Quality of service in public transport based on customer satisfaction surveys: A review and assessment of methodological approaches

By: Juan de Oña and Rocío de Oña

This document is a **post-print version** (ie final draft post-refereeing) of the following paper:

Juan de Oña and Rocío de Oña (in press) *Quality of service in public transport based on customer satisfaction surveys: A review and assessment of methodological approaches*. **Transportation Science,** DOI: 10.1287/trsc.2014.0544.

Direct access to the published version: http://dx.doi.org/10.1287/trsc.2014.0544

Quality of service in public transport based on customer satisfaction surveys: A review and assessment of methodological approaches

Juan de Oña* and Rocio de Oña

TRYSE Research Group. Department of Civil Engineering, University of Granada, ETSI Caminos, Canales y Puertos, c/ Severo Ochoa, s/n, 18071 Granada (Spain)

* Corresponding author. Phone: +34 958 24 99 79, jdona@ugr.es

Quality of service in public transport based on customer satisfaction surveys: A review and assessment of methodological approaches

ABSTRACT

The growth of literature in the field of quality of service in the public transport (PT) sector shows increasing concern for a better understanding of the factors affecting service quality (SQ) in PT organizations and companies. A large variety of approaches to SQ has been developed in recent years owing to the complexity of the concept; the broad range of attributes required to evaluate SQ; and the imprecision, subjectivity and heterogeneous nature of the data used to analyse it. Most of these approaches are based on customer satisfaction surveys. This paper seeks to summarize the evolution of research and current thinking as it relates to the different methodological approaches for SQ evaluation in the PT sector over the years, and provides a discussion of future directions.

Keywords: service quality; public transport; customer satisfaction surveys; derived importance; stated importance

1. INTRODUCTION

For a long time the performance evaluation of Public Transport (PT) has been carried out from the service managers' perspective, based on the cost efficiency and cost effectiveness of PT services and operations (e.g. Hensher and Daniels, 1995; Pullen, 1993). However, in the last few decades, Service Quality (SQ) has become a major area of attention for practitioners, managers and researchers, who have focused on the passengers' perspective.

Currently, researchers and managers in the PT sector strive to learn details about the main factors affecting SQ in their organizations for the obvious reasons of customer satisfaction, increased profitability, etc. An on-going enhancement of SQ represents an essential tool for transit agencies and transport planners in order to capture and retain more passengers. In fact, offering high quality transit services will encourage a modal shift from private modes to PT services and, consequently, it will promote a more sustainable mobility. SQ measures help transport managers to establish their strategic goals and to determine funding decisions. In this context, models gain specific importance as they not only help to learn the factors associated with SQ but also provide a direction for improvements.

Many authors have studied SQ in the PT sector from varying perspectives, using a range of different methodologies in recent years. The variety of existing approaches could be justified by the complexity of the SQ concept; the number of attributes used to evaluate it; the imprecision and subjectivity of the data used to analyse it, typically based on customer satisfaction surveys (CSS); and the heterogeneity of passenger perceptions.

The beginning of the 21st century saw an increase in the use of discrete choice models based on Stated Preference (SP) surveys (e.g. Hensher and Prioni, 2002; Hensher et al., 2003) to analyse PT service quality. Such methods are based on the assumption that although specific aspects of SQ may be particularly positive or negative in a passenger's satisfaction with a service, the overall level of passenger satisfaction is best measured by how an individual evaluates the total package of services on offer.

Nonetheless, models based on CSS have been and are the most widely adopted for analysing SQ in the PT sector. So, the aim of this research is to provide a review of contemporary thinking on PT quality of service-analysis field based on CSS and to highlight the main methodological approaches that have been used to address this issue. To this end, a varied amount of studies have been collected, ordered, categorized and explained seeking to understanding the reasons behind the varied and often complicated evaluation methods used in the transport sector, and identifying the particularities of its measurement. Moreover, the expected future steps on this research topic are pointed out, offering an insight to the pool of knowledge in this area.

This paper is structured in six sections. Section 2 discusses the general characteristics of SQ in the PT sector and methodological issues associated with its analysis. Section 3 provides a critical assessment of the various methodological approaches based on CSS that have been used to analyse SQ in the PT sector. Section 4 displays the approaches used for estimating the relative importance of the attributes characterizing the service. Section 5 follows with a discussion about the future directions for measuring and analysing service quality. And, finally, Section 6 concludes with a summary and main conclusions.

2. GENERAL CHARACTERISTICS OF SERVICE QUALITY IN PUBLIC TRANSPORT AND METHODOLOGICAL ISSUES

Past research has identified a number of characteristics and methodological issues that are critical considerations in the development and application of an appropriate methodology to analyse SQ in PT. This section presents a summary of these characteristics and methodological issues.

2.1. Complexity of the quality concept

The concept of SQ is complex, fuzzy and abstract, mainly because of the three properties of service: intangibility, heterogeneity and inseparability (Carman, 1990; Parasuraman et al., 1985).

Many authors (e.g. Grönroos, 1988; Parasuraman et al., 1985) maintain that the perception of SQ is the result of a comparison of consumer expectations with actual service performance perception. Other authors, however, do not take expectations into consideration (Cronin and Taylor, 1992). They are only interested in passengers' perceptions, or even the perception of transport companies and government managers (e.g. Eboli and Mazzulla, 2011; Nathanail, 2008).

There is no consensus on customer expectations. Certain models in the literature compare customer performance perception with ideal performance or quality (e.g. Mattsson, 1992); with desired quality (e.g. Gilbert and Wong, 2003); and with adequate or tolerable quality (Hu and Jen, 2006). Teas (1993) stated that expectations could be interpreted as predictions of service, as an ideal standard or as attribute importance. When analysing SQ in the PT sector, many researchers (e.g. Chen and Chang, 2005; Eboli and Mazzulla, 2010) have substituted importance measures for expectations, although there is no theoretical basis for this (Landrum and Prybutok, 2004). However, measuring which service attributes are important to customers may be more meaningful to managers than measuring customer expectations (Smith, 1995).

The relationship between SQ and satisfaction is not clear. In the literature, SQ usually accompanies satisfaction. This may be due to the similar nature of the two variables, which both derive from the disconfirmation theory (Parasuraman et al., 1988). Some authors think that customer satisfaction causes perceived quality and others consider that SQ is a vehicle for satisfaction (e.g. Chen, 2008; Chou and Kim, 2009). In recent years, a lot has been said about the "Service Quality–Satisfaction–Loyalty/Behavioural Intentions" paradigm (Jen et al., 2011). This paradigm suggests that satisfaction is the link between SQ and loyalty or behavioural intentions. Therefore, it would be on a "higher" attitude level with regards to SQ (Mattsson, 1992). Oliver (2010) defines SQ as a cognitive judgement (thinking/judging) that summarizes the exceptionally good (or bad) elements of the service, especially when compared with other direct alternatives; while in contrast, customer's satisfaction is an affective judgement (liking/pleasure) purely experiential, defined as the "Consumer's fulfillment response". However, both concepts are used interchangeably in much of the literature (Cavana et al., 2007; Oliver, 2010), although they are actually different.

Grönroos (1984,1988) and Lehtinen and Lehtinen (1991) support the three-dimensionality of SQ in terms of technical quality (the quality of what consumer actually receives), functional quality (how he gets the technical outcome) and image. Parasuraman et al. (1985, 1988) pointed out that service offers very few tangible elements, and therefore they focused on intangible elements (functional quality).

2.2. Service Quality Attributes

A very large number of attributes have been used to evaluate SQ (e.g. Murray et al. (2010) consider 166 attributes), so they are normally grouped into a smaller number, called dimensions. Although there is no general agreement as to the nature of SQ dimensions, there is a general recognition that service quality is a multidimensional construct (Parasuraman et al., 1985), and multilevel or hierarchical (Jen et al., 2011).

Parasuraman et al. (1988) defended the existence of a generic list of attributes and dimensions to analyse the SQ of any type of service. However, many authors criticized this generic list. Most authors agree that the attributes included in a survey must be selected to each specific case (Babakus and Boller, 1992;

Carman, 1990). In fact, Carrillat et al. (2007) demonstrated that the predictive value of the Parasuraman's model increased when the model's items were adapted to the study context. However, many items are repeated irrespective the type of service and context considered, due to the general importance for any of them. For example, for transit services those items are the frequency of the service, the punctuality, the comfort and cleanliness, the safety, the availability of information, the personnel courtesy, the fare and others. Therefore, additionally to these items, others aspects should be considered for each context-specific service, because the aspects appreciated by each user are highly dependent on the users' social and demographic characteristics; their context (i.e. geographical area, social class and type of service); the reason for travel; and the modes of transport used (e.g. Andreassen, 1995; Ganesan-Lim et al., 2008).

The selection is frequently made on the basis of an exhaustive study of which attributes are the most important in terms of evaluating SQ in the service under study. In the field of PT several methods are used to that end: literature review, survey of operators, focus groups, pilot users survey, statistical tests to identify whether an attribute should or should not be considered. In most cases, combinations of these methods are used (e.g. Chau and Kao, 2009; Dell'Olio et al., 2010; Liu and Gao, 2007). These methods are also used to simplify data collection by lowering the number of attributes. Therefore, ad-hoc surveys are the most appropriate tool for SQ analysis (it is not the same metropolitan services by bus than by metro, or regional services by train than by bus, etc.) and, although specific attributes are considered at each specific context, there are lots of attributes that are repeated among services. Then, comparisons should be made among services with similar characteristics and role.

Various papers point to the existence of several categories of attributes that have a greater or lesser impact on SQ and satisfaction. Philip and Hazlett (1997) propose a model with a hierarchical structure, based on three classes of attributes: pivotal, core and peripheral attributes. This model was subsequently contrasted for the rail transportation industry by Tripp and Drea (2002) who checked that the core attributes (e.g. the service announcements, seat comfort, rest room and café car) exerted the greatest influence on the passengers' satisfaction levels. The UNE-EN 13186 (2003) standard classifies the service characteristics into basic (e.g. punctuality, safety), proportional (e.g. comfort, cleanliness) and attractive (e.g. contactless cards, navigators), depending on how compliance and non-compliance affects customer satisfaction. The Transit Capacity and Quality of Service Manual (TRB, 2004) groups attributes into availability factors, which are more important to passengers (e.g. timetables, service coverage, information), and comfort and convenience factors, less important for passengers (e.g. service appearance, overcrowded, fare). Eboli and Mazzulla (2008) empirically demonstrated the existence of two categories of attributes (basic and not basic) from the preferences showed by users. Basic attributes compromise SQ when their level is low (e.g. punctuality, frequency, service coverage) and non-basic attributes (e.g. cleanliness, driver courtesy) are considered secondary service characteristics that affect SQ if they are present, but do not compromise it if they are absent.

2.3. Nature of the data

In passenger transport services, functional quality is more important than technical quality (Parasuraman et al., 1985; 1988) which gives the SQ concept a subjective nature, in so far as it is the result of passenger perceptions or its comparison with their expectations. Therefore, the evaluation process usually involves subjective assessments, resulting in qualitative and imprecise data being used. Several authors (e.g. Fernandes and Pacheco, 2010; Kuo and Liang, 2012; Kuo, 2011; Yeh et al., 2000) have used the fuzzy set theory as an effective method for handling the issue of subjective, qualitative and imprecise information inherent in the data used to assess SQ.

On the other hand, the subjective nature of this concept produce that the perceptions about different characteristics of the service are very different among users. Users' perceptions are heterogeneous because the qualitative nature of some PT service aspects, the different users' socioeconomic characteristics, and the diversity in tastes and attitudes towards PT.

To analyse this heterogeneity, one possibility is to stratify the sample and then build specific models. Segmentation is normally carried out in terms of the population's socioeconomic and demographic characteristics (i.e. income, gender, car availability, frequency, etc.) (e.g. Andreassen, 1995; Dell'Olio et al., 2010). However, other procedures are also used, such as cluster analysis (e.g. Wen et al., 2008).

In recent years, there has been an emerging debate on whether subjective data (customers' opinions) can be combined with objective data (technical data) on service performance to evaluate the global quality of PT. Some authors (e.g. Parasuraman et al., 1988; TRB, 2004) dismiss this approach because they consider that SQ is the quality perceived by the passengers' point of view. However, this type of measure suffers from a strong subjectivity and does not take into account non-users' perceptions. Moreover, if respondents are not correct sampled or users' judgements are too heterogeneous, considerable statistical errors could occur when analysing SQ (Eboli and Mazzulla, 2012a).

Thereby, in the past ten years several studies have begun to propose the combined use of subjective and objective measures (e.g. Eboli and Mazzulla, 2011; 2012a; Nathanail, 2008). In fact, Eboli and Mazzulla (2011) empirically identified some differences in the data obtained by both sources for defining the quality of various service attributes. Subjective rates showed high standards deviation among users' rates, while this variability was lower for the objective data. Most service attributes obtained a higher objective value than the average satisfaction rate expressed by the users, and only few subjective indicators reached a higher value than the objective one. This is very interesting because if users perceive a service aspect as satisfactory but it does not objectively reach appropriate standards of quality, transit agency may not invest further resources for improving that aspect. On the other hand, if users are unsatisfied with a particular service aspect but this aspect already offers good standards of quality, additional resources allocated for meeting customer requirements would be wasted (Eboli and Mazzulla, 2011). On the contrary, as the main cause could be due to the aspect is somewhat hidden from the travellers, the additional resources should be reallocated to communication and information activities, in order to achieve that customers perceive the high quality of it.

Therefore, objective indicators can provide clearer and less biased information, while passengers' perceptions represent the fundamental point of view for SQ evaluation. Therefore, both joined measures could provide a more useful and reliable measurement tool of transit SQ.

2.4. Surveys

User surveys are an essential tool for collecting the information used to analyse quality. As indicated in Section 1, CSS are widely adopted. These are questionnaires where customers are asked to rate satisfaction or performance perception on each key service attribute. In addition, customers are normally asked to answer other questions as well, depending on the methodological approach used for the subsequent data analysis (Section 3). They are often asked to rate also the importance of each attribute, or a ranking of them, and the global overall service satisfaction (e.g. Friman and Gärling, 2001; Joewono and Kubota, 2007a; 2007c; Koushki et al., 2003). In some cases, they are asked to rate on each attribute, in terms of both perceptions and expectations (e.g. Lin et al., 2008; Sultan and Simpson, 2000); or to rate global service, in terms of both perceptions and expectations (Eboli and Mazzulla, 2012b).

(Table 1)

In Table 1 it is possible to observe large part of the existing research for analysing SQ in PT based on CSS. The air transport services, as well as the urban and metropolitan PT services, are the ones with more research about SQ using CSS (Table1). Normally, ratings are expressed on two scales: numeric or linguistic, answering to questions in the following form: "How satisfied are you with…?". Numeric scales are more widely used and have a wider range: from 3- to 11-points. Table 1 shows that the 5-point Likert scales are the most widely adopted, while linguistic scales are used less and have a narrower range: from 3- to 7-points, defined from "not satisfied" to "very satisfied". High satisfaction ratings mean that a transport company is meeting or exceeding its passengers' expectations, and consequently, if they improve these satisfaction ratings, they could earn the trust and respect of its passengers. On the other hand, low satisfaction ratings mean an under-fulfilment of passengers' service requirements, and the need of urgent improvements. Attributes importance collected by importance rating, or ranking, gives information about passengers' underlying priorities, which provides support and justification for transport strategic decisions.

3. MODELING METHODS FOR ANALYZING SERVICE QUALITY

There are two main theoretical currents for analysing service quality: (a) performance perception and expectations approach (Parasuraman et al., 1985); and (b) only performance perception approach (Cronin and Taylor, 1992). Moreover, there are also two types of methodological approaches, depending on whether SQ is measured by disaggregation (i.e. service attributes are analysed individually) or aggregation (when an aggregate analysis of attributes is used to obtain an overall Service Quality Index, SQI, or a Customer Satisfaction Index, CSI). In this section, a review and critical discussion about the methodologies used for analysing SQ in the PT sector is presented. While disaggregated models help to set priorities for service improvements from among a long list of service attributes, aggregated models provide a SQI that permit service to be analysed over time and different services to be compared (e.g. territorial scope, suppliers, etc.). In some cases both approaches are used together to profit from their benefits: disaggregated models help managers to more optimally focus their organization's attention and resources, and aggregated models allow to globally analyse the level of SQ, and to determine the effect that service interventions have produced on the passengers' overall perception. In order to obtain the SQI, it is essential to know the weight or importance of each attribute in terms of global quality. The manners in which the weights can be obtained are approached in Section 4.

(Table 2)

3.1.- Aggregate performance-expectation models

Parasuraman et al. (1985) proposed that SQ is a function of the differences between expectation and performance, from a customer point of view. They developed a model based on gap analysis and the overall SQ was defined as:

$$SQ = \sum_{j=1}^{k}\left(P_{ij}-E_{ij}\right) \tag{1}$$

where k is the number of attributes; P_{ij} is performance perception of stimulus i with respect to attribute j; and E_{ij} is service quality expectation for attribute j that is the relevant norm for stimulus i. Parasuraman et al. (1988) developed the SERVQUAL (SERVice QUALity) scale for measuring customers' perception of SQ. A number of authors have used the SERVQUAL scale for analysing airline SQ (e.g. Abdlla et al., 2007; Kiatcharoenpol and Laosirihongthong, 2006) and Liu and Gao (2007) adapted the SERVQUAL scale for evaluating railway services.

Although this model has served as a starting point for SQ analysis for several decades because it provides an overall SQI that allows for analysis over the time or to compare different services, this approach does not help to set priorities for service attributes improvements. Also, Eq.1 implies that all the attributes are equally important in SQ, which is not the case.

Other authors proposed weighting each attribute by a weight that would take the importance of each attribute into consideration. Pakdil and Aydin (2007) used a weighted SERVQUAL for analysing airline SQ. Chou et al. (2011b) included fuzziness in SQ evaluation by using a fuzzy weighted SERVQUAL to evaluate airlines.

Another measure for SQ evaluation is provided by the Customer Satisfaction Index (CSI) (Hill et al., 2003). CSI represents a measure of SQ on the basis of attributes' importance and satisfaction rates (see Eq.2).

$$\mathrm{CSI} = \sum_{k=1}^{N} [\bar{S}_k . W_k] \qquad (2)$$

where S_k is the mean of the satisfaction rates expressed by users on the service quality k attribute; and W_k is a weight of the k attribute, calculated on the basis of the importance rates expressed by users. Specifically, it is the ratio between the mean of the importance rates expressed by users on the k attribute and the sum of the average importance rates of all the service quality attributes.

CSI represents a good measure of overall satisfaction because it summarizes customer judgments on several service attributes in a single score. However, customer satisfaction rates can be very heterogeneous among users. These heterogeneities cannot be taken into account by CSI. To overcome this lack, importance weights and satisfaction rates can be corrected according to their dispersion. Eboli and Mazzulla (2009) introduced these adjustments calculating a Heterogeneous Customer Satisfaction Index (HCSI) that they used to evaluate suburban bus lines. HCSI was calculated by Eq.3.

$$\mathrm{HCSI} = \sum_{k=1}^{N}[S_k^c . W_k^c] \text{ where } S_k^c = \bar{S}_k . \frac{\frac{\bar{S}_k}{\mathrm{var}(S_k)}}{\sum_{k=1}^{N}\frac{\bar{S}_k}{\mathrm{var}(S_k)}} \cdot N \text{ and } W_k^c = \frac{\frac{\bar{I}_k}{\mathrm{var}(\bar{I}_k)}}{\sum_{k=1}^{N}\frac{\bar{I}_k}{\mathrm{var}(\bar{I}_k)}} \quad (3)$$

where S^c_k is the mean of the satisfaction rates expressed by users on the k attribute corrected according to the deviation of the rates from the average value; and W^c_k is the weight of the k attribute calculated on the basis of the importance rates expressed by users, corrected according to the dispersion of the rates. HCSI introduces heterogeneity into user judgments: more significance is given to the attributes with homogeneous user judgments.

This group of methods can also include multicriteria analysis (MA) when customers are asked for their degree of satisfaction with a specific criterion or attribute. MA has been widely used to deal with problems involving multiple criteria/attributes, as in the case of PT quality of service. Frequently, MA has been combined with a fuzzy approach: Kuo et al. (2007) assessed SQ for interurban bus services, and several authors have used this approach for airlines. In the model proposed by Chang and Yeh (2002) subjectivity is considered in assessments in terms of attribute satisfaction and attribute importance. Liou and Tzeng (2007) take into account that attributes are not usually independent. Tsaur et al. (2002) and Nejati et al. (2009) ranked airlines' SQ factors using a fuzzy TOPSIS approach. The VIKOR method, which is based on an aggregate function representing "closeness to the ideal point" has also been adopted recently (e.g. Kuo and Liang, 2011; Liou et al., 2011b). Opricovic and Tzeng (2004) compared VIKOR and TOPSIS and demonstrated that TOPSIS does not consider the relative importance of attributes.

3.2.- Aggregate models based only on performance

Some researchers consider that models based only on performance perception are better than combined perception-expectation models (Babakus and Boller, 1992; Cronin and Taylor, 1992). Arguments in favour of only perceptions models are based on the notion that performance perceptions are already the result of customers' comparison of the expected and actual service. Cronin and Taylor (1992) proposed a measurement based only on performance perceptions (SERVPERF). Overall SQ is evaluated according to:

$$SQ = \sum_{j=1}^{k} P_{ij} \quad (4)$$

where k is the number of attributes; and P_{ij} is performance perception of stimulus i with respect to attribute j.

Eq.4 implies that all the attributes have the same weight in SQ. So, many authors propose weighting each attribute to take into consideration their respective importance in SQ. In the field of PT, Sánchez et al. (2007) proposed a weighted SERVPERF for assessment of urban bus services.

MA based on SERVPERF was also used for analysing SQ in the PT sector. Yeh et al. (2000) and Awasthi et al. (2011) used a fuzzy MA approach for evaluating urban transportation systems. This approach was also used for airlines (Kuo, 2011) and airports evaluation (e.g. Yeh and Kuo, 2003). Nathanail (2008) evaluated railways using a MA based on objective and subjective data from several sources: statistical data, mystery rider and CSS. In their opinion, it is impossible for a passenger to be able to provide a global performance grade of the itinerary based on a short experience, and therefore a combination of objective and subjective measurements is proposed (TRB, 1999).

3.3.- Disaggregated models based on performance only

The disaggregated models most widely used to evaluate SQ are based on Quadrant Analysis (QA) (Figure 1). The most widespread QA is Importance-Performance Analysis (IPA), which uses importance and performance as coordinates. This quadrant chart quantifies how important each attribute appears to be from a customer perspective (y-axis) and shows the average customer rating for each characteristic (x-axis).

This simple technique prescribes the prioritization of attributes for improvement, and provides guidance for strategy formulation (Slack, 1994). However, the IPA matrix is a visualizing method and how to determine the precise ranking of the priority of improving attributes remains ambiguous and unidentified (Abalo et al., 2007).

(Figure 1)

IPA has been applied to evaluate transit systems (Weinstein, 2000), high speed railways (Chou et al., 2011a) and airlines (Chen and Chang, 2005). This method is widely used by transport company managers in the metropolitan transport sector (Christopher et al., 1999; Figler et al., 2011; Foote and Stuart, 1998) owing to their simplicity.

Stradling et al. (2007) introduced the user disgruntlement measure, derived by cross-tabulating performance against importance rating for each attribute. They used this variation of the IPA to analyse different aspects of a particular service (e.g. user satisfaction with bus interchange), to compare across modes (e.g., user satisfaction with trips by car and bus), and within a mode across population sub-groups.

Eboli and Mazzulla (2011), following Nathanial (2008) and Tyrinopoulos and Aifadopoulou (2008), recently used a non-weighted disaggregated method, based on the use of both passenger perception and transit agency performance measures, to evaluate a suburban bus line. The method is based on each attribute having a subjective indicator (S) (calculated by the average of satisfaction rates expressed by a sample of users about the attribute) and an objective indicator (O) (obtained from performance indicators or, for the most qualitative attributes, calculated as the average of the scores assigned by operators or mystery riders to the parameters). Subsequently, through an optimization process, using the variance of S and O, a composite indicator (X) was obtained for each attribute. If the variance of the objective indicator is very low (close to 0) the X value coincides with the O indicator, by ignoring S indicator, and vice versa.

3.4.- Disaggregated performance-expectation models

Parasuraman et al. (1991) proposed the concept of the zone of tolerance (ZOT) of expectations. They thought expectation could be divided into two levels: desired service (DS) and adequate service (AS). ZOT is the difference between DS and AS, service superiority (SS) is the difference between DS and perceived service (PS), and service adecuacy (SA) is the difference between PS and AS (Zeithaml et al., 1993). Following Parasuraman et al. (1991), DS is the service the customer hopes to receive (it is a blend of what the customer believes “can be” and “should be”) and AS level is that which the customer finds acceptable.

Hu and Jen (2006) define SQ in terms of the difference between perceived quality and tolerable quality (AS following Parasuraman et al. (1991)) and apply it to evaluate the SQ of urban buses. Cavana et al (2007) use ZOT for managing passenger rail service quality. Recently, Chang et al. (2012) introduced the fuzzy ZOT concept and applied it in the airline cargo business.

IPA was also used interchanging performance by satisfaction (Mathisen and Solvoll, 2010; Wang et al., 2010). Recently, Tsai et al. (2011) combined AHP, VIKOR and IPA methods for considering airport passengers' preferences (importance) and satisfaction simultaneously. The AHP was employed to measure the relative importance of each attribute; then, the VIKOR method was used for computing the customer gaps of airport passenger service. Finally, IPA was used for improving (reducing the gaps) attributes with higher importance.

Based on ZOT and IPA, Hu (2010) proposed the concept of ZOT of expectation for evaluating SQ (ZSQ) and built an analytical framework for prioritizing attributes through a QA based on ZSQ and normalized importance (NIZSQ method). ZSQ is based on the concept of the 'performance ratio' in the customer satisfaction area (Vavra, 1997). The 'performance ratio' quantifies how much, from minimal to superior performance, an organization has progressed on a specific attribute. According to the same concept, ZSQ can show the 'SQ ratio'. Since DS, AS and PS can be seen as 'superior', 'minimum' and 'current', ZSQ can be expressed by the following equation:

$$\mathrm{ZSQ} = \frac{\text{PS-AS}}{\text{DS-AS}} = \frac{\mathrm{SA}}{\mathrm{ZOT}} \tag{5}$$

The meaning of SA divided by ZOT represents the performance ratio of SQ according to the customers' expectation. The smaller value of the service attribute's ZSQ means worse performance and should therefore have a higher priority to be improved.

After evaluating the ZSQ, managers need to consider the attribute's importance for judging the priority for improving attributes whose ZSQ values are between '0' and '1'. They only need to focus on values between '0' and '1' for two reasons (Hu, 2010):

- If ZSQ >1, PS is higher than DS and there is no need for improvement at the moment.
- If ZSQ < 0, the attribute must be improved immediately without any prioritizing analysis.

NIZSQ can be used for this purpose (Hu, 2010). NIZSQ method normalizes the importance data (NI) and replaces the x-axis in the IPA by ZSQ. Thus, NIZSQ analysis can be drawn as a two-dimensional diagram whose x-axis and y-axis have the same range (Figure 1): from '0' to '1' and they can be divided into four quadrants. The meanings of the four quadrants are the same as for the traditional IPA.

The top-right and bottom-left diagonal shows the ideal positions for attributes, which means that the performance of SQ is even with the importance (Slack, 1994). So, attributes on the left side of the diagonal need to be improved. The horizontal distance between attributes and the diagonal represents the improving space and the degree of urgency. The longer the distance is, the larger the space to improve is, and therefore the higher the priority to be improved. If the attributes have the same *d* value, they should be prioritized by their importance (Hu, 2010).

Hu (2010) used NIZSQ analysis to evaluate SQ of bus services and compared the results with a traditional IPA. While IPA may lead managers to focus only on some items and ignore others, NIZSQ analysis reminds managers that they should keep those items in mind. Furthermore, NIZSQ analysis is not only a QA, but also offers the improvement priority (d value) of each item based on the ZOT. Since prioritization is critical to managers' planning and they are usually unlikely to be able to focus on all items, *d* value can give them clear information regarding which items should be improved in priority and which items later.

3.5.- Other analyses

Finally, there are other studies in the literature that do not come under any of the methodological approaches indicated in Table 2.

Some studies are conducted in terms of verifying hypotheses on SQ based on data supplied by CSS. Some of them use standard statistical methods (e.g. t-test, ANOVA, MANOVA, etc.) to confirm the hypotheses (e.g. Drea and Hanna, 2000; Pedersen et al., 2011). Others, however, use more advanced methods, such as

Structural Equations Models (SEM) or Path Analysis for verification purposes (SEM and Path Analysis are described in Section 4).

Other papers study the differences in SQ perceived by different groups of individuals, services or companies, or before/after carrying out an action. Many authors (e.g. Kim et al., 2011; Paquette et al., 2012; Surovitskikh and Lubbe, 2008) compare SQ through different categories of users or population groups, using the standard statistical methods pointed out in the previous paragraph. This type of analysis has been also conducted using more advanced methods, including ordered choice models (Dell'Olio et al., 2010; Hensher et al., 2010; Huse and Evangelho, 2007), SEM (Andreassen, 1995; Friman et al., 2001) or Path Analysis (Ringle et al., 2011).

Most authors compare SQ in different services and companies to each other, using standard statistical methods (e.g. López-Bonilla and López-Bonilla, 2008; Ostrowski et al., 1993; Park, 2007), although comparisons have also been made using SEM (Chou and Kim, 2009; Chou et al., 2011b), Path Analysis (Forgas et al., 2010) and ordered choice models (Tyrinopoulos and Antoniou, 2008).

Finally, some studies analysed SQ before and after carrying out an action (Foote et al., 2001; Friman, 2004; Pedersen et al., 2011).

4. APPROACHES TO ESTIMATE THE RELATIVE IMPORTANCE OF EACH SERVICE QUALITY ATTRIBUTE

PT companies want to know not only how their customers rate them on detailed service attributes (attribute-performance ratings), but also the relative importance of these attributes (attribute-importance measures) to their customers. Section 3 shows that most approaches use the importance of each attribute for analysing SQ, due to not all the attributes equally affect customers' overall evaluation, but some of them are key drivers for generating their opinions. Considering the attributes-performance ratings joined to the attribute-importance measures provide a deeper understanding about how passengers' global evaluation is created.

At this section, there are displayed the different approaches used for estimating the attributes' importance in the PT sector, highlighting their main advantages and disadvantages. Table 3 collects several research carried out in the transport industry, classified by the method used to determine the weight of the attributes. The most common approach (Table 3) is asking customers to rate each attribute on an importance scale (Stated Importance), although methods that derive attribute importance by statistically testing the strength of the relationship of individual attributes with overall satisfaction (Derived Importance) are also widely used.

(Table 3)

4.1. Stated importance

This is the most intuitive and simplest of both methods: passengers are asked for the importance that attributes have for them. However, this approach has several disadvantages:

- it increases the length of the survey. This can depress the overall response rate and accuracy of the survey
- it yields insufficient differentiation among mean importance ratings, with customers rating nearly all of the measures near the top of the scale
- attributes may be rated as important even though they in fact have little influence on satisfaction.

In some cases (Liou and Tzeng, 2007; Tsai et al., 2011; Tsaur et al., 2002) more sophisticated processes are used, such as the Analytic Hierarchy Process (AHP).

4.2. Derived importance

It is common practice to include in CSS both questions about a customer's overall satisfaction with the service and detailed questions about specific characteristics of the service. The information gathered can be used in several statistical methods (e.g. bivariate correlations, multiple-regression analysis, SEM, etc.) for deriving the attributes importance from CSS.

Factor analysis (FA)

FA is a set of multivariate statistical techniques whose primary goal is to investigate whether a number of variables of interest are linearly related to a smaller number of unobservable factors. FA provides a better understanding of how customers perceive various service attributes by showing which attributes tend to be thought of similarly. FA is normally used as a preliminary step for other methods, such as multiple linear regression analysis (Kim and Lee, 2011; Weinstein, 2000), discriminant analysis (Aksoy et al., 2003) or SEM (Eboli and Mazzulla, 2012b). The factors provide a more manageable number of variables with which to carry the analysis to the next level.

Bivariate correlations

Bivariate correlations can be used as a tool for ranking the relative importance of each attribute (Figler et al., 2011; Weinstein, 2000). The main disadvantage of this method is that it disregards the correlation among attributes, so it is important not to interpret the coefficients too literally owing to the extensive collinearity among them.

Regression analysis

The purpose of regression analysis is to assess the relative importance of each factor and to test the overall explanatory power of the battery of factors as a whole. In the regression model, the factors serve as the independent variables (IV), whereas overall satisfaction, or SQ, serves as the dependent variable (DV). Regression analysis results in a best-fitting model in the form of an equation that expresses the DV as a combination of the IV. Several models of regression have been proposed to study satisfaction or SQ (Table 3).

Papers based on multiple linear regression models (Kim and Lee, 2011; Weinstein, 2000) do not take the categorical nature of the DV into consideration and are infrequently used in the literature.

The most widely used methods are the ones that take into account that the DV is categorical. Aksoy et al. (2003) propose using Discriminant Analysis (DA) to identify key service dimensions for predicting satisfaction in airlines. DA undertakes the same task as multiple linear regressions by predicting an outcome, but considering that the DV is categorical. Logistic regression and probit regression are similar to DA, as they also explain a categorical variable. However, these other methods are preferable in applications where it is not reasonable to assume that the independent variables are normally distributed, which is a fundamental assumption of the DA method. Several authors propose Ordered Logit, OL (e.g. Tyrinopoulos and Antoniou, 2008) and Ordered Probit, OP (e.g. Dell'Olio et al., 2010) models to study the relationship between overall satisfaction and each of the attributes under consideration. OL or OP models are extensions of the logistic or probit regression models, allowing for more than two (ordered) response categories, which is the situation encountered with the CSS. In OP models the unobserved terms are supposed to be distributed as standard normal instead of logistic, which is the hypothesis in OL. Recently, Hensher et al. (2010) proposed a Generalized Ordered Logit model that accounts for preference heterogeneity through random parameters.

Structural Equation Models (SEM)

SEM is a multivariate technique combining regression, factor analysis, and analysis of variance to estimate interrelated dependence relationships simultaneously. This approach allows the modelling of a phenomenon by considering both the unobserved "latent" constructs and the observed indicators that describe the phenomenon. SEMs are made up of two elements: the first describes the relationship between endogenous and exogenous latent variables, and permits the evaluation of both direction and strength of the causal effects among these variables (latent variable model); the second component describes the relationship between latent and observed variables (measurement model). The structural equation system is estimated by using different methods: maximum likelihood, weighted and un-weighted least squares, generalized least squares, and so on.

SEM have been adopted for describing customer satisfaction in several PT services: metropolitan public transportation (e.g. Karlaftis et al., 2001; Lai and Chen, 2011; Minser and Webb, 2010; Nurul-Habib et al., 2011; Stuart et al., 2000; de Oña et al. 2013); interurban bus services (e.g. Wen et al., 2005); rail transportation (e.g. Chou et al., 2011a; Tripp and Drea, 2002; Eboli and Mazzulla, 2012b) and airlines (e.g. Cheng et al., 2008; Park et al., 2006; Saha and Theingi, 2009; Yang et al., 2012).

Path analysis can be viewed as a special case of SEM: one in which only single indicators is employed for each of the variables in the causal model. That is, path analysis is SEM with a structural model, but no measurement model. Several authors (e.g. Forgas et al., 2010; Park et al., 2004) used this method for modelling airlines SQ. Jen and Hu (2003) and Lin et al. (2008) used path analysis for evaluating bus services. Finally, Joewono and Kubota (2007b) used it for analysing user perceptions of paratransit.

5. FUTURE DIRECTIONS

The first studies of SQ in the PT sector emerged in air transport sector and in urban and metropolitan PT in the late 20th century. However, such studies have increased considerably at the start of the 21st century, particularly in the field of airlines (e.g. American Airlines, 2010; Bowen and Headley, 2000) and urban and metropolitan transport (e.g. AVV Transport Research Centre, 2004; Department for Transport of England, 2005; Transport for London, 2006). These two sectors pose different problems but share the same goal: to increase the number of passengers. In the case of air transport, the deregulation and opening-up-the-sky policies of the airline industry have put pressure on airlines and airports to become more competitive. In the urban and metropolitan PT, companies and governments are highly interested in enhancing the quality in order to discourage the use of cars. There have been fewer documented studies on ground interurban PT (by bus or by train) (e.g. Tripp and Drea, 2002; Wen et al., 2005). It is to be hoped that SQ concerns in these sectors will grow if this kind of services are liberalized, as it is one of the main objectives of the European transport policy (CE, 2001).

In this context it is essential that transport managers can have a tool for measuring the quality of the service delivered, in order to formulate profitable funding strategies that improve the levels of performance of the service in harmony with the passengers' requirements.

It is not possible to identify the best method for measuring service quality. Each one of the methods shows advantages and disadvantages. Table 4 shows a summary of the key advantages and disadvantages of each one of the models.

(Table 4)

The approaches most widely used by practitioners, transport operators and governments are the ones based on CSS that use a quadrant analysis, such as the IPA and its variants (e.g. Christopher et al., 1999; Figler et al., 2011; Foote and Stuart, 1998). Such methods help managers to set priorities for service improvements among a list of service attributes. Although IPA is the most used, it is only a visualizing method, it does not provide a precise ranking of the priority of improving attributes, and it sometimes can lead to ignore attributes that should have a high priority to be improved. For example, if an attribute is not considered as important (e.g. cleanliness) and its perceived quality is not very low (although it is under an acceptable threshold for passengers), IPA will ignore it, while NIZSQ will focus on it immediately. Hu (2010) compared the results using IPA and NIZSQ, and found significant differences (e.g. when the attribute "The company deals with passengers opinions and complaints" was ranked in the third place of priority in the NIZSQ method, it was not consider of high priority using IPA). Then, NIZSQ could better provide this information. However, this technique is complex and uses two types of rates (expectation and importance of each attribute). This could lead to passengers' confusion, and probably will not be very used by practitioners in the future.

On the contrary, researchers and academics have sought to arrive at a global indicator (SQI or CSI) that could be used to compare different services and their development over time. The most widely used models in this case have been based on disconfirmation theory (Parasuraman et al., 1988). The traditional SERVQUAL model has been improved adding the importance of the attributes in the estimation of the index (weighted SERVQUAL and fuzzy weighted SERVQUAL) and also the CSI has been improved in the HCSI (Eboli and Mazzulla, 2009), which introduces the heterogeneity of the users' judgements. However, both approaches (aggregate and disaggregate) should be used together to complement each other.

Given the subjective, qualitative and imprecise nature inherent to SQ evaluation data, a growing number of studies are using fuzzy set theory as an effective way for formulating this kind of problems. Another emerging trend in recent years proposes the combined use of subjective information obtained from users (through CSS) and objective data on service performance supplied by transport companies (e.g. Eboli and Mazzulla, 2011; Nathanail, 2008; Yeh et al., 2000). This new approach is based on the consideration that passengers' perceptions alone can lead to many biases, especially when users' judgements are too heterogeneous, thereby both joined measures could provide a more useful and reliable measurement tool of transit SQ.

Analysing the heterogeneity of user perceptions has focused attention from the first studies on SQ in the PT sector (e.g. Glascock, 1997; Ritchie et al., 1980) in which the difference in the SQ perceived by different groups of individuals that had been previously segmented (in terms of socioeconomic or demographic data, travel habits, etc.) were compared. More recently, several works (Huse and Evangelho, 2007; Wen and Lai, 2010; Wen et al., 2008) proposed new approaches that use several methods to identify clusters. This new method can be used to study specific population segments whose behaviour could hardly be identified by conventional stratification based on socioeconomic and/or demographic factors. If transport managers and practitioners are able to handle this heterogeneity they could formulate more successful strategies. If they want to success they need to adapt their strategies to each one of the different groups of users (i.e. personalized marketing). HCSI (Eboli and Mazzulla, 2009) is also interesting with regards to taking into account the heterogeneity of user perceptions. It can be used to give more significance to the SQ attributes characterized by homogeneous user judgement, while less significance is given to the more heterogeneous attributes.

Moreover, as indicated in the introduction to this paper, discrete choice models based on SP surveys are becoming more widespread in the analyses of SQ in the PT sector. Several studies (e.g. Eboli and Mazzulla, 2008; Hensher et al., 2003; Gatta and Marcucci, 2007) assume that the overall level of passenger satisfaction is best measured by how an individual evaluates the total package of services offered. Appropriate weights attached to each service dimension will reveal the strength of positive and negative sources of overall satisfaction. The weights are estimated using several models based on SP surveys, such as: multinomial logit, hierarchical or nested logit and mixed logit models. Gatta and Marcucci (2007) point out that these methods overcome some critical factors pertaining to methods based on CSS, such as conceptual grounds, psychometric problems and troubles with Likert scales. The latter, in particular, have a well-documented tendency for respondents to choose central response options rather than extreme ones.

Other factors include the impact of the number of scale points used; the influence of the format and verbal labelling of the points; and the transformation from ordinal data to cardinal data. However, for the moment, SP surveys and discrete choice models are complex, and probably practitioners will not use them in the near future.

The drawbacks of the methods that determine the relative importance of each SQ attribute based on stated importance are many: increase in the length of the survey; insufficient differentiation among mean importance ratings, with customers rating nearly all of measures near the top; and attributes that are rated as important even though they have little influence. Nonetheless, there has been no increase in the methods used to derive importance by statistically testing the strength of the relationship of individual attributes with overall satisfaction. This may be largely due to the fact that although statistically inferred methods can overcome the shortcomings of stated importance ratings, most of them carry the assumptions of relatively normal data, linear relationships between independent and dependent variables, and the relatively low multi-collinearity between independent variables and, in customer satisfaction research, these assumptions are almost always violated (Garver, 2003). Lately, new methods are being proposed that can overcome these weaknesses, such as the AHP (Liou and Tzeng, 2007; Tsai et al., 2011; Tsaur et al., 2002) and others based on data mining (de Oña et al., 2012, de Oña et al. 2014a; de Oña et al. 2014b).

The methods based on data mining have the advantage of not needing assumptions or pre-defined underlying relationships between dependent and independent variables, and therefore they are now being used to study SQ in the PT sector. Liou et al. (2011a) use the dominance-based rough set approach for identifying a set of ‘‘*If-then*’’ rules for airport SQ improvement. An “*If-then*” rule is a conditional statement which provides a prediction of the target variable when a set of conditions are complied. For example, imagine a transport company that faces the following two rules:

IF (Frequency<6 AND Information<4) THEN (overall SQ=POOR)

IF (Frequency<6 AND Information>4) THEN (overall SQ=FAIR)

In this case, the company can decide the strategy based on its resources limitations. Maybe increasing the quality of Frequency is not affordable, while increasing the quality of Information is easier, achieving that POOR evaluations are removed. These rules allow considering more than one attribute at the same time, which is not the case for IPA. Recently, de Oña et al. (2012; 2014a; 2014b) and de Oña and de Oña (2013a; 2013b) use decision trees to infer the rules and the attributes that have the most influence on metropolitan PT quality of service and on a railway service. Owing to their success when applied to many other fields and the fact that they provide effective ‘‘If-then’’ rules that make the model very practical and easy to interpret from the perspective of management by PT operators and managers, it is to be hoped that such models will be further developed in the years to come.

6. SUMMARY AND CONCLUSIONS

As the preceding discussion indicates, SQ evaluation of PT poses formidable challenges: how to deal with such a complex, fuzzy and abstract concept as SQ; whether we should use performance perception only or also customers' expectations; which expectations should be considered (ideal, desired, adequate or tolerable quality); what is the relationship between SQ and satisfaction; how to identify the most relevant attributes that affect SQ; how to deal with subjective, qualitative and fuzzy data from surveys; possibility of using objective data (from transport companies) combined with subjective data for SQ analysis; customers satisfaction surveys limitations (maximum length of the survey, scale used, etc.); best ways to analyse heterogeneity; etc.

To deal with these challenges, innovative methodological approaches have been introduced from the beginning of 21st century. However, everybody will use not all the models. Researchers and academics probably will develop and use the most sophisticated and complex ones, e.g. SP surveys and discrete choice models, fuzzy logic, NIZSQ, etc. While practitioners and transport managers will use the simplest and most useful for their main objective, i.e. increase the passengers' perceived SQ for increasing profitability. Since SQ measurement is carried out for very practical reasons, we believe that, among the approaches that have appeared in recent years, there are three that probably will have a highest impact on practice: the combined use of objective and subjective information, cluster analysis and "If-then" rules. Since the subjective information is essential for PT managers, and it cannot be replaced by objective data, the combination of both kinds of data sources (objective and subjective information) could acquire notable importance in the future, and the identification of gaps among these different SQ evaluations is worthy of more research, in order to fulfil the need to provide a reliable as possible measurement tool of SQ for PT managers. Cluster analysis is very interesting for achieving a personalized marketing in PT. And finally the rules based on data mining are easy to understand by practitioners and do not need pre-assumptions between the variables relationships.

In summary, this paper tries to provide a useful and comprehensive review of the key issues associated with SQ evaluation in the PT sector as well as the reasons behind the different methodological approaches that have been used to address this issue. This paper tries to add useful insight to the knowledge in this area.

ACKNOWLEDGEMENTS

This paper benefited from the input of Dr. Laura Eboli and Dr. Gabriella Mazzulla. Their comments and suggestions are gratefully acknowledged. This study is sponsored by the Conserjería de Innovación, Ciencia y Economía of the Junta de Andalucía (Spain) through the Excellence Research Project denominated "Q-METROBUS-Quality of service indicator for METROpolitan public BUS transport services".

REFERENCES

Abalo J, Varela J, Manzano V. (2007) Importance values for importance-performance analysis: a formula for spreading out values derived from preference rankings. Journal of Business Research, 60(2), 115–121.

Abdlla G, Mohamed AR, Mekawy MA. (2007) Managing tourists' needs and expectations: An empirical analysis of the egyptian airline sector. Tourism, 55(3), 277-296.

Aksoy S, Atilgan E, Akinci S. (2003) Airline services marketing by domestic and foreign firms: differences from the customers' viewpoint, Journal of Air Transport Management, 9, 343–351.

American Airlines (2010). 2010 Corporate Responsability Report. (Avaliable at: http://www.americanairlines.es/content/images/aboutUs/corporate/AmericanAirlines2010CRR.PDF)

Andreassen TW. (1995) (Dis)satisfaction with public services: The case of public transportation. Journal of Services Marketing 9: 30–41.

AVV Transport Research Centre (2004) Passenger Transport in the Netherlands (Rotterdam: Ministry of Transport, Public Works and Water Management of Netherlands). Available at: http://www.rwsavv.nl/pls/portal30/docs/10197.PDF)

Awasthi A, Chauhan SS, Omrani H, Panahi A. (2011) A hybrid approach based on SERVQUAL and fuzzy TOPSIS for evaluating transportation service quality. Computers and Industrial Engineering, 61(3), 637-646.

Babakus E, Boller GW. (1992) An Empirical Assessment of the SERVQUAL Scale. Journal of Business Research, 24, 253-268.

Bowen B., Headley D. (2000) Air Travel Consumer Report: The Airline Quality Rating 2000. US Department of Transportation, Washington, DC.

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

Carrillat FA, Jaramillo F, Mulki JP. (2007) The validity of the SERVQUAL and SERVPERF scales A meta-analytic view of 17 years of research across five continents. International Journal of Service Industry Management, 18(5), 472-490.

Cavana RY, Corbett LM, Lo YL. (2007) Developing zones of tolerance for managing passenger rail service quality. International Journal of Quality & Reliability Management, 24(1), 7–31.

CE (2001) European transport policy for 2010: time to decide. White Paper, 124p.

Chang H, Liang G, Chu C, Chou C. (2012) Prioritizing service attributes for improvement using fuzzy zone of tolerance. International Journal of Innovative Computing, Information and Control, 8(1A), 75-89.

Chang YH, Yeh CH. (2002) A survey analysis of service quality for domestic airlines. European Journal of Operational Research, 139, 166-177.

Chau VS, Kao YY. (2009) Bridge over troubled water or long and winding road? Gap-5 in airline service quality performance measures. Managing Service Quality, 19(1), 106-134.

Chen C. (2008) Investigating structural relationships between service quality, perceived value, satisfaction, and behavioral intentions for air passengers: Evidence from Taiwan. Transportation Research Part A, 42(4), 709-717.

Chen FY, Chang YH. (2005) Examining airline service quality from a process perspective. Journal of Air Transport Management, 11, 79–87.

Cheng J-H, Chen F-Y, Chang Y-H. (2008) Airline relationship quality: an examination of Taiwanese passengers. Tourism Research, 29, 487-499.

Chou J, Kim C, Kuo Y, Ou N. (2011a) Deploying effective service strategy in the operations stage of high-speed rail. Transportation Research Part E, 47(4), 507-519.

Chou CC, Liu LJ, Huang SF, Yih JM, Han TC. (2011b) An evaluation of airline service quality using the fuzzy weighted SERVQUAL method. Applied Soft Computing, 11, 2117-2128.

Chou JS, Kim C. (2009) A structural equation analysis of the QSL relationship with passenger riding experience on high speed rail: An empirical study of Taiwan and Korea Expert Systems with Applications, 36, 6945–6955.

Christopher MK, Stuart D, Foote PJ. (1999) Structuring and assessing transit management response to customer satisfaction surveys. Transportation Research Record, 1669, 99-108.

Cronin J, Taylor S. (1992) Measuring service quality: a reexamination and extension. Journal of Marketing, 56, 55-68.

Dell'Olio L, Ibeas A, Cecin P. (2010) Modelling user perception of bus transit quality. Transport Policy, 17(6), 388-397.

Department for Transport of England (2005) Bus Quality Indicators: England. Available at: http://www.dft.gov.uk/stellent/groups/dft_transstats/documents/page/dft_transstats_041049.hcsp) (accessed 6 February 2006).

De Oña J and de Oña R. (2013a) Quality in public transport services for promoting a sustainable mobility: Case study. Carreteras 191 (Special Edition 17 th IRF World Meeting), 81-90

De Oña J, De Oña R, Calvo FJ. (2012) A classification tree approach to identify key factors of transit service quality. Expert Systems with Applications, 39, 11164–11171.

De Oña, J, de Oña, R, Eboli, L, Mazzulla, G. (2013) Perceived service quality in bus transit service: A structural equation approach. Transport Policy 29, 219-226.

De Oña, J, de Oña, R, Eboli, L, Mazzulla, G. (2014a) Heterogeneity in perceptions of service quality among groups of railway passengers. International Journal of Sustainable Transportation, in press. (DOI: 10.1080/15568318.2013.849318)

De Oña R and de Oña J. (2013b). Analyzing transit service quality evolution using decision trees and gender segmentation. WIT Transactions on the Built Enviroment 130, 611-621

De Oña, R, Eboli, L, Mazzulla, G. (2014b) Key factors affecting rail service quality in the northern Italy. A Decision Tree Approach. Transport, 29 (1), 75-83.

Drea JT, Hanna JB. (2000) Niche marketing in intrastate passenger rail transportation. Transportation Journal, 39(3), 33-43.

Eboli L, Mazzulla G. (2007) Service quality attributes affecting customer satisfaction for bus transit. Journal of Public Transportation, 10(3), 21-34.

Eboli L, Mazzulla G. (2008) A Stated Preference Experiment for Measuring Service Quality in Public Transport. Transportation Planning and Technology, 31(5), 509-523.

Eboli L, Mazzulla G. (2009) A new customer satisfaction index for evaluating transit service quality. Journal of Public Transportation, 12(3), 21-37.

Eboli L, Mazzulla G. (2010) How to capture the passengers' point of view on a transit service through rating and choice options. Transport Review, 30(4), 435-450.

Eboli L, Mazzulla G. (2011) A methodology for evaluating transit service quality based on subjective and objective measures from the passenger's point of view. Transport Policy, 18, 172–181.

Eboli L, Mazzulla G. (2012a) Performance indicators for an objective measure of public transport service quality. European Transport, 51.

Eboli L, Mazzulla G. (2012b) Structural Equation Modelling for Analysing Passengers' Perceptions about Railway Services. Procedia: Social & Behavioral Sciences, 54, 96-106.

Fernandes E, Pacheco RR. (2010) A quality approach to airport management. Quality and Quantity, 44(3), 551-564.

Figler SA, Sriraj PS, Welch EW, Yavuz N. (2011) Customer loyalty and Chicago, Illinois, transit authority buses. Transportation Research Record, 2216, 148-156.

Foote PJ, Stuart DG. (1998) Customer satisfaction contrasts express versus local bus service in Chicago's North Corridor. Transportation Research Record, 1618, 143–152.

Foote PJ, Stuart DG, Elmore-Yalch R. (2001) Exploring customer loyalty as a transit performance measure. Transportation Research Record, 1753, 93–101.

Forgas S, Moliner MA, Sánchez J, Palau R. (2010) Antecedents of airline passenger loyalty: Low-cost versus traditional airlines. Journal of Air Transport Management, 16(4), 229-233.

Friman M, Gärling T. (2001) Frequency of negative critical incidents and satisfaction with public transport services, II. Journal of Retailing and Consumer Services, 8(2), 105-114.

Friman M. (2004) Implementing quality improvements in public transport. Journal of Public Transportation, 7(4), 49–65.

Friman M, Edvardsson B, Garling T. (2001) Frequency of negative critical incidents and satisfaction with public transport services, I. Journal of Retailing and Consumer Service, 8(2), 95-104.

Ganesan-Lim C, Russell-Bennett R, Dagger T. (2008) The impact of service contact type and demographic characteristics on service quality perceptions. Journal of Services Marketing, 22(7), 550–561.

Garver MS. (2003) Best practices in identifying customer-driven improvement opportunities. Industrial Marketing Management, 32(6), 455–466.

Gatta V, Marcucci E. (2007) Quality and Public Transport Service Contracts. European Transport, 36, 92-106.

Gilbert D, Wong RKC. (2003). Passenger expectations and airline service: a Hong Kong based study. Tourism Management, 24, 519–532.

Glascock J. (1997) Research on customer requirements for transit service design and delivery. Transportation Research Record, 1604, 121–127.

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

Grönroos C. (1988) Service Quality: The six criteria of good service quality, Review of Business. New York: St. John's University Press.

Hensher DA, Mulley C, Yahya N. (2010) Passenger experience with quality-enhanced bus service: The Tyne and Wear 'superoute' services. Transportation, 37(2), 239-256.

Hensher DA, Daniels R. (1995) Productivity measurement in the urban bus sector. Transport Policy, 2(3), 179–194.

Hensher DA, Prioni P. (2002) A Service Quality Index for Area-wide Contract Performance Assessment. Journal of Transport Economics and Policy, 36(1), 9-113.

Hensher DA, Stopher P, Bullock P. (2003) Service quality - developing a service quality index in the provision of commercial bus contracts. Transportation Research Part A, 37(6), 499-517.

Hill N, Brierley G, MacDougall R. (2003) How to Measure Customer Satisfaction. Gower Publishing, Hampshire.

Hu K-C. (2010) Evaluating City Bus Service Based on Zone of Tolerance of Expectation and Normalized Importance. Transport Reviews, 30 (2), 195–217.

Hu K-C, Jen W. (2006) Passengers' Perceived Service Quality of City Buses in Taipei: Scale Development and Measurement. Transport Reviews, 26 (5), 645–662.

Huse C, Evangelho F. (2007) Investigating business traveller heterogeneity: Low-cost vs full-service airline users? Transportation Research Part E, 43, 259–268.

Jen W, Hu K-C. (2003) Application of perceived value model to identify factors affecting passenger's repurchase intentions on city bus: a case study of the Taipei metropolitan area. Transportation, 30, 307–327.

Jen W, Tu R, Lu T. (2011) Managing passenger behavioral intention: an integrated framework for service quality, satisfaction, perceived value, and switching barriers. Transportation, 38, 321–342.

Joewono TB, Kubota H. (2007a) Exploring negative experiences and user loyalty in paratransit. Transportation Research Record, 2034, 134-142.

Joewono TB, Kubota H. (2007b) User perceptions of private paratransit operation in Indonesia. Journal of Public Transportation, 10(4), 99-118.

Joewono TB, Kubota H. (2007c) User satisfaction with paratransit in competition with motorization in Indonesia: Anticipation of future implications. Transportation, 34(3), 337-354.

Karlaftis MG, Golias J, Papadimitriou E. (2001) Transit quality as an integrated traffic management strategy: Measuring perceived service. Journal of Public Transportation, 4(1).

Kiatcharoenpol T, Laosirihongthong T. (2006) Innovations in service strategy: An evaluation of quality in airline service operations by using SERVQUAL Model. In: IEEE international conference on management of innovation and technology, Singapore, China (pp. 748–752).

Kim YK, Kim YB, Lee YI. (2011) Perceived service quality for south korean domestic airlines. Total Quality Management and Business Excellence, 22(10), 1041-1056.

Kim YK, Lee HR. (2011) Customer satisfaction using low cost carriers. Tourism Management, 32(2), 235-243.

Koushki PA, Al-Saleh OI, Al-Lumaia M. (2003) On management's awareness of transit passenger needs. Transport Policy, 10, 17–26.

Kuo M-, Liang G-. (2011) Combining VIKOR with GRA techniques to evaluate service quality of airports under fuzzy environment. Expert Syst Appl. 38(3):1304-1312.

Kuo M, Wu J, Pei L. (2007) A soft computing method for selecting evaluation criteria of service quality. Applied Mathematics and Computation, 189(1), 241-254.

Kuo M. (2011) A novel interval-valued fuzzy MCDM method for improving airlines' service quality in chinese cross-strait airlines. Transportation Research Part E, 47(6), 1177-1193.

Lai W, Chen C. (2011) Behavioral intentions of public transit passengers-the roles of service quality, perceived value, satisfaction and involvement. Transport Policy, 18(2), 318-325.

Landrum H, Prybutok VR. (2004) A service quality and success model for the information service industry. European Journal of Operational Research, 156, 628–642.

Lehtinen U, Lehtinen J. (1991) Two approaches to service quality dimensions. The Service Industries Journal, 11(3), 287-303.

Lin JH, Lee TR, Jen W. (2008) Assessing asymmetric response effect of behavioral intention to service quality in an integrated psychological decision making process model of intercity bus passengers: a case of Taiwan. Transportation, 35(1), 129–144.

Liou JH, Tzeng GH. (2007) A non-additive model for evaluating airline service quality. Journal of Air Transport Management, 13, 131–138.

Liou JJH, Tang C, Yeh W, Tsai C. (2011a) A decision rules approach for improvement of airport service quality. Expert Systems with Applications, 38(11), 13723-13730.

Liou JJH, Tsai C, Lin R, Tzeng G. (2011b) A modified VIKOR multiple-criteria decision method for improving domestic airlines service quality. Journal of Air Transport Management, 17(2), 57-61.

Liu J, Gao H. (2007) Study on railway transport service quality evaluation. 2007 International Conference on Wireless Communications, Networking and Mobile Computing, WiCOM 2007, 3308-3311.

López-Bonilla JM, López-Bonilla LM. (2008) Influence of the state-owned airlines on passenger satisfaction. Journal of Air Transport Management, 14(3), 143-145.

Mathisen TA, Solvoll G. (2010) Service quality aspects in ferry passenger transport - examples from Norway. European Journal of Transport and Infrastructure Research, 10(2), 142-157.

Mattsson J. (1992) A service quality model based on ideal value standard. International Journal of Service Industry Management, 3(3), 18-33.

Minser J, Webb V. (2010) Quantifying the benefits: Application of customer loyalty modeling in public transportation context. Transportation Research Record, 2144, 111-120.

Murray SJ, Walton D, Thomas JA. (2010) Attitudes towards public transport in New Zealand. Transportation, 37(6), 915-929.

Nathanail E. (2008) Measuring the quality of service for passengers on the hellenic railways. Transportation Research Part A, 42, 48–66

Nejati M, Nejati M, Shafaei A. (2009) Ranking airlines' service quality factors using a fuzzy approach: study of the Iranian society. International Journal of Quality & Reliability Management, 26(3), 247-260.

Nurul-Habib KM, Kattan L, Islaam T. (2011) Model of personal attitudes towards transit service quality. Journal of Advanced Transportation, 45, 271-285.

Oliver RL. (2010). Satisfaction: A Behavioral Perspective on the Consumer. ME Sharpe

Opricovic S, Tzeng GH. (2004) Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS. European Journal of Operational Research, 156(2), 445–455.

Ostrowski PL, O'Brien TV, Gordon GL. (1993) Service quality and customer loyalty in the commercial airline industry. Journal of Travel Research, 32, 16–24.

Pakdil F, Aydin Ö. (2007) Expectations and perceptions in airline services: An analysis using weighted SERVQUAL scores. Journal of Air Transport Management, 13, 229–237.

Paquette J, Bellavance F, Cordeau J, Laporte G. (2012) Measuring quality of service in dial-a-ride operations: The case of a canadian city. Transportation, 39(3), 539-564.

Parasuraman A, Zeithaml VA, Berry LL. (1985) A conceptual model of service quality and its implications for future research. Journal of Marketing, 49, 41–50.

Parasuraman A, Zeithaml VA, Berry LL. (1988) SERVQUAL: a multiple-item scale for measuring consumer perceptions of service quality. Journal of Retailing, 64(1), 12-40.

Parasuraman A, Zeithaml VA, Berry LL. (1991) Understanding customer expectations of service. Sloan Management Review, 32(3), 39–48.

Park J. (2007) Passenger perceptions of service quality: Korean and Australian case studies. Journal of Air Transport Management, 13, 238–242.

Park J, Robertson R, Wu CL. (2004) The effect of airline service quality on passengers' behavioural intentions: a Korean case study. Journal of Air Transport Management, 10, 435–439.

Park J, Robertson R, Wu CL. (2006) Modelling the impact of airline service quality and marketing variables on passengers' future behavioural intentions. Transportation Planning and Technology, 29(5), 359–381.

Pedersen T, Friman M, Kristensson P. (2011) Affective forecasting: Predicting and experiencing satisfaction with public transportation. J Appl Soc Psychol., 41(8), 1926-1946.

Philip G, Hazlett SA. (1997) The measurement of service quality: a new P-C-P attributes model. International Journal of Quality & Reliability Management, 14(3), 260-286.

Pullen WT. (1993) Definition and measurement of quality of service for local public transport management. Transport Reviews, 13(3), 247–264.

Ringle CM, Sarstedt M, Zimmermann L. (2011) Customer satisfaction with commercial airlines: The role of perceived safety and purpose of travel. Journal of Marketing Theory and Practice, 19(4), 459-472.

Ritchie JRB, Johnston EE, Jones VJ. (1980) Competition fares and fences perspective of the air traveler. Journal of Travel Research 18(3), 17–25

Saha GC, Theingi. (2009) Service quality, satisfaction, and behavioural intentions: A study of low-cost airline carriers in thailand. Managing Service Quality, 19(3), 350-372.

Sánchez M, Gázquez JC, Marín GM, Sánchez R. (2007) Effects of service quality dimensions on behavioural purchase intentions. A study in public-sector transport. Managing Service Quality, 17(2), 134-151

Slack N. (1994) The importance-performance matrix as a determinant of improvement priority. International Journal of Operations & Production Management, 14(5), 59–75.

Smith AM. (1995) The consumer's evaluation of service quality: an examination of the ServQual methodology. Doctoral dissertation, University of Manchester, British Thesis Service, D189377.

Spreng RA, Mackoy RD. (1996) An empirical examination of a model of perceived service quality and satisfaction. Journal of retailing, 722, 201-214.

Stradling S, Anable J, Carreno M. (2007) Performance, importance and user disgruntlement: a six method for measuring satisfaction with travel modes. Transportation Research Part A, 41(1), 98–106.

Stuart KR, Mednick M, Bockman J. (2000) Structural equation model of customer satisfaction for the New York city subway system. Transportation Research Record, 1735, 133–137.

Sultan F, Simpson MC. (2000) International service variants: airline passenger expectations and perceptions of service quality. Journal of Services Marketing, 14(3), 188-216.

Surovitskikh S, Lubbe B. (2008) Positioning of selected middle eastern airlines in the south african business and leisure travel environment. Journal of Air Transport Management, 14(2), 75-81.

Teas RK. (1993) Expectations, performance evaluation, and consumers' perceptions of quality. Journal of Marketing, 57(4), 18-34.

Transportation Research Board (1999) A Handbook for Measuring Customer Satisfaction and Service Quality, TCRP Report 47.

Transportation Research Board (2004) Transit Capacity and Quality of Service Manual, second ed.

Transport for London (2006) Network Performance of London Buses. Available at: http://www.tfl.gov.uk/buses/about/performance/service-report.asp) (accessed 6 February 2006).

Tripp C, Drea JT. (2002) Selecting and promoting service encounter elements in passenger rail transport. The Journal of Services Marketing, 16(5), 432-442.

Tsai W, Hsu W, Chou W. (2011) A gap analysis model for improving airport service quality. Total Quality Management and Business Excellence, 22(10), 1025-1040.

Tsaur SH, Chang TY, Yen CH. (2002) The evaluation of airline service quality by fuzzy MCDM. Tourism Management, 23, 107–115.

Tyrinopoulos Y, Aifadopoulou G. (2008) A complete methodology for the quality control of passenger services in the public transport business. European Transport, 38, 1-16.

Tyrinopoulos Y, Antoniou C. (2008) Public transit user satisfaction: Variability and policy implications. Transport Policy, 15(4), 260–272.

UNE-EN 13186 (2003) Transportation. Logistics and Services. Public Passenger Transport. Service Quality Definition, Targeting and Measurement. AENOR.

Vavra TG. (1997) Improving Your Measurement of Customer Satisfaction: A Guide to Creating, Conducting, Analysis, and Reporting Customer Satisfaction Measurement Programs (Milwaukee, WI: ASQ Quality Press).

Wang S, Feng C, Hsieh C. (2010) Stakeholder perspective on urban transport system service quality. Total Quality Management and Business Excellence, 21(11), 1103-1119.

Weinstein A. (2000) Customer satisfaction among transit riders. How customer rank the relative importance of various service attributes. Transportation Research Record, 1735, 123–132.

Wen CH, Lai SC, Yen W. (2008) Segmentation and positioning analysis for international air travel market: Taipei-to-Tokyo route. Transportation Research Record, 2052, 46-53.

Wen CH, Lai SC. (2010) Latent class models of international air carrier choice. Transportation Research Part E, 46, 211-221.

Wen CH, Lan LW, Cheng HL. (2005) Structural equation modeling to determine passenger loyalty toward intercity bus services. Transportation Research Record, 1927, 249–255.

Yang K, Hsieh T, Li H, Yang C. (2012) Assessing how service quality, airline image and customer value affect the intentions of passengers regarding low cost carriers. Journal of Air Transport Management, 20, 52-53.

Yeh CH, Deng H, Chang YH. (2000) Fuzzy multicriteria analysis for performance evaluation of bus companies. European Journal of Operational Research, 126, 459-473.

Yeh CH, Kuo YL. (2003) Evaluating passenger services of Asia-Pacific international airports. Transportation Research Part E, 39, 35–48.

Zeithaml VA, Berry LL, Parasuraman A. (1993) The nature and determinants of customer expectations of service. Journal of the Academy of Marketing Science, 12(1), 1–12.

List of Figures:

List of Tables:

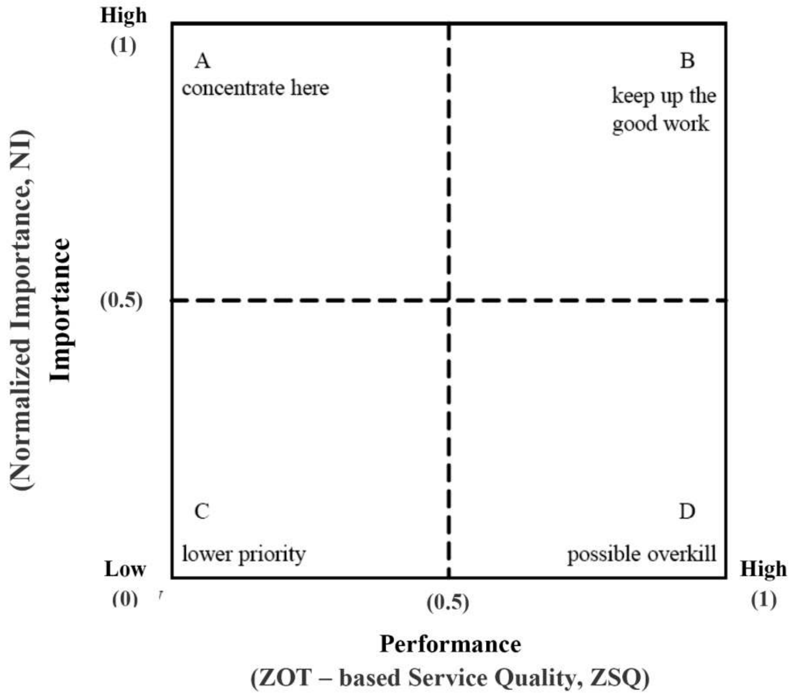

Figure 1. IPA matrix and NIZSQ matrix

TABLE 1.- Summary of previous research: PT industry, regional context, valid surveys and scale used [a].

References	PT industry	Regional context	Valid surveys	Scale used
Airlines and Airports				
Abdlla et al., 2007	Airlines	Egypt	474	9-point Likert
Aksoy et al., 2003	Airlines	Istanbul Airport (Turkey)	1.014	7-point Likert
Chang and Yeh, 2002	Airlines	Taiwan	354	11-point scale
Chau and Kao, 2009	Airlines	Taipei (Taiwan) and London (UK)	161 and 102	5-point Likert
Chen and Chang, 2005	Airlines	Taiwan	470	5-point Likert
Chen, 2008	Airlines	Taiwan	245	5- & 7-point Likert
Cheng et al., 2008	Airlines	Taiwan	252	5-point Likert
Chou et al., 2011a	Airlines	Taiwan	329	5-point linguistic
Forgas et al., 2010	Airlines	Barcelona-London corridor	1.700	n.a.
Gilbert and Wong, 2003	Airlines	Hong Kong Airport	365	8-point scale
Huse and Evangelho, 2007	Airlines	Santos Dumont Airport (Brasil)	88	10-point Likert
Kiatcharoenpol and Laosirihongthong. 2006	Airlines	Developing country	n.a.	5-point Likert
Kim and Lee, 2011	Airlines	South Korea	244	5-point Likert
Kim et al., 2011	Airlines	South Korea	231	5-point Likert
Kuo, 2011	Airlines	China-Taiwan corridor	1.635	7-point linguistic
Liou and Tzeng, 2007	Airlines	Taiwan	408	11-point scale
Liou et al., 2011b	Airlines	Taiwan	5.553	5-point Likert
López-Bonilla and López-Bonilla, 2008	Airlines	Spain	3,000 and 1,911	5-point Likert
Nejati et al., 2009	Airlines	Teheran (Iran)	231	7-point Likert
Ostrowski et al., 1993	Airlines	USA	6.000	4-point scale
Pakdil and Aydin, 2007	Airlines	Turkey	298	5-point Likert
Park et al, 2004	Airlines	Korea	592	7-point Likert
Park et al., 2006	Airlines	Sydney Airport (Australia)	501	7-point Likert
Park, 2007	Airlines	Incheon International Airport (Korea) and Sydney Airport (Australia)	592 and 501	7-point Likert
Ringle et al., 2011	Airlines	International Airport (Western Europe)	1.031	n.a.
Ritchie et al., 1980	Airlines	Calgary (Canada)	150	7-point Likert
Saha and Theingi, 2009	Airlines	Thailand	1.212	7-point Likert
Sultan and Simpson, 2000	Airlines	North Transaltantic corridor	1.956	7-point Likert
Surovitskikh and Lubbe, 2008	Airlines	Middle Eastern Airlines in South Africa	410	7-point Likert
Tsaur et al., 2002	Airlines	Taiwan	211	5-point linguistic
Wen et al., 2008	Airlines	Taipei-Tokio corridor	381	7-point Likert
Yang et al., 2012	Airlines	Taiwan	458	5-point Likert
Fernandes and Pacheco, 2010	Airports	Brazil	947	3-point linguistic
Kuo and Liang, 2011	Airports	Northeast-Asian region	23 and 26	7-point linguistic
Liou et al., 2011a	Airports	Taoyuan International Airport (Taiwan)	503	3-point Likert
Tsai et al., 2011	Airports	Taoyuan International Airport (Taiwan)	204	n.a.
Yeh and Kuo, 2003	Airports	Taiwan	15	5-point Likert
Urban and Metropolitan Public Transport				
Andreassen, 1995	Bus and rail services	Oslo Area (Norway)	1.000	7-point Likert
Christopher et al., 1999	Bus and rail services	Chicago (USA)	>2,400	5-point Likert
Foote et al., 2001	Bus and rail services	Chicago (USA)	2.464	5-point Likert
Karlaftis et al., 2001	Bus and rail services	Athens (Greece)	n.a.	n.a.
Minser and Webb, 2010	Bus and rail services	Chicago (USA)	264	5-point Likert
Tyrinopoulos and Antoniou, 2008	Bus and rail services	Athens and Thessaloniki (Greece)	1.474	4- & 5-point Likert
Figler et al., 2011	Bus services	Chicago (USA)	364	5-point Likert
Foote and Stuart, 1998	Bus services	Chicago (USA)	4.191	5-point Likert & 11-point scale
Friman, 2004	Bus services	Sweden	2.797	9-point Likert
Glascock, 1997	Bus services	Seattle (USA)	485	n.a.
Hensher et al., 2010	Bus services	Tyne and Wear area (UK)	310	5-point Likert
Jen and Hu, 2003	Bus services	Taipei (Taiwan)	235	5-point Likert

Koushki et al., 2003	Bus services	Kuwait	679	5-point Likert
De Oña et al., 2012	Metropolitan bus services	Granada (Spain)	858	3-point Likert
Eboli and Mazzulla, 2007	Metropolitan bus services	University of Calabria, Cosenza (Italy)	763 (students)	10-point Likert
Eboli and Mazzulla, 2009	Metropolitan bus services	Cosenza, Calabria (Italy)	218	10-point Likert
Eboli and Mazzulla, 2011	Metropolitan bus services	Cosenza and Rende (Italy)	123	11-point scale
Hu, 2010	Metropolitan bus services	Taipei (Taiwan)	292	7-point Likert
Friman and Gärling, 2001	PT services	Sweden	95	Number from 10 (very dissatisfied) to 90 (very satisfied)
Friman et al., 2001	PT services	Sweden	997	9-point Likert
Pedersen et al., 2011	PT services	Stockholm (Sweden)	1,007 and 169	5-point Likert & 11-point scale
Awasthi et al., 2011	Railways (subway)	Montreal (Canada)	60	5-point Likert
Lai and Chen, 2011	Railways (subway)	Kaohsiung (Taiwan)	763	5-point Likert
Stuart et al., 2000	Railways (subway)	New York (USA)	1.075	11-point scale
Weinstein, 2000	Rapid-transit system (trains)	Bay Area District, San Francisco (USA)	4.150	5-point linguistic & 7-point Likert
Dell'Olio et al., 2010	Urban bus services	Santander (Spain)	768	5-point Likert
Hu and Jen, 2006	Urban bus services	Taipei (Taiwan)	3 data collection (1: 244; 2: 292; 3: 235)	7-point Likert
Sánchez et al., 2007	Urban bus services	Almeria (Spain)	1.000	5-point Likert
Yeh et al., 2000;	Urban bus services	Taipei (Taiwan)	n.a.	5-point linguistic
Murray et al., 2010	Urban public services	Auckland, Wellington and Christchurch (New Zealand)	639	5- & 7-point Likert
Nurul-Habib et al., 2011	Urban public services	Calgary (Canada)	500	5-point Likert
Wang et al., 2010	Urban public services	Taipei Metropolitan Area	510 and 103	5-point Likert
Interurban Public Transport				
Jen et al., 2011	Interurban bus services	Taiwan	747	7-point Likert
Kuo et al., 2007	Interurban bus services	Taiwan	60	7-point linguistic
Lin et al., 2008	Interurban bus services	Taiwan	385	5-point Likert
Wen et al., 2005	Interurban bus services	Taiwan	600	5-point Likert
Cavana et al., 2007	Railways	Wellington (New Zealand)	340	9-point Likert
Drea and Hanna, 2000	Railways	USA	2.369	n.a.
Ganesan-Lim et al., 2008	Railways	Queensland (Australia)	224	7-point Likert
Liu and Gao, 2007	Railways	China	168	n.a.
Nathanail, 2008	Railways	Greece	n.a.	5-point Likert
Tripp and Drea, 2002	Railways	Illinois (USA)	2.529	7-point Likert
Chou and Kim, 2009 ; Chou et al., 2011b	Railways, high speed	Taiwan and Korea	418 and 414	10-point Likert
Others				
Paquette et al., 2012	Dial-a-ride services	Montreal (Canada)	331	10-point Likert
Mathisen and Solvoll, 2010	Ferry passenger	Norway	1.734	5-point Likert
Joewono and Kubota, 2007a ; 2007b ; 2007c	Paratransit	Indonesia	980	5-point Likert
Stradling et al., 2007	Three different transport industries	Scotland (UK)	213, 666 and 1,101	5-point Likert

n.a.: not available

[a] This is a representative, but not a comprehensive list of references.

TABLE 2. Summary of previous research on PT analyzing SQ by model type [a]

	Disaggregated models		Aggregated models [b]
	NO Importance	With Importance	
Performance perceptions and expectations [c]	Cavana et al., 2007; Chang et al., 2012; Hu and Jen, 2006	Hu, 2010; Mathisen and Solvoll, 2010; Tsai et al., 2011; Wang et al., 2010	Abdlla et al., 2007; Chang and Yeh, 2002; Chau and Kao, 2009; Chou et al., 2011a; Eboli and Mazzulla, 2009; Kiatcharoenpol and Laosirihongthong. 2006; Kuo and Liang, 2011; Kuo et al., 2007; Liou and Tzeng, 2007; Liou et al., 2011b; Liu and Gao, 2007; Nejati et al., 2009; Pakdil and Aydin, 2007; Sultan and Simpson, 2000; Tsai et al., 2011; Tsaur et al., 2002
Only performance perceptions		Chen and Chang, 2005; Christopher et al., 1999; Chou et al., 2011b; Eboli and Mazzulla, 2011; Figler et al., 2011; Foote and Stuart, 1998; Stradling et al., 2007; Weinstein, 2000	Awasthi et al., 2011; Fernandes and Pacheco, 2010; Kuo, 2011; Nathanail, 2008; Sánchez et al., 2007; Yeh and Kuo, 2003; Yeh et al., 2000

[a] This is a representative, but not a comprehensive list of references; [b] Most of them try to develop a CSI or SQI; and [c] Based on disconfirmatory theory (Parasuraman et al., 1988)

TABLE 3. Summary of previous research on PT classified by the method used to determine the weight of the attributes used in SQ analyses [a]

TECHNIQUE	PREVIOUS RESEARCH IN PUBLIC TRANSPORT
Asking for importance directly through Customer Satisfaction Surveys	
Abdlla et al., 2007; Aksoy et al., 2003; Awasthi et al., 2011; Cavana et al., 2007; Chang and Yeh, 2002; Chen and Chang, 2005; Chou et al., 2011a; Christopher et al., 1999; Eboli and Mazzulla, 2007; 2009; 2011; Fernandes and Pacheco, 2010; Foote and Stuart, 1998; Gilbert and Wong, 2003; Glascock, 1997; Hensher et al., 2010;Hu, 2010; Huse and Evangelho, 2007; Kuo and Liang, 2012; Kuo, 2011; Liou et al., 2011b; Liou and Tzeng, 2007; Mathisen and Solvoll, 2010; Nathanail, 2008; Nejati et al., 2009; Ostrowski et al., 1993; Paquette et al., 2012; Ritchie et al., 1980; Sánchez et al., 2007; Stradling et al., 2007; Sultan and Simpson, 2000; Surovitskikh and Lubbe, 2008; Tsai et al., 2011; Tsaur et al., 2002; Tyrinopoulos and Aifadopoulou, 2008; Tyrinopoulos and Antoniou, 2008; Wang et al., 2010; Wen et al., 2008; Yeh and Kuo, 2003; Yeh et al., 2000	
Model deduction from Customer Satisfaction Surveys	
Bivariate Pearson correlations	Figler et al., 2011; Weinstein, 2000
Regression analysis	***Multiple Linear Regression***: Kim and Lee, 2011; Weinstein, 2000 ***Discriminant Analysis (DV is categorical)***: Aksoy et al., 2003 ***Ordered Logit (logistic distribution)***: Tyrinopoulos and Aifadopoulou, 2008;Tyrinopoulos and Antoniou, 2008 ***Ordered Probit (normal distribution)***: Dell'Olio et al., 2010; Huse and Evangelho, 2007 ***Generalised Ordered Logit (account heterogeneity)***: Hensher et al., 2010
Structural Equation Model	Andreassen, 1995; Chen, 2008; Cheng et al., 2008; Chou and Kim, 2009; Chou et al., 2011b; Eboli and Mazzulla, 2007; 2012b; Friman and Gärling, 2001; Friman et al., 2001; Jen et al., 2011;Joewono and Kubota, 2007a; 2007c; Karlaftis et al., 2001; Kim and Lee, 2011; Lai and Chen, 2011; Minser and Webb, 2010; Nurul-Habib et al., 2011; Park et al., 2006; Saha and Theingi, 2009; Stuart et al., 2000; Tripp and Drea, 2002; Wen et al., 2005; Yang et al., 2012
Path analysis	Forgas et al., 2010; Jen and Hu, 2003; Joewono and Kubota, 2007b; Lin et al., 2008; Park et al., 2004; Ringle et al., 2011

[a] This is a representative, but not a comprehensive list of references.

Table 4. Summary of key advantages and disadvantages

Model	Advantages	Disadvantages
SERVQUAL	Most basic model; allow to obtain an overall index	Use many concepts (may be confusing and increase surveys' length); all attributes are equally important; do not account for heterogeneity; changes in individual components may be masked
Weighted SERVQUAL	Different weight for each attribute; allow to obtain an overall index	Use many concepts (may be confusing and increase surveys' length); do not account for heterogeneity; changes in individual components may be masked
Fuzzy weighted SERVQUAL	Different weight for each attribute; handle subjective information; allow to obtain an overall index	Use many concepts (may be confusing and increase surveys' length); do not account for heterogeneity; changes in individual components may be masked; complex process
CSI	Different weight for each attribute; allow to obtain an overall index	Do not account for heterogeneity; changes in individual components may be masked
HCSI	Different weight for each attribute; account for heterogeneity; allow to obtain an overall index	Changes in individual components may be masked
Multicriteria Analysis (MA) (Satisfaction)	Allow to obtain an overall index	Do not account for heterogeneity; changes in individual components may be masked
MA – TOPSIS	Allow to obtain an overall index	All attributes are equally important; do not account for heterogeneity
MA – Fuzzy TOPSIS	Handle subjective information; allow to obtain an overall index	All attributes are equally important; do not account for heterogeneity; complex process
MA – VIKOR	Different weight for each attribute; allow to obtain an overall index	Do not account for heterogeneity
SERVPERF	Most basic model; allow to obtain an overall index	All attributes are equally important; do not account for heterogeneity; changes in individual components may be masked
Weighted SERVPERF	Different weight for each attribute; allow to obtain an overall index	Do not account for heterogeneity; changes in individual components may be masked
Fuzzy weighted SERVPERF	Different weight for each attribute; handle subjective information; allow to obtain an overall index	Do not account for heterogeneity; changes in individual components may be masked; complex process
IPA	Most basic model; easily-interpreted (graphical tool); different weight for each attribute; allow to set priorities for improvements	Visualizing method with no precise ranking of priority; do not account for heterogeneity; if importance is stated by passengers, almost all attributes are crowd together at the top of the grid
n.a. (Eboli and Mazzulla, 2011)	Set priorities for improvements; account for heterogeneity; use conjointly subjective and objective data	All attributes are equally important
ZOT	Set priorities for improvements	Use many concepts (may be confusing and increase surveys' length); all attributes are equally important; do not account for heterogeneity
Fuzzy ZOT	Set priorities for improvements; handle subjective information	Use many concepts (may be confusing and increase surveys' length); all attributes are equally important; do not account for heterogeneity; complex process
NIZSQ	Different weight for each attribute; set a precise ranking of the priority of improvements based on the ZOT and normalized importance	Use many concepts (may be confusing and increase surveys' length); do not account for heterogeneity; complex process