alone (Project Officer of Centre B). To facilitate understanding, some recommendations were made, such as: a meeting with the analysts to ask questions and clarify suspicious points (Project Officer of Centre B) and simplifying the technical concepts within the report (General Manager of Centre A).

• ***Concise.*** Due to time constraints, it is important that the report can be skimmed by a manager within a short period of time (Contract Manager of Centre C) and it is not necessary to show all the survey results (Chief Executive of Centre A). Some suggestions were made to make the report a more friendly and accessible document, such as: provide a summary of findings in the first few pages (Deputy Manager of Centre B) and simplify the technical concepts (Contract Manager of Centre C).

Indeed, some authors have discussed the design of performance measurement models and suggested that it should be transparent, simple to understand, have visual impact and visible to all (e.g. Neely *et al.*, 1997). The results basically coincide with these findings.

8.4.2 Model's Implementation

As argued by Neely *et al.* (2000), the process of designing a measurement model is intellectually challenging, however there is increasing anecdotal evidence to show that the real challenge is the implementation. The following three issues were revealed by the managers which may be the main barriers to implementation.

• ***Commitment.*** Top management commitment was found to be critical to the success of the model's implementation. In Centre B, the Project Officer has emerged to promote the use of the model. It is argued that the leadership must have conviction in the validity of the model in the first place, and then the engagement of senior managers will help to facilitate its use.

• ***Time.*** The availability of management time was identified across three centres as the main reason for slow progress. Joining more than one performance measurement scheme, the managers stated that they do not have enough time to read so many reports (Deputy Manager of Centre B) and to put everything into action (Chief Executive of Centre A).

• ***Priority.*** Given that the length of time between the intervention (i.e. workshop) and evaluation (i.e. post process interview) lasted more than eight months, it is highly likely that management found other more pressing issues and the enthusiasm for change declined. As such, the distraction of other events (Centre B) and relatively lower priority to use the new model (Centre C) were identified as the reasons for not contacting with the benchmarking partners.

8.4.3 Concluding Remarks

The above findings are mainly related to performance measurement initiatives but has wider implications for management commitment in change management. The literature highlights many of the issues affecting the management of change within organisations. The results provide a structured view of the factors affecting the model's implementation and basically coincide with some previous researches. For instance, top management commitment was identified as a key driver for successful performance measurement initiatives (e.g. Bourne *et al.*, 2002; Bauer *et al.*, 2004; Bernstein, 2001) and implementation of change (e.g. Frizelle, 1991; Kotter, 1995). The availability of management time was also found to be essential to implement a new performance measurement systems (e.g. Bourne *et al.*, 2000). It appears that if a long-term plan is required, there is a need to maintain enthusiasm and momentum for the duration of the plan. Thus, there is a need to continually communicate the actual and potential benefits of the change with the managers. If they are committed and enthusiastic, it is easier to gain the support of other levels of the organisation.

CHAPTER 9

Conclusion

Road map of the thesis

Introduction	Theoretical context	Methodology	Results	Conclusion
Ch1 ▶	Ch2 Performance measurement overall Ch3 Efficiency & Equity Ch4 Customer service	▶ Ch5 ▶	Ch6 Efficiency & Equity Ch7 Customer service Ch8 Action research	▶ Ch9

The thesis is concluded by specifying the two original contributions of the thesis and the research implications: (1) to present the value and critically evaluate the practicability of aggregate performance analysis; and (2) to expand the theoretical literature on performance measurement in the public leisure sector. Finally, limitations and proposals for future research are also discussed.

9.1 Contributions and Implications

9.1.1 Value and practicability of aggregate performance analysis

Although widely advocated by academics, seldom has research been done to evaluate the practicability of aggregate performance analysis. To the author's knowledge, this thesis is the first attempt, at least in the field of sport/leisure

management, to present the value of aggregate performance analysis and simultaneously evaluate its practicability. Even though the two analytical techniques used in this study- data envelopment analysis (DEA) and multivariate statistical analysis (MSA) have been applied in some previous sport/leisure related studies, most of these studies aimed to provide a comprehensive aggregate model, but ignored to examine how these techniques can contribute to managerial practices and be accepted by the practitioners. Some researchers (e.g. Chen, 2003; Smith and Street, 2005) have compared the usefulness of 'global measures' and 'partial measures', but none of them has taken steps to evidence their propositions in real-life organisations. The first original contribution of this study is therefore to provide knowledge on how aggregate analysis can be used to support current performance measurement practices in English public sport facilities, and to critically evaluate the extent to which global and partial measures outperform or complement each other.

According to Chen (2003), a good system of performance measurement should have both high 'inclusiveness' (i.e. all aspects of the organisation should be considered) and high 'convergence' (i.e. consistent and simple information should be provided to facilitate decision making). The research findings demonstrate that the 'convergence' strength of aggregate analysis can complement the 'inclusiveness' strength of the NBS ratio analysis, particularly in efficiency measurement and customer segmentation analysis. The value and practicability of the two aggregate models are summarised and discussed in turn.

• *Efficiency and equity measurement*

As discussed in Chapter 3, the major functions of DEA include: overall performance measurement, clarifying strengths and weaknesses, and identifying benchmarking partners.

In terms of overall performance measurement, as argued by Charnes and Cooper (1994) and Athanassopoulos and Ballantine (1995), rather than being based on the assessment of individual performance indicators, all-round performance is taken into account while applying DEA. It therefore gives a more balanced approach to performance measurement. The research findings in Chapter 6 show that, at the industrial level, DEA allows differentiation between overall efficiency and inefficiency, which is unavailable from ratio analysis, especially when conflicting signals exist among various indicators. However, this function is less useful for the practitioners because more detailed benchmarking data are needed for local managerial or operational purposes.

While measuring equity (i.e. the application of DEA effectiveness model), the application of DEA without weight restrictions may provide a misleading measurement of overall performance. This deficiency has been pointed out by previous researchers, e.g. Thanassoulis (1996). The results of Chapter 6 (Table 6.13) demonstrate that certain facilities achieve full marks simply by placing a lot of weight on one dimension in which they perform well and assigning a zero weight to the dimensions in which they perform poorly. It is argued that this kind of unusual mix of outputs is far from achieving equity because the access of certain target groups is ignored.

The problem of weighting flexibility has led to the development of methods for using weight restrictions in DEA assessments to ensure that the DEA efficiencies estimated reflect the performance of units on all, and not just a subset, of the input and output variables. In the action research part of this research, the author tried to introduce another subjective weighting scheme- Analytic Hierarchy Process (AHP) - into the conventional DEA model. While it is one way of overcoming the problem of weighting flexibility, it is not is not a straightforward issue for the analysis at the industrial level. According to Boussofiance (1991), the reason is that individual units may have their own priorities and therefore it may be difficult to derive a common set of weights.

Regarding the function of clarifying strengths and weaknesses, since contradictory information is often found while examining a series of efficiency ratios, virtual weights provided by DEA allow clear distinction between the strengths and weaknesses. However, the performance indices provided by the NBS can still give an insight into the gaps in specific performance dimensions. This is an area where the two approaches can complement each another. From the practitioners' points of view, identifying the strengths and weaknesses is also the most important function of DEA. The reason, which has been demonstrated by Staat and Maik (2000), is that the comparison of virtual weights provided by DEA need not refer to complicated mathematical concepts and therefore are comparatively easy for the lay person to understand.

However, this function is highly restricted in the measurement of equity, because the results of DEA are basically consistent with the NBS results. Thus there is no reason for the practitioners to adopt a more complicated model.

Both the analyses at the industrial and individual facility levels confirm this point. When measuring efficiency, contradictory information usually exists between different efficiency ratios, as they link the interaction between different sets of inputs and outputs. Efficiency can be achieved through either output maximisation or input minimisation. As for effectiveness, the measurement is more straightforward since output maximisation is the only target. Consequently, the identification of strengths and weaknesses in effectiveness PIs can simply examine their values relevant to the national benchmarks.

Third, the benchmarking partners suggested by the NBS are based on those who perform well in specific performance indicators. However, DEA can complement the NBS by selecting benchmarking partners which have better all-round performance. Moreover, the NBS type of benchmarking partners are determined on the basis of a similar operating context. Nevertheless, DEA can enhance the NBS by providing benchmarking partners with similar strengths and weaknesses, but it must be noted that this function in equity measurement is again restricted due to the weighting flexibility of DEA, which will provide inappropriate benchmarking partners with unbalanced performance and ignore the access of certain target groups.

Finally, although the DEA effectiveness model has been applied by some researchers to measure the performance of regional development (Chang *et al.*, 1995), with reference to the financial performance of banks (Al-Shammari and Salimi, 1998; Halkos and Salamouris, 2004) and hospitals (Ozcan and McCue, 1996) etc., none of them tried to critically evaluate the suitability of the DEA effectiveness model. As such, another innovative part of this research is to demonstrate that the practicability of DEA is mainly constrained in the measurement of efficiency.

• *Customer service measurement*

As discussed in Chapter 5, the value of using MSA to measure customer service quality and segment customer base is to reduce the complexity of data.

Regarding service quality measurement, this benefit was realised by clarifying the industry's strengths and weaknesses as well as justifying the validity of the gap analysis model, which is the most popular approach in the industry. It means that gap analysis is not only highly acceptable to practitioners due to its simplicity, but its validity has also been demonstrated through this study. However, at the individual facility level, the major pitfall of MSA lies in the fact that detailed information is lost while aggregating various attributes into factors. For the practitioners, more detailed information is required to understand customers' views about each single service attribute because management decisions relate to individual attributes. Therefore, the value and practicability of aggregate analysis is much lower than that of traditional approaches, such as grid analysis.

Although the above pitfall also exists in the customer segmentation analysis, factor-cluster analysis is probably the simplest way to provide managers with a holistic view about the priorities of different target groups. The reason is that, to derive marketing and managerial implications, it is necessary to cross-tabulate the results of cluster analysis against a series of customers' demographics and participation profiles. Aggregating individual service attributes into dimensions (factors) can reduce the complexity of data and facilitate data interpretation.

9.1.2 Performance measurement in the public leisure sector

The second original contribution of this thesis is to expand the theoretical literature on performance measurement in the public leisure sector. Previous research mainly focused on the application of DEA or MSA in the private sector. However, there is some evidence to indicate differences between the public and private sector in the nature of service delivery (e.g. Crompton and MacKay, 1989) as well as the motivation for benchmarking (e.g. Bowerman *et al.*, 2002; Kouzmin *et al.*, 1999). Also, the acceptability of the quantitative models may not solely rely on the models themselves but also the characteristics of the public sector leisure provision. Certainly, any attempt to assess the value and practicability of aggregate performance analysis in the context of public leisure service has to recognise these problems. In order to clarify the theoretical implications of this research, it is useful to answer the following two questions launched by Smith and Street (2005) in Chapter 2.

- Is it legitimate to seek to develop global measures of organisational performance?

- Why should policy makers have any interest in global measures of organisational performance?

• ***Criteria of performance measures***

The first question is related to the criteria of good performance measures. The results of action research show that 'grid analysis' benefits the facility managers more than the two aggregate models. As discussed in Chapter 8, two major benefits of grid analysis pointed out by the interviewees are: a snapshot of overall performance and a clear-cut line demarcating the strengths and weaknesses. Actually, these two benefits

can also be realised by DEA, but the acceptability of DEA is much lower than that of grid analysis.

According to the 'inclusiveness' and 'convergence' criteria proposed by Chen (2003), DEA can be regarded as good in 'convergence' (i.e. providing a snapshot of overall performance) but weak in 'inclusiveness' (i.e. presenting detailed results). However, the action research results demonstrate that 'simplicity' is much more important than 'convergence' in the research context, since the weighting procedure for yielding a single global measure is complicated.

For example, although DEA can provide managers with a snapshot of centres' overall performance, its analytical process is less understandable and, in turn, makes the results less transparent and less acceptable. By contrast, by applying grid analysis, 'inclusiveness' and 'simplicity' are ensured simultaneously since it visualises complicated data and provides simple information to facilitate decision making. The application of either DEA or MSA requires levels of knowledge in mathematics and statistics which most managers in the public leisure sector probably do not possess. This research therefore concludes that 'simplicity' is more important than 'convergence' as a criterion of good performance measurement.

Other evidence to support the author's argument is that, at the industrial level or from a strategic regulator's point of view, global measures may be useful. For instance, it might identify beacons of good practice. However, at the individual facility level, it is more important to examine specific organisational functions rather than be given a summary measure of overall performance, i.e. much more detailed benchmarking data is needed for local managerial purposes (Smith and Street, 2005). This is the reason why traditional ratio analysis- an inclusive and easily interpretable

approach- continues to be the method of choice, at least for the UK government, in reporting performance in the delivery of public services (Thanassoulis *et al.*, 1996).

To sum up, for practitioners in the public leisure sector, not only does performance data need to be inclusive, the analytical process also needs to be simple and understandable.

- ***Willingness to apply the proposed model***

The second question is related to the willingness of facility managers to apply aggregate performance analysis. The research findings demonstrate the following two factors which determinate the feasibility of introducing the proposed model into the public leisure sector.

First, since the facility managers generally lack sufficient analytical skills, the practicability of aggregate performance analysis relies highly on the communication and interaction between practitioners and analysts. By adopting action research, the author played a proactive role in introducing the new model. However, without the support of leading bodies (e.g. Department for Culture, Media and Sport [DCMS] or Sport England), it is less likely to that other local authorities would be stimulated to accept a new model in the same way. One example which highlights this point in the UK was the promotion of the Public Services Productivity Panel to introduce DEA and Stochastic Frontier Analysis (SFA) to measure the efficiency of police forces (Spottiswoode, 2000).

Second, although the usefulness of DEA in identifying benchmarking partners was appreciated by the facility managers, no real action was taken to contact the benchmarking partners identified by the model. Time constraints were frequently

cited as the reason by the managers, but this finding has wider implications about the motivation of benchmarking in the public sector. According to Bowerman *et al.* (2002), whilst information generated through benchmarking in the private sector is confidential, public agencies have no competitive drawbacks to fear from passing information on to peer organisations. Consequently, the public sector may have greater potential for benchmarking than the private sector due to the availability of a wide choice of benchmarking partners. However, in the private sector, benchmarking is often undertaken in order to be the best. By contrast, public sector organisations may strive, through benchmarking, to be 'good enough' or merely to demonstrate that they are not the worst. In this case, benchmarking results become more important than acting on those results in order to close performance gaps (Bowerman *et al.*, 2002). It is therefore concluded that if 'data benchmarking' is an end in itself and local authorities have little incentive to conduct 'process benchmarking', the feasibility of the proposed model will be seriously constrained.

To sum up, more work is still required before the potential benefits of aggregate analysis in supporting performance measurement in public sport facilities can be fully realised.

9.2 Limitations and Extensions

In terms of efficiency and equity measurement, since the major problem revealed in this research is the weighting flexibility of DEA, there are at least two ways in which this research might be extended. First, one may use a Delphi-like technique or AHP, which are techniques for collecting and organising expert opinion in decision

making, to obtain subjective weights to be included in the DEA models (Zhu, 2004). Alternatively, the design of the current CPA framework is to set thresholds for each PI. The same logic can be also applied to set upper and lower thresholds to restrict the weighting flexibility of DEA. There are two ways to achieve this: absolute and relative weight restrictions (Allen *et al.*, 1997).

In terms of service quality, some models (e.g. Johnston, 1995a; Johnston, 1995b; Johnston and Heineke, 1998) have been proposed to improve its measurement. For example, while some attributes may be necessary to generate a good perception of quality, improvements in these attributes will not generate corresponding improvements in the service quality perception. That is, although there will be a desired level to which the organisation should aim, customers will accept a lower degree of performance. This is known as a 'zone of tolerance' and it will vary between service attributes and between different customers. Johnston (1995a) argued that once outside the 'zone of tolerance' there could be a disproportionate impact on perceptions, i.e. relatively small changes in performance could have a large impact on how the service is viewed. The nature and size of the 'zone of tolerance' of specific factors is worth investigating further.

In terms of customer segmentation, one of main deficiencies of the factor-cluster approach is that six original NBS attributes did not load on any factor and therefore were excluded from the analysis. In fact, segmentation is now becoming more sophisticated with the advancements in modelling software. These new techniques allow segments to be developed by computers without the bias of human judgments. For instance, some researchers have applied non-traditional segmentation methods such as chi-squared automatic interaction detection (CHAID) model (e.g. Chen, 2003),

artificial neural networks (ANNs) (e.g. Bloom, 2004), conjoint analysis (e.g. Becker-Suttle *et al.*, 1994) or logistic regression (e.g. Taks and Scheerder, 2006). These models therefore suggest opportunities for further research in the context of public sport facilities.

Finally, there are also two limitations on the action research part of this thesis. First, the scope of the problems addressed in the study was mainly confined to the operational level of the organisation. Although the model's practicability may be similar at the strategic level, it is yet to be tested. One may argue that another limitation of the study is the difficulty in drawing general theoretical conclusions from a few applications of the approach. In order to draw further conclusions outside this sample, future research should test the model in more case studies.

References

Adelman, C. (1993) Kurt Lewin and the origins of action research, *Educational Action Research*, 1, 7-24.

Agrell, P.J. and West, B.M. (2001) A caveats on the measurement of productivity efficiency, *International Journal of Production Economics*, 69, 1-14.

Allen, R., Athanassopoulos, A., Dyson, R. G. and Thanassoulis, E. (1997) Weights restrictions and value judgements in data envelopment analysis: evolution, development and future directions, *Annals of Operations Research*, 73, 13-34.

Al-Shammari, M. and Salimi, A. (1998) Modeling the operating efficiency of banks: a nonparametric methodology, *Logistics Information Management*, 11(1), 5–17.

Ammons, D.N. (1999) A proper mentality for benchmarking, *Public Administration Review*, 59(2), 105-109.

Andrews, R. (2004) Analysing deprivation and local authority performance: the implications for CPA, *Public Money and Management*, January, 19-26.

Argyris, C., Putnam, R. and Smith, D.M. (1985) *Action Science*, Oxford, Jossey-Bass.

Athanassopoulos. A. D. and Ballantine, J. A. (1995) Ratio and frontier analysis for assessing corporate performance: evidence from the grocery industry in the UK, *Journal of the Operational Research Society*, 46, 427–440.

Atilgan, E, Akinci, S. and Aksoy, S. (2003) Mapping service quality in the tourism industry, *Managing Service Quality*, 13(5), 412-422.

Audit Commission (1995), *People, Pay and Performance: Paying the Piper*, London, Audit Commission.

Audit Commission (2003) *Comprehensive Performance Assessment*, London, Audit Commission,.

Audit Commission (2005) *Service Assessment Framework: Technical Guide to CPA 2005 for Single Tier and County Councils*, London, Audit Commission.

Avkiran, N. (2001) Investigating technical and scale efficiencies of Australian universities through data envelopment analysis, *Socio-Economic Planning Sciences*, 35, 57–80.

Bagozzi, R. P., Rosa, J. A., Celly, K. S., and Coronel, F. (1998) *Marketing Management*, NJ: Prentice Hall, Englewood Cliffs.

Bailey, R. (2005) Evaluating the relationship between physical education, sport and social inclusion, *Educational Review*, 57(1), 71-90.

Ball, A, Bowerman, M. and Hawksworth, S. (2000) Benchmarking in local government under a central government agenda, *Benchmarking: An International Journal*, 7(1), 20-34.

Banker, D. and Morey R. (1986) Efficiency analysis for exogenously fixed inputs and outputs, *Operations Research*, 34(4), 513–521.

Banker, R.D. and Morey, R.C. (1989) Incorporating value judgments in efficiency analysis, *Research in Government and Nonprofit Accounting*, 5, 245–67.

Banker, R.D., Charnes, A. and Cooper, W.W. (1984) Models for estimation of technical and scale inefficiencies in data envelopment analysis, *Management Science*, 30, 1078-1092.

Baskerville, R. and Pries-Hejeb, J. (1999) Grounded action research: a method for understanding IT in practice, *Accounting, Management and Information Technology*, 9, 1–23.

Baskerville, R.L. and Wood-Harper, A.T. (1996) A critical perspective on action research as a method for information systems research, *Journal of Information Technology*, 11, 235-46.

Bauer, J., Tanner, S.J. and Neely, A. (2004) Developing a performance measurement audit template – a benchmarking study, *Measuring Business Excellence*, 8 (4), 17-25.

Beaver, W. (1968) Alternative financial ratios as predictors of failure, *Accounting Review*, 43(1), 113-122.

Becker-Suttle, C.B., Weaver, P. and Crawford-Welch, S. (1994) A pilot study

utilizing conjoint analysis in the comparison of age based segmentation strategies in the full service restaurant market, *Journal of Restaurant and Food service Marketing*, 1(2), 71-91.

Bell, R.A. and Morey, R.C. (1994) The search for appropriate benchmarking partners: a macro approach and application to corporate travel management, *Omega*, 22(5), 477-490.

Bennet, G., Henson, R. K., and Zhang, J. (2003) Generation Y's perceptions of the action sports industry segment, *Journal of Sport Management*, 17(2), 95-115.

Bernstein, D.J. (2001) Local government measurement use to focus on performance and results, *Evaluation and Program Planning*, 24, 95-101.

Bhargava, M., Dubelaar, C. and Ramaswami, S. (1994) Reconciling diverse measures of performance: A conceptual framework and test of a methodology, *Journal of Business Research*, 31, 235-246.

Bloom, J.Z. (2004) Tourist market segmentation with linear and non-linear techniques, *Tourism Management*, 25, 723–733.

Bodet, G. (2006) Investigating customer satisfaction in a health club context by an application of the Tetraclasse model, *European Sport Management Quarterly*, 6(2), 149-165.

Booms, B. and Bitner, M. (1981) Marketing strategies and organizational structures for service firms, in Donnelly, J. and George, W. (eds.) *Marketing of Services*, Chicago, American Marketing Association, 47-51.

Borrie, W.T., Christensen, N., Watson, A.E., Miller, T.A. and McCollum, D.W. (2002) Public purpose recreation marketing: a focus on the relationships between the public and public lands, *Journal of Park and Recreation Administration*, 20(2), 49-68.

Bourne, M., Mills, J., Wilcox, M., Neely, A. and Platts, K. (2000) Designing, implementing and updating performance measurement systems, *International Journal of Operations and Production Management*, 20(7), 754-771.

Bourne, M., Neely, A., Platts, P. and Mills, J. (2002) The success and failure of performance measurement initiatives: Perceptions of participating managers,

International Journal of Operations and Production Management, 22(11), 1288 -1310.

Boussofiance, A., Dyson, R. G. and Thanassoulis, E. (1991) Applied data envelopment analysis, *European Journal of Operational Research*, 52, 1-15.

Bowerman, M. and Ball, A. (2000) Great expectations: Benchmarking for best value, *Public Money and Management*, 20(2), 21-26.

Bowerman, M. and Ball, A. (2001) *Authorities on Benchmarking? The State of Benchmarking in the UK Local Government Sector*, London, CIMA.

Bowerman, M., Ball, A. and Francis, G.A.J. (2001) Benchmarking as a tool for the modernisation of local government, *Financial Accountability and Management*, 17(4), 321-329.

Bowerman, M., Ball, A. and Francis, G.A.J. (2002) The evolution of benchmarking in UK local authorities, *Benchmarking: An International Journal*, 9(5), 429-449.

Bowlin, W.F. (1999) An analysis of the financial performance of defense business segments using data envelopment analysis, *Journal of Accounting and Public Policy*, 18, 287-310.

Bradish, C. and Lathrop, A.H. (2001) Girl power: examining the female pre-teen and teen as a distinct segment of the sport marketplace, *Sport Marketing Quarterly*, 10(1), 19-24.

Brassington, F., and Pettit, S. (2000) *Principles of Marketing*, 2nd ed., Harlow, Pearson Education.

Broadbent, J. (2003) Comprehensive Performance Assessment: the crock of gold at the end of the performance rainbow? *Public Money and Management*, January, 5-9.

Brockett, P.L. and Golany, B. (1996) Using rank statistics for determining programmatic efficiency differences in data envelopment analysis, *Management Science*, 42, 466-472.

Bruijn, H. (2002) Performance measurement in the public sector: strategies to cope with the risks of performance measurement, *The International Journal of Public Sector Management*, 15(7), 578-594.

Bryman, A. (1989) *Research Methods and Organisation Studies*, London, Urwin Hyman.

Brysland, A. and Curry, A. (2001) Service improvements in public services using SERVQUAL, *Managing Service Quality*, 11(6), 389-401.

Burdge, R. (1969) Levels of occupational prestige and leisure activity, *Journal of Leisure Research*, 1, 262-274.

Burns, R., Graefe, A. and Absher, J. (2003) Alternate measurement approaches to recreational customer satisfaction: satisfaction-only versus gap scores, *Leisure Sciences*, 25, 363–380.

Burns, R.B. (1997) *Introduction to research methods*, 3rd ed., Melbourne, Longman.

Burrell, G. and Morgan, G. (1979) *Sociological Paradigms and Organizational Analysis*, Aldershot, Ashgate Publishing Company.

Carman, J. (1990) Consumer perceptions of service quality: an assessment of the SERVQUAL dimensions, *Journal of Retailing*, 66(1), 33-55.

Carr, W. and Kemmis, S. (1983) *Becoming Critical: Knowing through Action Research*, Geelong, Deakin University.

Cha, S., McCleary, K.W. and Uysal, M. (1995) Travel motivations of Japanese overseas travelers: a factor- cluster segmentation approach, *Journal of Travel Research*, 34(1), 33-39.

Chang, P. L., Hwang, S. N. and Cheng, W. Y. (1995) Using data envelopment analysis to measure the achievement and change of regional development in Taiwan, *Journal of Environmental Management*, 43, 49-66.

Charmaz, K. (2000) Grounded theory: Objectivist and constructivist methods. in N.K. Denzin and Y.E. Lincoln (eds.) *Handbook of qualitative research*, 2nd ed., CA, Sage.

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

Charnes, A., and Cooper, W.W. eds. (1994) *Data Envelopment Analysis: Theory, Methodology and Applications*, Boston, Kluwer.

Checkland, P. (1991) From framework through experience to learning: the essential nature of action research, in Nissen, H.E. *et al.* (eds.), *Information Systems Research: Contemporary Approaches and Emergent Traditions*, Amsterdam, Elsevier.

Chen, J.S. (2003) Developing a travel segmentation methodology: a criterion-based approach, *Journal of Hospitality and Tourism Research*, 27(3), 310-327.

Chen, W. C. (2003) *Integrating approaches to efficiency and productivity measurement*. PhD Thesis, Atlanta, Georgia Institute of Technology.

Chen, Y. and Ali, A. I. (2002) Output-input ratio analysis and DEA frontier. *European Journal of Operational Research*, 142, 476-479.

Child, D. (1990) *The Essentials of Factor Analysis*, London, Cassell.

Chung, K.Y., Oh, S.Y., Kim, S.S. and Han, S.Y. (2004) Three representative market segmentation methodologies for hotel guest room customers, *Tourism Management*, 25, 429–441.

Citizen's Charter Unit (1992) *Raising the Standard: Britain's Citizen's Charter and Public Service Reform*, London, Foreign and Commonwealth Office.

Clarke, A. (1956) The use of leisure and its relations to levels of occupational prestige, *American Sociological Review*, 21, 301-132.

Coalter, F. (1993) Sports participation: price or priorities? *Leisure Studies*, 12(3), 171-182.

Coalter, F. (1998) Leisure studies, leisure policy and social citizenship: the failure of welfare or the limits of welfare? *Leisure Studies*, 17, 21–36.

Coalter, F. (2000) Public and commercial leisure provision: active citizens and passive consumers? *Leisure Studies*, 19, 163–181.

Coelli, T. (1996) A guide to DEAP version 2.1: a data envelopment analysis (computer) program *CEPA Working Paper 96/08*, Australia, University of New England.

Coelli, T. J., Prasada Rao, D. S., O'Donnell, C. J. and Battese, G. E. (2005) *An Introduction to Efficiency and Productivity Analysis*. 2nd edition, New York,

Springer.

Collins, M. F. (2003) *Sport and Social Exclusion*. London, Routledge.

Collins, M. F. (2004) Sports participation in decline, *Recreation*, November, 26-27.

Collins, M. F., Henry, I. and Houlihan, B. (1999) *Sport and Social Inclusion: A Report to the Department of Culture, Media and Sport*. Loughborough, Institute of Sport and Leisure Policy, Loughborough University.

Commission for Health Improvement (2003) *Rating the NHS*, London, Commission for Health Improvement.

Connolly, S. (2002) Comprehensive performance assessment, *Recreation*, December, 34-38.

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

Cox, A. and Thompson, I. (1998) On the appropriateness of Benchmarking, *Journal of General Management*, 23, 1-20.

Craig, S. (1995) Performance indicators, *Recreation Management Fact File*, London, Sports Council,

Crawford-Welch, S. (1991) Market segmentation in the hospitality industry, *Hospitality Research Journal*, 14(2), 295-308.

Crompton J.L. and Lamb, C.W. (1983) The importance of the equity concept in the allocation of public services, *Journal of Macromarketing*, 3(1), 28-48.

Crompton, J. and Love, L. (1995) The predictive validity of alternative approaches to evaluating quality of a festival, *Journal of Travel Research*, 34(1), 11–24.

Crompton, J. and MacKay, K. (1989) Users' perceptions of the relative importance of service quality dimensions in selected public recreation programs, *Leisure Sciences*, 11(4), 367–375.

Crompton, J. L. and Lue, C. C. (1992) Patterns of equity preferences among Californians for allocating park and recreation resources, *Leisure Sciences*, 14, 227-246.

Crompton, J. L. and Wicks, B. E. (1986) Citizen and administrator perspectives of equity in the delivery of park services, *Leisure Sciences*, 8(4), 341-365.

Crompton, J. L. and Wicks, B. E. (1988) Implementing a preferred equity model for the delivery of leisure services in the US context, *Leisure Studies*, 7, 287–304.

Cronin, J and Taylor, S. (1994) SERVPERF versus SERVQUAL: Reconciling performance-based perceptions minus expectations measurement of service quality, *Journal of Marketing*, 58(1), 125-131.

Cronin, J., and Taylor, S. (1992) Measuring service quality: a re-examination and extension, *Journal of Marketing*, 56(3), 55–68.

Day, J., Reynolds, P. and Lancaster, G. (1998) A marketing strategy for public sector organisations compelled to operate in a CCT environment, *International Journal of Public Sector Management*, 11(7), 583-595.

DCMS (1999) *Arts and Sports: A Report to the Social Exclusion Unit*. London, Department for Culture, Media and Sport (DCMS).

De Borger, B. and Kerstens K. (1996) Cost efficiency of Belgian local governments: a comparative analysis of FDH, DEA, and econometric approaches, *Regional Science and Urban Economics*, 26, 145–170.

Denzin N. K. and Lincoln Y. S. eds. (1994) *Handbook of Qualitative Research*, CA, Thousand Oaks, Sage Publications.

Despic, O. (2004) DEA-R relative efficiency model: some theoretical and practical considerations, *Aston Business School Research Papers*, Birmingham, Aston University.

DETR (2000) *Guide to quality schemes and Best Value*, London, DETR.

Díaz-Martín, A., Iglesias, V., Vazquez, R., and Ruiz, A. (2000) The use of quality expectations to segment a service market, *Journal of Services Marketing*, 14(2), 132-146.

Dibb, S. (1998) Market segmentation: strategies for success, *Marketing Intelligence and Planning*, 16(7), 394–406.

Doble, M. (1995) Measuring and improving technical efficiency in UK post o ffice

counters using data envelopment analysis, *Annals of Public and Cooperative Economics*, 66(1), 31-64.

Doherty, N.F., Saker, J. and Smith, I.G. (1998) Developing 'appropriate' marketing within the public leisure and library sectors: a comparative study, *Managing Leisure: An International Journal*, 3, 149-161.

Dolnicar, S. (2002) A review of data-driven market segmentation in tourism, *Journal of Travel and Tourism Marketing*, 12(1), 1-22.

Eisenhardt K M, (1989) Building theories from case study research, *Academy of Management Review* 14(4), 532-550.

Elden, M. and Chisholm, R.F. (1993) Emerging varieties of action research: introduction to the special issue, *Human Relations* 46, 121-142.

Farrell (1957) The measurement of productive efficiency, *Journal of Royal Statistical Society*, 120(3), 253-290.

Favret, L. (2000) Benchmarking, annual library plans and best value: the implications for public libraries, *Library Management* 21(7), 340-348.

Fern, E.F. (2001) *Advanced Focus Group Research*, London, Sage Publications.

Floyd, M. (1999) Race, ethnicity and use of the national park system, *Social Science Research Review*, 1, 1-23.

Foot, J. (1998) *How to do Benchmarking: A Practitioner's Guide*, London, Inter-Authorities Group.

Fountain, J.E. (2001) Paradoxes of public sector customer service, *Governance: An International Journal of Policy and Administration*, 14(1), 55-73.

Frizelle, G.D.F.M. (1991) Deriving a methodology for implementing CAPM systems, *International Journal of Production Management*, 11(7), 6-26.

Funk, D.C. (2002) Consumer based marketing: the use of micro-segmentation strategies for understanding sport consumption, *International Journal of Sports Marketing and Sponsorship*, 4(3), 231-256.

Funk, D.C. and Pastore, D.L. (2000) Equating attitudes to allegiance: the usefulness of selected attitudinal information in segmenting loyalty to professional sports

teams, *Sport Marketing Quarterly*, 9(4), 175-184.

Glaser,B.G. (1992) *Emergency Versus Forcing: Basics of Grounded Theory Analysis*, CA, Mill Valley, Sociology Press.

Golany, B. and Roll, Y. (1989) An application procedure for DEA, *Omega*, 17(3), 237-250.

Gratton, C. and Taylor, P. (1985) *Sport and Recreation: An Economic Analysis*, London, Spon.

Gratton, C. and Taylor, P. (1988) *Economics of Leisure Services Management*, Essex, Longman/ILAM.

Gratton, C. and Taylor, P. (2000) *Economics of Sport and Recreation*, London, Spon.

Green, P.E. and Krieger, A.M. (1995) Alternative approaches to cluster-based market segmentation, *Journal of the Market Research Society*, 37(3), 221-239.

Greenwell, T. C., Fink, J. S., and Pastore, D. L. (2002) Perceptions of the service experience: using demographic and psychographic variables to identify customer segments, *Sport Marketing Quarterly*, 11(4), 234-242.

Greenwood, D. J., and Levin M. (1998) *Introduction to Action Research: Social Research for Social Change*, London, Sage Publications.

Grönroos, C. (1984) A service quality model and its marketing implications. *European Journal of Marketing*, 18(4), 36-44.

Grosskopf, S. (1993) Efficiency and productivity. In Harold et al., eds., *The Measurement of Productive Efficiency: Techniques and Applications*, chapter 4, 160-194, New York, Oxford University Press.

Guba, E. G. and Lincoln, Y. S. (1994) Competing paradigms in qualitative research, in N. K. Denzin and Lincoln Y. S. eds., *Handbook of Qualitative Research*, CA, Sage, Thousand Oaks, Publications.

Guest, C. and Taylor, P. (1999) Customer oriented public leisure services in the United Kingdom, *Managing Leisure: An International Journal*, 4, 94-106.

Hackman, S.T., Frazelle, E.H., Griffin, P.M., Griffin, S.O. and Vlasta, D.A. (2001)

Benchmarking warehouse and distribution operations: An input-output approach, *Journal of Productivity Analysis*, 16, 79-100.

Hair, J., Anderson, R., Tathain, R. and Black, W. (1995) *Multivariate Data Analysis*, 4th ed., NJ, Prentice-Hall.

Haley, R. (1985) *Developing Effective Communication Strategies: A Benefit Segmentation Approach*, New York, Wiley.

Halkos, G. E. and Salamouris, D. S. (2004) Efficiency measurement of the Greek commercial banks with the use of financial ratios: a data envelopment analysis approach, *Management Accounting Research*, 15, 201–224.

Hass, D.A. and Murphy, F.H. (2003) Compensating for non-homogeneity in decision-making units in data envelopment analysis, *European Journal of Operational Research*, 144, 530–544.

Haywood-Farmer, J. and Stuart, I. (1990) An instrument to measure the 'degree of professionalism' in a professional service, *The Service Industries Journal*, 10(2), 336–347.

Heale, G. (2003) Applying theory to practice: an action research resource pack for professionals, *Clinnical Chiropractic*, 6, 4-14.

Her Majesty's Treasury (2003) *Public Services: Meeting the Productivity Challenge*, London, Stationery Office.

Heron, John, and Reason, P (2001) The practice of co-operative inquiry: Research with rather than 'on' People, *Handbook of Action Research: Participative Inquiry and Practice*, Reason and Bradbury eds., London, Sage.

Hibbard, J., Stockard, J. and Tusler, M. (2003) Does publicizing hospital performance stimulate quality improvement efforts? *Health Affairs*, 22, 84-98.

Hinton, M., Francis, G. and Holloway, J. (2000) Best practice benchmarking in the UK, *International Journal of Benchmarking*, 7(1), 52-61.

Howat, G., Absher, J., Crilley, G. and Milne, I. (1996) Measuring customer service quality in sports and leisure centres, *Managing Leisure: An International Journal*, 1(2), 77-89.

Hsu, H.C. and Lee, E.J. (2002) Segmentation of senior motorcoach travellers, *Journal of Travel Research*, 40, 364-373.

Hudson, S., Hudson, P. and Miller, G. (2004) The measurement of service quality in the tour operating sector: a methodological comparison, *Journal of Travel Research*, 42(3), 305-312.

Hult, M. and Lennung, S. (1980) Towards a definition of action research: a note and a bibliography, *Journal of Management Studies* 17, 241-50.

Johnson, R. A., and D. A. Wichern (1998) *Applied Multivariate Statistical Analysis*, 4th ed., Englewood Cliffs, NJ, Prentice Hall.

Johnston, R. (1995a) The zone of tolerance: exploring the relationship between service transactions and satisfaction with the overall service, *International Journal of Service Industry Management*, 6(2), 46-61.

Johnston, R. (1995b) The determinants of service quality: satisfiers and dissatisfiers, *International Journal of Service Industry Management*, 6(5), 53-71.

Johnston, R. and Heineke, J. (1998) Exploring the relationship between perception and performance: priorities for action, *The Service Industries Journal*, 18(1), 101-112.

Jonsonn, S. (1991) Action research, in H. Nissen, H.K. Klein and R. Hirschheim eds., *Information Systems Research: Contemporary Approaches and Emergent Traditions*, New York, North-Holland.

Joseph, M., Yakhou, M. and Stone, S. (2005) An educational institution's quest for service quality: customers' perspective, *Quality Assurance in Education*, 13(1), 66-82.

Kaplan, R.S. and Norton, D.P. (1992), The balanced scorecard ± measures that drive performance, *Harvard Business Review*, January-February, 71-79.

Kelly, J. (1996) *Leisure*. 3rd edition, NJ, Prentice Hall.

Kennerley, M. P. and Neely, A. D. (2000) Performance Measurement Frameworks - A Review, *Proceedings of the 2nd International Conference on Performance Measurement*, Cambridge, 291-298.

Kennett, P. A., Sneath, J. Z. and Henson, S. (2001) Fan satisfaction and segmentation: a case study of minor league hockey spectators, *Journal of Targeting, Measurement and Analysis for Marketing*, 10(2), 132-142.

Ketchen D.J. and Shook, C.L. (1996) The application of cluster analysis in strategic management research: an analysis and critique, *Strategic Management Journal*, 17, 441-458.

Kim, D., and Kim, S. (1995) QUESC: An instrument for assessing the service - quality of sport centres in Korea, *Journal of Sport Management*, 9(2), 208-220.

Kim, S., Yoo, E. and Pederson, P.M. (2007) Market segmentation in the K-League: an analysis of spectators of the Korean Professional Soccer League, *International Journal of Sport Marketing and Sponsorship*, 8(2), 141-158.

Kloot, L. (1999) Performance measurement and accountability in Victorian local government, *The International Journal of Public Sector Management*, 12(7), 565-583.

Kock N., and Lau F. (2001) Information systems action research: serving two demanding masters, *Information Technology and People* 14(1), 6 –12.

Kock, N. (2003) The three threats of action research: a discussion of methodological antidotes in the context of an information systems study, *Decision support system*, 37, 265-286.

Kotler, P. (1994) *Marketing Management*, 8th ed., Englewood Cliffs, NJ, Prentice Hall.

Kotler, P. and Andreasen, A.R. (1991) *Strategic Marketing for Non-profit Organisation*, 4th ed., Englewood Cliffs, NJ, Prentice Hall.

Kotter, J.P. (1995) Leading change: why transformation efforts fail, *Harvard Business Review*, 73(2), 59-67.

Kouthouris C. and Alexandris, K. (2005) Can service quality predict customer satisfaction and behavioral intentions in the sport tourism industry? An application of the SERVQUAL model in an outdoors setting, *Journal of Sport Tourism*, 10(2), 101–111.

Kouzmin, A., Lof er, E., Klages, H. and Korac-Kakabadse, N. (1999) Benchmarking and performance measurement in public sectors: towards learning for agency effectiveness, *The International Journal of public Sector Management*, 12(2), 121-144.

Krueger, R. A., and M. A. Casey (2000) *Focus groups: A Practical Guide for Applied Research*, 3rd ed., London, Sage Publications.

Laing, A. (2002) Marketing in the public sector: towards a typology of public services, *Marketing Theory*, 3(4), 427-445.

Lawder, K. (1989) Ratios 101: Back to the basis of financial analysis, *Business Credit*, June, 28-30.

Lee, J. H., Scott, D. and Floyd, M. F. (2001) Structural inequalities in outdoor recreation participation: A multiple hierarchy stratification perspective, *Journal of Leisure Research*, 33(4), 427-450.

Lentell, R. (2000) Untangling the tangibles: 'physical evidence' and customer satisfaction, *Managing Leisure: An International Journal*, 5(1), 1-16.

Liljander V. and Strandvik, T. (1997) Emotions in service satisfaction, *International Journal of Service Industry Management*, 8(2), 148-169.

LIRC (2002) *Sport England Benchmarking Service for Sports Halls and Swimming Pools- National Benchmarks 2001*, Unpublished report, Sheffield, Leisure Industries Research Centre (LIRC), University of Sheffield.

Loker, L.E. and Perdue, R. (1992) A benefit-based segmentation of a non-resident summer travel market, *Journal of Travel Research*, 31(1), 30-35.

Lovell, K. and Pastor J. (1999) Radial DEA models without inputs or without outputs, *European Journal of Operational Research*, 118, 46-51.

Luna-Arocas, R. and Tang, L.P. (2005) The use of cluster analysis to segment clients of a sport centre in Spain, *European Sport Management Quarterly*, 5(4), 381-413.

Lyons, D.M. (1995) *Performance measurement in urban transit: A comparative analysis of single and partial measures of transit performance*, PhD thesis, MD,

College Park, The University of Maryland.

MacKay, K. and Crompton, J. (1989) Users' perceptions of the relative importance of service quality dimensions in selected public recreation programs, *Leisure Sciences*, 11(4), 367-375.

Magd, H. and Curry, A. (2003) Benchmarking: achieving best value in public-sector organisations, *Benchmarking: An International Journal* 10(3), 261-286.

Mahony, D.F., Madrigal, R. and Howard, D. (2000), Using the psychological commitment to team (PCT) scale to segment sport consumers based on loyalty, *Sport Marketing Quarterly*, 9(1), 15-25.

Martilla, J. and James, J. (1977) Importance-performance analysis, *Journal of Marketing*, 41(1), 77–79.

McAdam, R. and O'Neill, L. (2002) Evaluating best value through clustered benchmarking in UK, *The International Journal of Public Sector Management*, 15(6), 438-457.

McAdam, R. and O'Neill, L. (2002) Evaluating best value through clustered benchmarking in UK local government: building control services, *The International Journal of Public Sector Management* 15(6), 438-457.

McDougall, D.H.G. and Levesque, T.J. (1994) Benefit segmentation using service quality dimensions: an investigation in retail banking, *International Journal of Bank Marketing*, 12(2), 15-23.

Metzger, L. M. (1993) Measuring quality cost effects on productivity using data envelope analysis, *Journal of Applied Business Research*, 9(3), 69-81.

Miles, M. and Huberman, A. M. (1994) *Qualitative Data Analysis: An Expanded Sourcebook*, 2nd ed., London, Sage.

Mintel (2005) *Health and Fitness Clubs-UK*, Market research report, London, Mintel.

Mintel (2006) *Leisure Centres and Swimming Pools-UK*, Market research report, London, Mintel.

Mori (2002) *Public Service Reform: Measuring and Understanding Customer*

Satisfaction, Mori review, London, Mori.

Mullin, B., Hardy, S., and Sutton, W. (2000) *Sport Marketing*, IL, Champaign, Human Kinetics.

Murray, D. and Howat, G. (2002) The relationships among service quality, value, satisfaction, and future intention of customers at an Australian sports and leisure centre, *Sport Management Review*, 5, 25-43.

Neely, A. and Bourne, M. (2000) Why measurement initiatives fail? *Measuring Business Excellence*, 4(4), 3-6.

Neely, A., Mills, J., Platts, K., Richards, H., Gregory, M., Bourne, M. and Kennerley, M. (2000) Performance measurement system design: developing and testing a process-based approach, *International Journal of Operations and Production Management*, 20(10), 1119-1145.

Neely, A.., Richards, A.H., Mills, J.F., Platts, K.W. and Bourne, M.C.S. (1997) Designing performance measures: a structured approach, *International Journal of Operations and Production Management*, 17(11), 1131-1153.

Nicholls, S. (2001) Measuring the accessibility and equity of public parks: a case study using GIS, *Managing Leisure: An International Journal*, 6, 201–219.

Novatorov, E.V. and Crompton, J.L. (2001) Reformulating the conceptualization of marketing in the context of public leisure services, *Leisure Studies*, 20, 61–75.

O'Neill, M. and Palmer, A. (2004) Importance-performance analysis: a useful tool for directing continuous quality improvement in higher education, *Quality Assurance in Education*, 12(1), 39-52.

Ogden, M.S. and Wilson, P. (2000) Bridging the quality gaps: implementing benchmarking to deliver best value, *Public Management: An International Journal of Research and Theory* 2 (4), 525-46.

Ogden, S. and Booth, D. (2001) Benchmarking of public leisure services: A tool for efficiency, effectiveness or equity? in: LSA (2001), *Leisure and Social Inclusion: New Challenges for Policy and Provision*, 73, 103-122.

Ogden, S. and Wilson, P. (2001) Beyond data benchmarking: the challenge of managing a benchmarking network in the UK public leisure sector, *Managing*

Leisure 6(2), 95-108.

Oh, H. (2001) Revisiting importance-performance analysis, *Tourism Management*, 22(6), 617–627.

Oh, H. and Parks, C. (1997) Customer satisfaction and service quality: a critical review of the literature and research implications for the hospitality industry, *Hospitality Research Journal*, 20(3), 36-64.

Ozcan, Y. A. and McCue, M. J. (1996) Development of a financial performance index for hospitals: DEA approach. *The Journal of Operational Research Society*, 47(1), 18-26.

Paradi, J. and Schaffnit, C. (2004) Commercial branch performance evaluation and results communication in a Canadian bank—a DEA application, *European Journal of Operational Research*, 156, 719–735.

Parasuraman, A., Zeithaml, V. and Berry, I. (1985) A conceptual model of service quality and implications for future research, *Journal of Marketing*, 49(3), 41–50.

Pitt, L., Morris, M.H. and Oosthuizen, P. (1996) Expectations of service quality as an industrial market segmentation variable, *Service Industries Journal*, 16(1), 1-9.

Ramanathan, R. (2003) *An Introduction to Data Envelopment Analysis*, New Delhi, Sage Publications.

Rapoport, R. (1970) Three dilemmas of action research, *Human Relations*, 23(6), 499-513.

Rawls (1971) *A Theory of Justice*, London, Oxford University Press.

Rickards, R.C. (2003) Setting benchmarks and evaluating balanced scorecards with data envelopment analysis, *Benchmarking: An International Journal* 10(3), 226-245.

Robinson, L. (1999) Following the quality strategy: the reasons for the use of quality management in UK public leisure facilities, *Managing Leisure: An International Journal*, 4(4), 201-217.

Robinson, L. (2002) Is quality management appropriate for public leisure services?

Managing Leisure: An International Journal, 7(1), 33-40.

Robinson, L. (2003) Committed to quality: the use of quality schemes in UK public leisure services, *Managing Service Quality*, 13(3), 247-255.

Robinson, L. (2004) Public leisure facilities: managing customer expectations, *Municipal Engineer*, 157(2), 129-133.

Robinson, L. (2006) Customer expectations of sport organisations, *European Sport Management Quarterly*, 6(1), 67-84.

Robinson, L. and Taylor, P. (2003) The performance of local authority sports halls and swimming pools in England, *Managing Leisure: An International Journal*, 8(1), 1-16.

Rohm, A.J., Milne, G.R. and McDonald M.A. (2006) A mixed-method approach for developing market segmentation typologies in the sports industry, *Sport Marketing Quarterly*, 15, 29-39.

Root-Bernstein, R. S. (1989) *Discovering: Inventing and Solving Problems at the Frontiers of Scientific Knowledge*, MA, Cambridge, Harvard University Press.

Sarrico, C.S. and Dyson, R.G. (2004) Restricting virtual weights in data envelopment analysis, *European Journal of Operational Research*, 159, 17-34.

Scheel, H. (2000) *EMS: Efficiency Measurement System*, available at: http://www.wiso.uni-dortmund.de/lsfg/or/scheel/ems, accessed 15 December 2005.

Schefczyk, M. (2003) Industrial benchmarking: A case study of performance analysis techniques, *International Journal of Production Economics*, 32, l-l 1.

Sharma, A. and Lambert, D.M. (1994) Segmentation of markets based on customer service, *International Journal of Physical Distribution and Logistics Management*, 24(4), 50-58.

Singleton, Royce A., and Bruce C. Straits. (1999) *Approaches to Social Research*, 3rd ed., New York, Oxford University Press.

Skivington, M., Christie, M. and Young G. (2003) Disabled and disadvantaged? *Recreation*, June, 28-31.

Smith J (1995) Semi-structured interviewing and qualitative analysis, in: Smith JA, Harre R, Van Langengrove L eds. *Rethinking Methods in Health Psychology*, London, Sage.

Smith, P. (1990) Data envelopment analysis applied to financial statements, *OMEGA*, 18, 131-138.

Smith, P. C. and Street, A. (2005) Measuring the efficiency of public services: the limits of analysis, *Journal of Royal Statistical Society A*, 168(2), 401-417.

Smith, W.R. (1956) Product differentiation and market segmentation as alternative marketing strategies, *Journal of Marketing*, July, 3-8.

Sport England (1999a) *Best Value through Sport: A survey of Sports Halls and Swimming Pools in England*, London, Sport England.

Sport England (1999b) *Best Value through Sport: The Value of Sport*, London, Sport England.

Sport England (2000a) *Performance Measurement for Local Authority Sports Halls and Swimming Pools*, London, Sport England.

Sport England (2000b) *Sports Participation and Ethnicity in England: National Survey 1999/2000*, London, Sport England.

Sport England (2001a) *Performance Measurement for the Development of Sport: A Good Practice Guide for Local Authorities*, London, Sport England.

Sport England (2001b) *Disability Survey 2000: Young People with a Disability*, London, Sport England.

Sport England (2001c) *National benchmarking services for sports halls and swimming pools*, London, Sport England.

Sport England (2002a) *Participation in Sport in England: Sport Equity Index 2002*, London, Sport England.

Sport England (2002b) *Adults with a Disability and Sport: National Survey 2000-2001*, London, Sport England.

Sport England (2004) *Participation in Sport: Results of the General Household*

Survey 2002, London, Sport England.

Sport England (2006a) *Understanding Participation in Sport: What Determines Sports Participation among Recently Retired People?* London, Sport England.

Sport England (2006b) *No limits: Sport England's Equity Policy*, London, Sport England.

Sport England (2006c) *Comprehensive Performance Assessment*, available at: http://www.sportengland.org/index/get_resources/local_authorities/comprehensive_performance_assessment.htm, accessed 10 August 2006.

Spottiswoode, C. (2000) *Improving Police Performance: A New Approach to Measuring Police Efficiency*, London, HM Treasury Public Services Productivity Panel.

Staat, M. and Maik, H. (2000) Benchmarking the health sector in Germany: an application of data envelopment analysis, *Research Paper*, Institute for Market-Oriented Management, Mannheim, University of Mannheim.

Stake, R. (1994) Case studies, in N. K. Denzin and Y. S. Lincoln eds., *Handbook of Qualitative Research*, London, Sage.

Stone, M. (2002) Can public service efficiency measurement be made a useful tool of government? Lessons from the Spottiswoode Report, *Public Money Management*, 22, 33-40.

Strauss, A. and Corbin, J. (1998) *Basics of Qualitative Research*, London, Sage.

Stringer, E.T. (1996) *Action Research: A Handbook for Practitioners*, London, Sage.

Sueyoshi, T and Aoki, S, (2001) A use of a nonparametric statistic for DEA frontier shift: The Kruskal and Wallis rank test, *Omega*, 29(1), 1-18.

Susman, G.I. and Evered, R.D. (1978) An assessment of the scientific merits of action research, *Administrative Science Quarterly* 23, 582-603.

Syrjänen, M.J. (2004) Non-discretionary and discretionary factors and scale in data envelopment analysis, *European Journal of Operational Research*, 158, 20-33.

Taks, M. and Scheerder. J. (2006) Youth sports participation styles and market segmentation profiles: evidence and applications, *European Sport Management*

Quarterly, 6(2), 85-121.

Tarrantm M. and Smith, E. (2002) The use of a modified importance-performance framework to examine visitor satisfaction with attributes of outdoor recreation settings, *Managing Leisure: An International Journal*, 7(2), 69-82.

Tashakkory, A. and Teddlie, C. (1998) *Mixed Methodology: Combining Qualitative and Quantitative Approaches*, London, Sage.

Taylor, P. (2004) The power of positive thinking, *Recreation*, December, 35-37.

Taylor, P. and Godfrey, A. (2003) Performance measurement in English local authority sport facilities. *Public Performance and Management Review*, 26(3), 251-262.

Taylor, P., Robinson, L., Kung, S., Gratton, C., Kokolokakis T. and Bovaird, A. (2000) *Performance Measurement for Sport Halls and Swimming Pools*, London, Sport England.

Teas, R. (1993) Expectation, performance evaluation and consumers' perceptions of quality, *Journal of Marketing*, 57(4), 18–34.

Thanassoulis, E., Boussofiane, A. and Dyson, R.G. (1996) A comparison of data envelopment analysis and ratio analysis as tools for performance assessment, *Omega*, 24(3), 229-244.

Tomas, S., Scott, S. and Crompton, J. (2002) An investigation of the relationships between quality of service performance, benefits sought, satisfaction and future intention to visit among visitors to a zoo, *Managing Leisure: An International Journal*, 7(4), 239-250.

Tsitskari, E. Tsiotras, D. and Tsiotras, G. (2006) Measuring service quality in sport services, *Total Quality Management*, 17(5), 623–631.

Vaughn, S., Schumm, J. S. and Sinagub, J. M. (1996) *Focus Group Interviews in Education and Psychology*, London, Sage.

Veal, A.J. (2002) *Leisure and Tourism Policy and Planning*, 2nd edition, Oxford, CABI International.

Voss, C. A., and P. Ahlstrom (1997) Benchmarking and operational performance:

some empirical results, *International Journal of Operations and Production Management*, 17, 1046-1058.

Walsh, K. (1991) Citizens and customers: marketing and public sector management, *Public Money and Management*, June-August, 9–16.

Walsh, K. (1994) Marketing and public sector management, *European Journal of Marketing*, 28(3), 63-71.

Webster, C. (1989) Can customers be segmented on the basis of their service quality expectations? *Journal of Services Marketing*, 3(2), 35-53.

Wedel, M., and Kamakura, W. A. (1998) *Market Segmentation: Conceptual and Methodological Foundations*, Boston, Kluwer Academic Publishers.

Weinstein, A. (1994) *Market segmentation: Using Demographics, Psychographics and Other Niche Marketing Techniques to Predict and Model Customer Behaviour*, Chicago, Irwin

Westbrook, R. (1995) Action research: a new paradigm for research in production and operations management, *International Journal of Operations and Production Management*, 15(12), 6-20.

Whittington, G. (1980) Some basic properties of accounting ratios, *Journal of Business Finance and Accounting* 7, 219-232.

Williams, C. (1998) Is the SERVQUAL model an appropriate management tool for measuring service quality in the UK leisure industry? *Managing Leisure: An International Journal*, 3(2), 98-110.

Wind, Y. (1978) Issues and advances in segmentation research, *Journal of Marketing Research*, 15, 317-337.

Wisniewski, M. (2001) Using SERVQUAL to assess customer satisfaction with public sector services, *Managing Service Quality*, 11(6), 380-388.

Woo, K.S. (1998) Using quality perceptions to segment customers in services, *Marketing Intelligence and Planning*, 16(7), 418–424.

Worthington, A. and Dollery, B. (2000) An empirical survey of frontier efficiency measurement techniques in local government, *Local Government Studies*, 26(2),

23-52.

Worthington, A.C., (1998) The application of mathematical programming techniques to financial statement analysis: Australian gold production and exploration, *Australian Journal of Management* 23(1), 97-114.

Yeh, Q. J. (1996) The application of data envelopment analysis in conjunction with financial ratios for bank performance evaluation, *Journal of the Operational Research Society*, 47, 980–988.

Yin, R. K. (1989) *Case Study Research: Design and Methods*, London, Sage.

Young, S. T. (1992) Multiple productivity measurement approaches for management, *Health Care Management Review*, 17(2), 51-58.

Yüksel, A. and Rimmington, M. (1998) Customer-satisfaction measurement, *Cornell Hotel and Restaurant Administration Quarterly*, 39(6), 60-70.

Zeithaml, V., Berry, L. L., and Parasuraman, A. (1993) The nature and determinants of customer expectations of services, *Journal of the Academy of Marketing Science*, 21(1), 1-12.

Zhu, J. (2002) *Quantitative Models for Performance Evaluation and Benchmarking: Data Envelopment Analysis with Spreadsheets and DEA Excel Solver*, Boston, Springer.

APPENDIX A

National Benchmarking Services

	Performance indicators	Dimensions
ACCESS	1. Youth	Equity
(12 PIs)	2. Disadvantaged groups	Equity
	3. Ethnic minorities	Equity
	4. Adults	Equity
	5. Older people	Equity
	6. First visits	Equity
	7. Discount card	Equity
	8. Discount card by DG	Equity
	9. Females	Equity
	10. Disabled < 60	Equity
	11. Disabled 60+	Equity
	12. Unemployed	Equity
FINANCIAL	1. Cost recovery	Efficiency
(15 PIs)	2. Subsidy per visit	Efficiency
	3. Subsidy per m^2	Economy
	4. Subsidy per hour	Economy
	5. Subsidy per resident	Economy
	6. Operating cost per visit	Efficiency
	7. Operating cost per m^2	Economy
	8. Operating cost per hour	Economy

	9. Maintenance cost per m^2	Economy
	10. Energy cost per m^2	Economy
	11. Income per visit	Effectiveness
	12. Income per m^2	Effectiveness
	13. Income per hour	Effectiveness
	14. Direct income per visit	Effectiveness
	15. Secondary income per visit	Effectiveness
UTILISATION (8 PIs)	1. Visit per m^2	Efficiency
	2. Visit per hour	Efficiency
	3. % casual visit	Effectiveness
	4. % unused program time	Effectiveness
	5. % unused unable program time	Effectiveness
	6. visiting hall as % of catchment population (cp)	Effectiveness
	7. visiting pools as % of cp	Effectiveness
	8. visiting other as % of cp	Effectiveness
CUSTOMER SERVICE * (25 PIs)	I. Accessibility	Effectiveness
	II. Availability of facilities	Effectiveness
	III. Quality of facility/services	Effectiveness
	IV. Cleanliness	Effectiveness
	V. Staff	Effectiveness
	VI. Value for money	Effectiveness

* Customer satisfactions are measured by 5-point Likert scale (from 'very unsatisfied' to 'very satisfied'). The 25 attributes can refer to p.243.

APPENDIX B

Aggregate Performance Analysis model

1. Introduction

Based on the performance data of Centre X and the 2001 database of Nation Benchmarking Service (NBS), the aim of this report is to provide an aggregate analysis of Centre X's performance. The analytic techniques used include data envelopment analysis, multivariate statistical analysis, gap analysis and importance-performance analysis. However, to facilitate better understanding, only the results of analysis are summarised in this report. This report consists of the following four parts of performance measurement:

- Operational efficiency: 'financial' and 'utilisation' in the NBS.
- Sport equity: 'access' in the NBS.
- Service quality: based on the user survey
- Customer segmentation: based on the user survey

2. Operational Efficiency

By adopting data envelopment analysis, efficiency is measured by selecting visits and income as the output variables and cost, facility area and opening hours as the input variables. It is assumed that output maximisation is the major concern of management, i.e. the maximisation of either visits or income. The aim is to provide benchmarking partners so that the strengths and weaknesses of Centre X can be identified and performance targets can be set. The results are shown in Table 1. By

adopting data envelopment analysis, eight benchmarking partners are identified. These eight centres have the highest similarity with Centre X in terms of facility profiles (i.e. the NBS families) and production structures (i.e. strengths and weaknesses). The eight original NBS efficiency PIs of Centre X and its benchmarking partners are shown to identify the relative strengths and weaknesses.

• Visits (strength): Centre X outperforms four of the eight benchmarking partners and 'visits per sq. m.' is in the highest quartile, so visits can be regarded as Centre X's strength.

• Subsidy (weakness): Centre X's performance is generally worse than its benchmarking partners in the three subsidy-related PIs, so subsidy is regarded as Centre X's weakness.

• Income (strength): Centre X outperforms three of the eight benchmarking partners in two income-related PIs. Also, 'income per sq. m.' and 'income per visit' are in the 4th and 3rd quartile respectively. As such, income is deemed as Centre X's strength.

• Cost (weakness): Compared to income, Centre X's performance in cost is relative weak. Although Centre X outperforms four of the eight benchmarking partners in two cost-related PIs, these two PIs are located in lower quartiles- 3rd for 'cost per visit' and 2nd for 'cost per sq. m.'.

In sum, the relatively higher cost offsets the strength brought by higher income, and in turn, results in higher subsidy required. Finally, Centre A is suggested as the best benchmarking partner for Centre X as it resembles the most to Centre X in terms of facility profiles (DE 15-20% and trust) and has only one PI worse than Centre X (income per visit). While setting performance targets, the performance of Centre A

could be a good reference.

2. Sport Equity

Equity is measured by selecting the five proposed CPA PIs. The analytic approach and aim of this section is basically the same as the previous section. The only difference is that, the measurement is based on so-called effectiveness model where only outputs are taken into account. The results are shown in Table 2. The strengths and weaknesses of Centre X are identified by examining the quartiles and the PIs relative to eight benchmarking partners.

- Weaknesses: disabled < 60 years; social class DE and 11-19 years.
- Strengths: Ethnic minority and 60+ years.

Centre A is suggested as the best benchmarking partner for Centre X, followed by Centre B and C, since they have better performance in all of the five PIs and also have the most similar facility profiles.

Table 1 Efficiency measurement

Centre	DEA score	PIs: visits /sq. m. [a]	Subsidy /visit [a]	Subsidy /sq. m.	Cost recovery	Income /visit	Income /sq. m.	Cost /visit	Cost /sq. m.	Family: Type	Location (DE)	Size [b]	Management
Centre X	**0.79**	**120 (4^{th})**	**0.56 (3^{rd})**	**29 (3^{rd})**	**102 (4^{th})**	**2.3 (3^{rd})**	**275 (4^{th})**	**2.2 (3^{rd})**	**269 (2^{nd})**	**Dry**	**15-20%**	**Medium**	**Trust**
Centre A	1.00	142	-0.63	-90	144	~~2.1~~	296	1.5	206	Mixed	**15-20%**	Large	**Trust**
Centre B	1.00	123	-0.75	-92	116	5.5	669	~~4.7~~	~~578~~	Mixed	<15%	**Medium**	Commercial
Centre C	1.00	221	-0.17	-38	106	2.8	628	~~2.7~~	~~590~~	Mixed	<15%	Large	Commercial
Centre D	1.00	~~103~~	-0.72	-74	141	2.4	~~251~~	1.7	177	Mixed	20%+	Large	In-house
Centre E	1.00	~~110~~	-0.52	-57	134	~~2.1~~	~~226~~	1.5	169	Mixed	20%+	Large	Commercial
Centre F	0.89	~~112~~	0.22	-24	108	2.9	319	~~2.6~~	~~295~~	Mixed	**15-20%**	Large	Commercial
Centre G	0.88	158	-0.02	3	~~99~~	2.3	361	~~2.3~~	~~364~~	Mixed	**15-20%**	Large	In-house
Centre H	0.87	~~109~~	-0.21	-23	111	~~2.1~~	~~224~~	1.9	202	Wet	20%+	Small	**Trust**

Note: Number with bracket is quartile adopted by the NBS (4^{th} is the highest and 1^{st} is the lowest); number with delete line means that Centre X has a better performance than its benchmarking partners.

[a] CPA PIs

[b] Small= <1,500 sq. m.; medium= 1,500 to 3,000 sq. m.; large= 3,000+ sq. m.

Table 2 Equity measurement

Centre	DEA score	PIs					Family			
		Social class DE	11-19 years	Ethnic minority	60+ years	Disabled <60 years	Type	Location (DE)	Size	Management
Centre X	**0.68**	**0.20 (2^{nd})**	**0.60 (2^{nd})**	**2.60 (3^{rd})**	**0.50 (3^{rd})**	**3.30 (1^{st})**	**Dry**	**15-20%**	**Medium**	**Trust**
Centre A	0.82	0.37	1.14	2.89	0.70	6.57	Wet	**15-20%**	**Medium**	**Trust**
Centre B	0.93	0.32	0.96	4.77	0.76	4.72	Mixed	**15-20%**	**Medium**	Other
Centre C	0.81	0.28	0.55	3.01	0.64	9.57	Mixed	**15-20%**	**Medium**	Commercial
Centre D	1.00	0.39	0.93	8.38	~~0.32~~	5.99	Wet	**15-20%**	**Medium**	Commercial
Centre E	0.72	0.30	1.54	~~0.21~~	0.63	3.69	Mixed	**15-20%**	**Medium**	In-house
Centre F	0.82	0.32	1.83	~~2.53~~	~~0.08~~	8.23	**Dry**	**15-20%**	**Medium**	Other
Centre G	0.84	0.40	~~0.54~~	~~0.84~~	0.72	9.39	Wet	**15-20%**	Large	**Trust**
Centre H	0.79	0.47	~~0.45~~	~~2.53~~	0.54	7.51	Wet	20%+	**Medium**	**Trust**

Note: Number with bracket is quartile adopted by the NBS (4^{th} is the highest and 1^{st} is the lowest); number with delete line means that Centre X has a better performance than its benchmarking partners.

4. Service quality

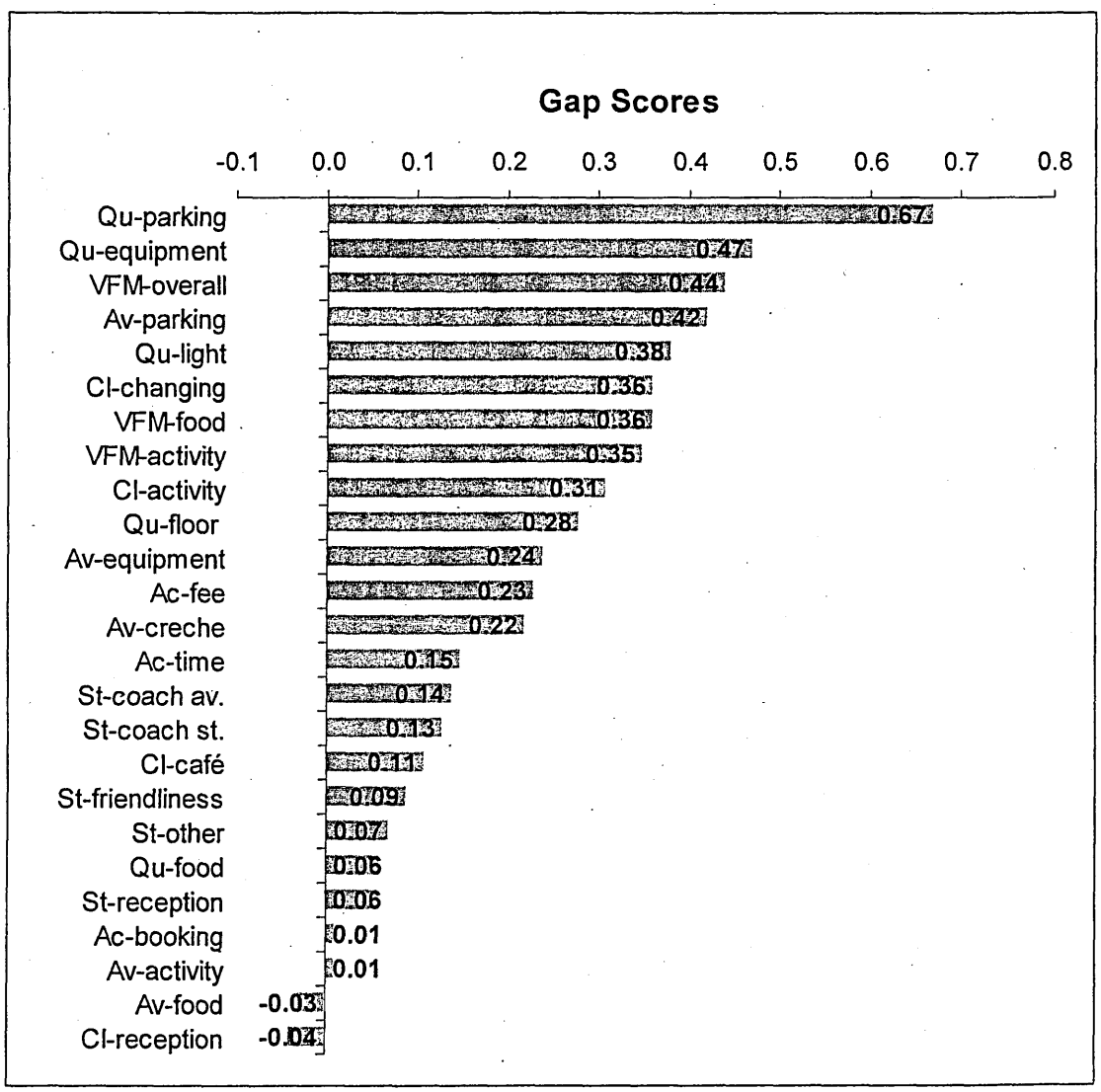
Customer service quality is measured by the following three approaches:

• Gap analysis (Figure 1): Gap analysis is conducted to identify the absolute gap of each attribute. Gap scores are determined by subtracting the 'satisfaction' scores from the 'importance' scores. A positive gap score reveals that the performance rating is lower than the importance rating, whereas a negative gap score indicates a good satisfaction relative to the importance rating. 'Quality of car parking', 'quality of equipment' and 'overall value for money' are the top three attributes with the largest service quality gaps. By contrast, 'cleanliness of reception' and 'availability of foods/drinks' have negative gap scores indicating good performance.

• Importance-performance analysis (IPA) (Figure 2): While gap analysis focuses more on the absolute gap of each attribute, IPA combines all attributes' importance and performance ratings into a two-dimensional grid in an effort to ease data interpretation and provides a picture of all selected attributes' relative weight. The centre of this grid where the two axes intersect represents the mean value of 25 attributes. Grid analysis generates four different suggestions based on four quadrants in the grid. Theoretically, the centre needs to focus on improving its performance on the attributes located in fourth quadrant (the bottom right hand), where the importance is high but the performance is low. Generally speaking, Centre X has better performance in 'staff' and 'accessibility' because most attributes of these two dimensions are located in the first quadrant where importance and performance are both high. On the other hand, although services related to catering and crèche have lower performance levels but they are also less important to the customers. The areas where Centre X needs to focus are: 'quality of equipment', 'quality of light in sports

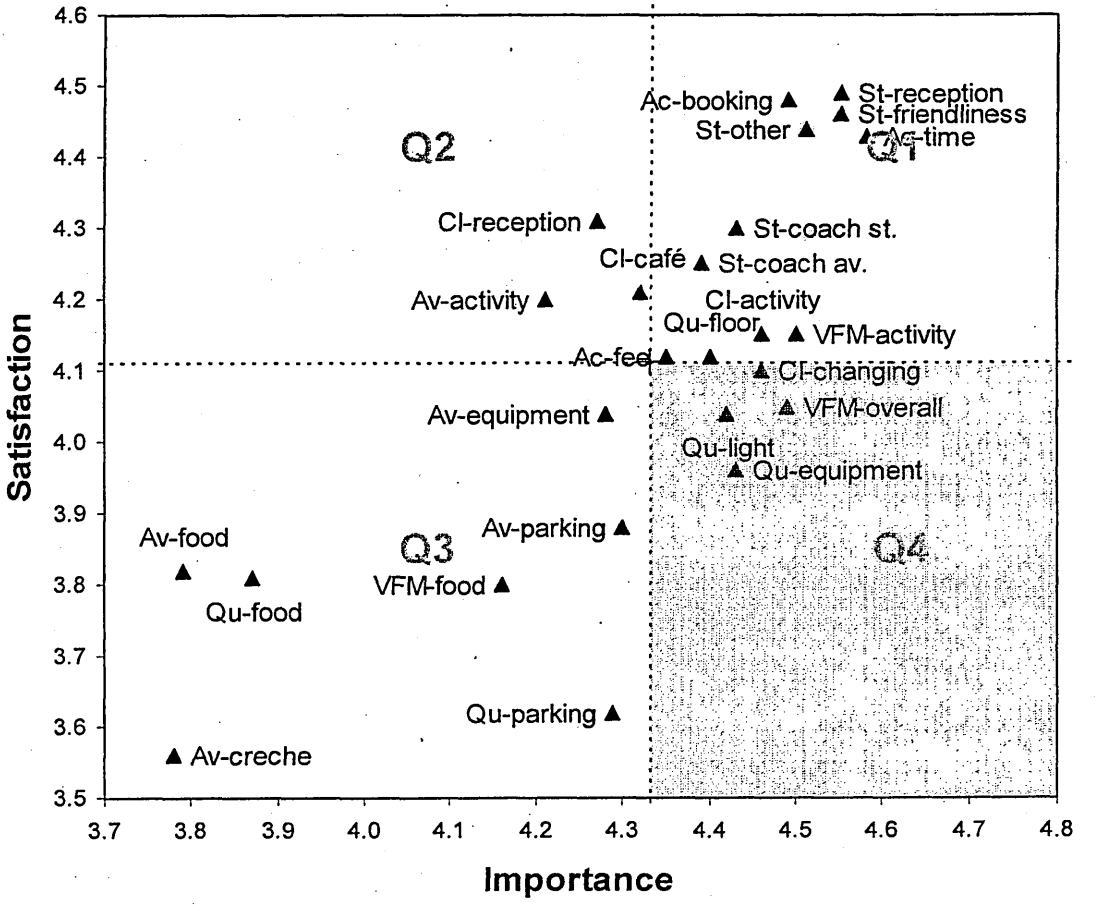
hall', 'overall value for money' and 'cleanliness of changing area'. In addition, although the availability and quality of car parking are located in the third quadrant, they are quite close to the fourth quadrant; especially the satisfaction to the quality of car parking is relatively low. Therefore, they are also important.

Figure 1 Gap analysis



Note: The codes used are explained in the appendix.

Figure 2 Importance-performance analysis



In sum, the strengths and weaknesses of Centre X are deemed as:

• Strengths: catering, staff and accessibility (but the accessibility is worse than the industrial average).

• Weaknesses: Value for money, facility quality (parking, equipment and light) and cleanliness of changing area.

5. Customer segmentation

Based on customer's perceived importance of service quality, this section attempts to identify customer segmentation by adopting multivariate statistical analysis. The result may help managers to develop specific promotional or quality improvement schemes to meet the needs of targeted customer segments. As indicated in the previous section, the 25 service attributes are grouped into five service dimensions. Then three customer segments with different priorities are presented (as shown in Table 3).

Table 3 Segment priorities

Dimensions	Mean importance scores							
	Overall		Segment 1 (48%)		Segment 2 (18%)		Segment 3 (34%)	
Accessibility	4.55	(1^{st})	4.83	(2^{nd})	3.73	(3^{rd})	4.64	(1^{st})
Staff	4.51	(2^{nd})	4.92	(1^{st})	4.06	(1^{st})	4.20	(3^{rd})
Facility & Cleanliness	4.43	(3^{rd})	4.82	(3^{rd})	3.66	(4^{th})	4.31	(2^{nd})
Value for money	4.38	(4^{th})	4.81	(4^{th})	3.77	(2^{nd})	4.17	(4^{th})
Catering	3.96	(5^{th})	4.59	(5^{th})	2.95	(5^{th})	3.68	(5^{th})

Note: The number with bracket is the mean ranking across factors.

• Segment 1 (48%): This segment represents the largest sample of the respondents. Customers in this segment score the highest on 'staff', followed by 'accessibility' and 'facility quality and cleanliness'.

• Segment 2 (18%): This segment represents the smallest sample of the respondents. Similar to segment 1, 'staff' is the most important dimension to the customers. However, it differs from segment 1 in that 'value for money' is also

important to the customers.

• Segment 3 (34%): This segment is characterised by the highest importance rating on 'accessibility', followed by 'facility quality and cleanliness' and 'staff'.

To provide practical information to formulate strategy, the three customer segments are cross-tabulated against customers' demographic profiles, including gender, ethnicity, age and occupation (as shown in Table 4). To avoid problems with different sample sizes across segments and among profiles, the data in Table 4 are normalised by using 'representativeness' ratio, where the percentage of customers with different characteristics in one segment is compared to the share of that segment in the whole sample. A ratio over 1.0 indicates well-represented and less than 1.0 indicates under-represented. For example, '<19 years' group is well-represented in segment 3, with a ratio of 1.4, meaning that the percentage of '<19 years' in segment 3 (47.37%) is 39% higher than the segment percentage as a whole (34.15%), i.e. 47.37% ÷ 34.15%= 1.39.

• Segment 1: It tends to group more '60+ years' (1.20) and 'retired' (1.40).

• Segment 2: It tends to group more '60+ years' (1.17), 'retired' (1.21) and unemployed (1.19). By contrast, the segment constitutes very few '<19 years' (0.29) and 'student' (0.32).

• Segment 3: 'Unemployed' (2.18) is strongly represented in this segment, followed by '<19 years' (1.39) and 'student' (1.38). By contrast, 'retired' (0.33) and '60+ years' (0.43) are relatively under-represented in this segment.

In sum, 'staff' is the major concern of older people, whereas 'accessibility' is the major concern of younger and unemployed people. There are no significant differences between segments for other targeted groups, such as female and ethnic

minority.

Table 4 Segment profiles

	Profiles		Segment 1 (48%)	Segment 2 (18%)	Segment 3 (34%)
Gender	Male	51%	0.91	1.09	1.09
	Female	49%	1.11	0.90	0.90
Ethnicity	White	88%	0.97	1.04	1.03
	Ethnic minority	12%	1.10	0.89	0.79
Age	<19 years	7%	1.00	0.29	1.39
	19-59 years	83%	0.98	1.04	1.00
	60+ years	10%	1.20	1.17	0.43
Occupation	Student	7%	0.99	0.32	1.38
	Work FT	62%	1.01	1.02	0.97
	Work PT	13%	0.74	1.33	1.19
	Retired	8%	1.40	1.21	0.33
	Unemployed	2%	0.53	1.19	2.18
	Housewife	8%	1.05	0.91	0.98

Abbreviations

Code	Attribute
Ac-time	= Activity available at convenient times
Ac-booking	= Ease of booking
Ac-fee	= The activity charges/fee
Av-parking	= Availability of car parking on site
Av-creche	= Availability of creche facilities
Av-food	= Availability of food and drink
Av-activity	= The range of activities available
Av-equipment	= Availability of equipment
Qu-floor	= Quality of flooring in the sports hall
Qu-light	= Quality of lighting in the sports hall
Qu-equipment	= Quality of equipment
Qu-food	= Quality of food and drink
Qu-parking	= Safety and security of car parking
Cl-reception	= Cleanliness of reception area
Cl-changing	= Cleanliness of changing area
Cl-activity	= Cleanliness of activity spaces
Cl-café	= Cleanliness of cafe/bar
St-reception	= Helpfulness of reception staff
St-other	= Helpfulness of other staff
St-friendliness	= Friendliness of staff
St-coach av.	= A vailability of coaching/instruction
St-coach st.	= Standard of coaching/instruction
VFM-activity	= Value for money of activities
VFM-food	= Value for money of food/drinks
VFM-overall	= Value for money of overall

APPENDIX C

Interview Questions

Background

Q1: Can you tell me a little about your background, like your position, how long have you been in this position?

Q2: How long have you adopted the NBS? How have you used the NBS results, e.g. reporting or performance management?

Q3: What are your own expectations of performance measurement? What effect do you expect performance measurement to have?

Importance and quality of model

Q4: What do you think about the importance of overall performance measurement and the identification of benchmarking partners?

Q5: To what extent does the model provide incremental and useful information which is not readily available from NBS type evidence?

Q6: Did you understand the content of the model? Has anyone ever questioned its practicability?

Actual use of the model

Q7: How does the model compare with the previous use of the NBS?

Q8: Have you used the information provided by the model for decisions or communications? If not, why not? If yes, what has changed as a result of using the model in your centre?

Q9: Overall, what effect do you think that using the model has had or has the potential to have?

Factors hinder or help in the use of the model

Q10: What are the greatest problems that have to be overcome to understand and use the model?

Q11: What are the greatest problems that have met to communicate the results in the report to other staff?

Q12: If you were making a list of lessons learned from the use of model, what would be on the list? What would facilitate greater understanding and more use of such a model?