Expert Systems with Applications 41 (2014) 3761–3768

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Benchmarking of service quality with data envelopment analysis

Hakyeon Lee [a], Chulhyun Kim [b,*]

[a] Department of Industrial and Systems Engineering, Seoul National University of Science and Technology, 172 Gongreung 2-dong, Nowon-gu, Seoul 139-746, Republic of Korea
[b] Department of Technology and Systems Management, Induk University, Choansan-ro 14, Nowon-gu, Seoul 139-050, Republic of Korea

ARTICLE INFO

Keywords:
Benchmarking
Service quality
Data envelopment analysis (DEA)
SERVQUAL
SERVPERF

ABSTRACT

This paper proposes a data envelopment analysis (DEA) approach to measurement and benchmarking of service quality. Dealing with measurement of overall service quality of multiple units with SERVPERF as multiple-criteria decision-making (MCDM), the proposed approach utilizes DEA, in particular, the pure output model without inputs. The five dimensions of SERVPERF are considered as outputs of the DEA model. A case study of auto repair services is provided for the purpose of illustration. The current practice of benchmarking of service quality with SERVQUAL/SERVPERF is limited in that there is little guidance to whom to benchmark and to what degree service quality should be improved. This study contributes to the field of service quality benchmarking by overcoming the above limitations, taking advantage of DEA's capability to handle MCDM problems and provide benchmarking guidelines.

1. Introduction

Service quality has consistently been at the core of research into service industries, since it is recognized as a critical determinant of business performance and a strategic tool for firms wishing to gain long-term viability (Gale, 1994). The prerequisite for achieving a high level of service quality is to be able to measure it. During the past two decades, determining the best way to measure service quality has been a matter of concern for both practitioners and researchers. There is an extensive body of knowledge on measuring service quality, which has also been a continued focus of research in terms of definition, typology, models, and operationalization (Seth, Deshmukh, & Vrat, 2005).

Unquestionably, the most popular measure of service quality is SERVQUAL developed by Parasuraman, Zeithaml, and Berry (1988). SERVQUAL is a multi-item instrument for measuring service quality based on the gap model, in which service quality is a function of the difference between perception and expectation (Parasuraman, Zeithaml, & Berry, 1985). SERVQUAL has enjoyed a number of applications in a variety of settings, but many researchers have also tackled its operationalization (Carrillat, Jaramillo, & Mulki, 2007). In an effort to address deficiencies in SERVQUAL, Cronin and Taylor (1992) developed the SERVPERF instrument, which uses customers' perceived performance as a direct measure of service quality.

Regardless of whether SERVQUAL or SERVPERF is used, what is of particular interest in this study is analysis of survey data in measuring service quality. The original instrument of both SERVQUAL and SERVPREF (henceforth, SERVQUAL/SERVPERF) comprises of five dimensions with 22 items (44 items in SERVQUAL, for measuring both perception and expectation). Analysis of SERVQUAL/SERVPERF can take several forms, including item-by-item analysis, dimension-by-dimension analysis, or computation of a single measure of overall service quality (Buttle, 1996). The single measure can also be obtained in various ways, such as a simple sum or average, a weighted sum, or a weighted average, with weights assigned to each dimension or item. One of the primary reasons for producing a single measure of overall service quality across dimensions is to enable benchmarking through comparison. One of the practical values of SERVQUAL/SERVPERF lies in its ability to establish best practices by comparing overall quality scores of service units and then to improve the performance of units that are falling behind (Camp, 1989; Kettinger & Lee, 1997). However, benchmarking based on a simple aggregated measure has the limitation that there is little guidance to whom to benchmark and to what degree service quality should be improved.

To address this limitation, this paper proposes a data envelopment analysis (DEA) approach to computation of a single measure of overall service quality and benchmarking in measuring service quality with the five dimensions of SERVQUAL/SERVPERF. DEA is a linear programming model for measuring the relative efficiency of decision-making units (DMUs) with multiple inputs and outputs (Cooper, Seiford, & Tone, 2000). DEA has various advantages: it can handle multiple inputs and outputs; it does not require prescribed functional forms of production as well as prescribed weights to be attached to each input and output. The greatest merit of DEA is that it provides benchmarking guidelines for inefficient DMUs (Paradi & Zhu, 2013). For each inefficient DMU, DEA identifies a set of

* Corresponding author. Tel.: +82 2 970 6469, fax: +82 2 974 2849.
E-mail address: stddevs@induk.ac.kr (C. Kim).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.12.008

efficiency units called the reference set, which constitutes its benchmark, with information on how much should be improved to be efficient. Thus, producing a single measure of overall service quality with DEA automatically draws up guidelines about how to conduct service quality benchmarking in terms of each dimension.

DEA has widely and successfully been employed to measure performance across various service industries, mainly in banking, health care, transportation, and education (Liu, Lu, Lu, & Lin, 2013). Most early studies have paid attention to the operational efficiency or profitability of service units, rather than the quality aspect (Soteriou & Stavrinides, 1997). Recent studies have started to look at quality measures as outputs, such as grades in education (Olesen & Petersen, 1995), the ratio of actual deaths to predicted deaths in healthcare (Morey, Fine, Loree, Retzlaff-Roberts, & Tsubakitani, 1992), and the number of satisfied customers in the airline industry (Adler & Berechman, 2001). Some researchers have attempted to measure both quality efficiency and operating efficiency using DEA (Soteriou & Stavrinides, 1997; Kamakura, Mittal, de Rosa, & Mazzon, 2002; Sherman & Zhu, 2006; Shimshak, Lenard, & Klimberg, 2009). However, the quality measures utilized in those studies are merely proxies, not direct indicators of service quality. This study adopts direct measures of service quality as the output variables of DEA, the five dimensions of SERVQUAL/SERVPERF. Another problem of the previous studies considering service quality variables as outputs of DEA is its implicit assumption that inputs are transformed into quality output measures. However, it does not make sense that quality-type outputs, as opposed to quantitative physical outputs, increase with additional inputs (Salinas-Jimenez & Smith 1996; Shimshak et al., 2009). To deal with the lack of connection between amount of inputs and quality-type outputs, this study adopts the pure output DEA model proposed by Lovell and Pastor (1997). The pure output model only includes several outputs without input variables.

The remainder of this paper is organized as follows. Section 2 reviews the theoretical background of service quality measurement, with a focus on SERVQUAL/SERVPERF. DEA models are explained in Section 3. Section 4 shows how to apply DEA to measurement and benchmarking of service quality with a case study of the auto repair services. The paper ends with conclusions and directions for future research in Section 5.

2. Measuring service quality: SERVQUAL and SERVPERF

In the pioneering work by Parasuraman et al. (1985), ten dimensions of service quality were proposed with the gap model in which service quality is a function of the difference between perceptions and expectations of a service. Parasuraman et al. (1988) developed a scale composed of 22 items for measuring service quality, called SERVQUAL, in which the original ten dimensions of service quality are collapsed into five: tangibles, reliability, responsiveness, assurance, and empathy, as presented in Table 1. The SERVQUAL instrument includes 22 items for measuring expectations (E) and 22 corresponding items for measuring perceptions (P). For each item, based on the gap model, a quality score (Q) is obtained as the difference between the perception (P) and expectation (E) ratings; that is, $Q = P - E$.

The development of SERVQUAL has spawned a considerable amount of related research on its practical applications as well as theoretical discussions. A number of applications of SERVQUAL have been reported in a variety of settings (Ladhari, 2009), but it has also been criticized on theoretical and operational grounds (Jain & Gupta, 2004). The biggest issue that has been raised by many researchers is its operationalization—namely, the use of the gap score (Calvo-Porral, Lévy-Mangin, & Novo-Corti, 2013). Contrary to the original work by Parasuraman et al. (1988), the convergent validity of SERVQUAL has often not been confirmed in subsequent studies. Many studies have found that service quality measured with SERVQUAL is not significantly related to that measured directly through a single-item scale (Babakus & Boller, 1992; Carman, 1990). Van Dyke, Kappelman, and Prybutok (1997) insisted that separately measuring expected and perceived level of service quality and subtracting one score from the other is too simplistic to capture the complex cognitive process of perceiving service quality because one's perception of service quality already entails the expectation of a service. The conceptualization of "expectation" has also been under attack because it is subject to multiple interpretations (Teas, 1993; Teas, 1994).

Although Parasuraman, Zeithaml, and Berry (1994) provided thorough rebuttals to the critics on the use of gap scores, many researchers posit that a simple performance-based measure is a preferable means of measuring service quality (Babakus & Boller, 1992; Bolton & Drew, 1991; Cronin & Taylor, 1992; Cronin & Taylor, 1994). Raising fundamental criticisms against SERVQUAL, Cronin and Taylor (1992) proposed SERVPERF, which directly assesses customers' perceived performance. The SERVPERF instrument discards the expectation component and only includes 22 items for measuring performance (P). SERVPERF assumes that higher perceived performance implies higher service quality; that is, $Q = P$. Obviously, the SERVPERF scale is more efficient than the SERVQUAL scale because it reduces by half the number of items to be measured in SERVQUAL. Cronin and Taylor (1992) also empirically showed the theoretical superiority of the SERVPERF scale over the SERVQUAL scale.

Since the advent of SERVPERF, much vigorous debate has been taking place on whether SERVQUAL or SERVPERF should be used for measuring service quality. Numerous attempts have been made to compare the two scales on such various criteria as reliability, content validity, predictive validity, convergent validity, and diagnostic power (Babakus & Boller, 1992; Brady, Cronin, & Brand, 2002; Brown, Churchill, & Peter, 1993; Carrillat et al., 2007; Cui, Lewis, & Park, 2003; Hudson, Hudson, & Miller, 2004; Jain & Gupta, 2004; Kettinger & Lee, 1997; Mukherje & Nath, 2005; Quester & Romaniuk, 1997; Zhou, 2004). However, the question is still controversial, and general agreement does not exist about which is the better. Most researchers have upheld the idea that SERVEPRF is a better alternative than SERVQUAL in terms of validity and explanatory power (Babakus & Boller, 1992; Brady et al., 2002; Brown et al., 1993; Kettinger & Lee, 1997; Zhou, 2004) while Quester and Romaniuk (1997) support the superiority of SERVQUAL in

Table 1
Dimensions of SERVQUAL (Parasuraman et al., 1988).

Dimension	Definition	Number of items
Tangibles	Physical facilities, equipment, and appearance of personnel	4
Reliability	Ability to perform the promised service dependably and accurately	5
Responsiveness	Willingness to help customers and provide prompt service	4
Assurance	Knowledge and courtesy of employees and their ability to inspire trust and confidence	4
Empathy	Caring, individualized attention the firm provides its customers	5