Transport Policy 49 (2016) 68–77

Contents lists available at ScienceDirect

Transport Policy

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

The importance of service quality attributes in public transportation: Narrowing the gap between scientific research and practitioners' needs

Begoña Guirao [a,*], Antonio García-Pastor [b], María Eugenia López-Lambas [a]

[a] Departamento de Ingeniería Civil: Transportes y Territorio, ETSI Caminos, Canales y Puertos, Universidad Politécnica de Madrid, Avenida Profesor Aranguren, s/n., 28040 Madrid, Spain
[b] AVANZA Grupo, C/San Norberto, 48, 28021 Madrid, Spain

ARTICLE INFO

Article history:
Received 5 January 2016
Received in revised form
11 March 2016
Accepted 4 April 2016

Keywords:
Public transport
Customer satisfaction Surveys (CSS)
Service Quality (SQ)
User Perception
Factorial Analysis
MIMIC models

ABSTRACT

Customer Satisfaction Surveys (CSS) have become an important tool for public transport planners, as improvements in the perceived quality of certain service attributes can lead to greater use of public transport and lower traffic pollution. The literature shows that the importance of quality attributes has until now been estimated indirectly, as they are derived from the Customer Satisfaction Index using various different and complex techniques. Little work has been dedicated to its direct estimation (stated importance) by designing ad-hoc surveys, an approach that represents a considerable reduction in the length of the questionnaire.

This paper contributes to the limited existing literature by developing a survey technique based on hierarchy processes to estimate the stated importance of quality attributes, and compares the results with the derived importance obtained using conventional surveys with the same sample. The added value of this research is that it provides the first comparison between two quality survey methods using the same real case study in Madrid (Spain). The results achieved using this pioneer survey method (293 valid questionnaires) were validated using conventional face-to-face surveys (520 valid questionnaires). Factorial analysis, multiple regression analysis and Multiple Indicators Multiple Causes (MIMIC) models were applied to the conventional survey sample to analyse and derive the importance of the attributes. The results clearly show that, after a few teething troubles, the stated importance of quality attributes can be estimated directly, thus providing transport management companies with a simple and useful tool to implement in their Customer Satisfaction Surveys (CSS), and narrowing the gap between practitioners' needs and scientific research.

1. Introduction

The analysis of Service Quality (SQ) is of vital importance for both operators and public transport authorities, as the increase in SQ in public transport has been shown to play a key role in attracting new passengers from private cars to the public transport system and in reducing traffic pollution as a result (Transportation Research Board, 1999). The literature reveals a significant gap between the scientific research and practitioners' needs. Scientific research regards the concept of SQ as complex, fuzzy and abstract, mainly due to the three aspects of service: intangibility, heterogeneity for each individual, and the inseparability of production and consumption (Parasuraman et al., 1985). Most scientific methodologies for analysing SQ are applied only to customer perception surveys, although a number of authors (Grönroos, 1988) differentiate between consumer expectations and perception of service during the trip, and maintain that the perception of SQ is the result of the comparison of consumer expectations with actual service performance. Other authors such as Hu (2010) define service quality in terms of the difference between perceived and tolerable quality.

Leaving aside this theoretical approach, the study of quality took a giant stride forwards in linking the fields of research and practice with the QUATTRO project entitled “Quality Approach in Tendering Urban Public Transport Operations” (European Commission, 1998), whose objective was to define and introduce quality indicators into tendering and contracting in public transport services. The QUATTRO project was also the basis for the European Standard EN 13816 Quality of service in passenger transport services (2003), but provided a more practical concept of SQ. Four quality levels were identified in the QUATTRO project: expected quality, perceived quality, targeted quality and delivered quality. The level of quality desired (expected) by passengers and

* Corresponding author.
E-mail addresses: bguirao@caminos.upm.es (B. Guirao), agarciapa@avanzagrupo.com (A. García-Pastor), mariaeugenia.lopez@upm.es (M.E. López-Lambas).

http://dx.doi.org/10.1016/j.tranpol.2016.04.003
0967-070X/

citizens in general may be different from the perceived quality – observed with varying degrees of objectivity – by the passengers during their journeys. The level of quality the company wishes to achieve (targeted quality) is determined by external and internal pressures, expected quality, budgetary constraints and competitors' performance. Finally, the delivered quality is the level of quality obtained on a daily basis in real operating conditions.

Likewise, the emphasis in the US was also on measuring service quality through customer satisfaction, as evidenced by the Handbook for Measuring Customer Satisfaction and Service Quality (TRB, 1999) and the Transit Capacity and Quality of Service Manual (TRB, 2004).

The only objective information for operating companies is the provided quality, normally established in the concession contracts; however Customer Satisfaction Surveys (CSS) are also conducted on a yearly or six-monthly basis to monitor the users' perception of the service. The data collected from the CSS are used to analyse the company's operations, and provide useful information on overall service quality. They are also used in research and by academics to focus on the mathematical analysis of perceived quality, and test a large number of indicators: from simple indices such as SERVQUAL (Chau and Kao, 2009; Chou et al., 2011), SERVPERF (Sánchez et al., 2007), Customer Satisfaction Index CSI (Hill et al., 2003) and Heterogeneous Customer Satisfaction Index HCSI (Eboli and Mazzulla, 2009) to other more complex indices obtained by applying econometric models to satisfaction rates, such as Structural Equation Models (SEM) and discrete choice models (Hensher at al., 2003; Román et al., 2014). Some authors (Del Castillo and Benitez, 2013; Celik et al., 2014) have used mixed methodologies to determine the quality of the bus service. Del Castillo and Benitez (2013) used three models simultaneously (weighted means, a multivariate discrete distribution and a generalised linear model), while Celik et al. (2014) integrated statistical analysis, SERVQUAL, interval type-2 fuzzy sets and VIKOR (Opricovic and Tzeng, 2004) to evaluate customer satisfaction with the rail transit network in Istanbul.

The more complex the indicators (and the techniques to obtain them), the less likely it is that practitioners will be able to understand and use them in practice. Due to its simplicity, CSI, based on the importance of "attributes" and satisfaction rates, is the most widely applied index – even by operating companies – to determine service quality in public transportation. From a marketing point of view, an attribute is a characteristic or feature of a product that is thought to appeal to customers. In the public transport sector, the term service attribute is commonly used to refer to cleanliness, on-time performance, availability, comfort or security, constituting the criteria applied to assess customer service quality.

As the main tools for analysing service quality in public transport are based on CSS, the design of the questionnaire is absolutely crucial and depends strongly on the service attributes to be considered and on the approach used to estimate the relative importance of the attributes to the customers. This relative importance is another key point, as once a group of attributes is selected for a specific survey, public transport operators and service industries need to know not only how the users rate the service in terms of detailed service attributes (attribute-performance rating), but also the relative importance of these attributes to their customers (attribute-importance measures). The CSS results can help managers choose from a long list of service attributes (e.g. cleanliness, on-time performance, availability, comfort or security) so they can target their organisation's attention and resources more effectively. The rates are normally expressed in two scales, numeric and linguistic. Numeric scales are more commonly used and have a wider range, from 3 to 11 points. Linguistic scales are used less, and have a narrower range, from 3 to 7 points (the 5-point Likert scales are the most widely adopted).

The survey is usually designed by the operating companies, and the resulting database is used first by the companies and then passed on to researchers. This highlights the need to narrow the gap between theory and practice. There is currently no proper debate on the design and format of the survey, and in most situations researchers use only part of the survey results, as their modelling tools are only suited to a specific database. There is also a problem from the scientific point of view due to the lack of critical comparison – using the same case study – between the competing techniques in order to analyse user perception and attribute importance.

This paper contributes to the limited existing literature by developing a survey technique based on hierarchy processes to estimate the stated importance of quality attributes, and compares the results with the derived importance obtained with the same sample using conventional surveys. The added value of this research resides in the fact that it is the first comparison between two quality survey methods using data from the same case study, in Madrid (Spain). The article has been divided into the following parts in order to describe the research as a whole: introduction (Section 1); a review of the literature on the methods used to estimate attribute importance (Section 2); a description of the case study and conventional survey campaign (Section 3); the proposed new stated survey method and its application to the case study (Section 4); the validation of the stated importance survey using the conventional survey (Section 5); and finally, the presentation of the most important conclusions and recommendations (Conclusions section).

2. Estimation of attribute importance using CSS

The literature shows that a considerable number of attributes are used to evaluate SQ, so they are normally grouped into a smaller number, called dimensions. Although there is no general agreement as to the nature or content of SQ dimensions, it is generally recognised that service quality is a multidimensional (Lehtinen and Lehtinen, 1982), multilevel or hierarchical (Brady and Cronin, 2001) construct. Various papers (Eboli and Mazzulla, 2007) have pointed to several categories of attributes that have a greater or lesser impact on SQ and satisfaction. In 2002 the European Committee for Standardization CEN (2003) established a quality standard – EN 13816 Service Quality Standard for Public Transport – in connection with the QUATTRO research, and published a final report. The UNE-EN 13186 standard classifies the characteristics of a service into basic, proportional and attractive, depending on how compliance and non-compliance affects customer satisfaction. In the US, the Transit Capacity and Quality of Service Manual TCQSM (Transportation Research Board, 2004) groups attributes into availability factors, and comfort and convenience factors. The primary distinction made by the TCQS is whether a transit service is offered, and if it is, customers then consider both the type of availability (e.g. frequency or access), and its comfort and convenience. In practice, the choice of variables is far from straightforward and usually derives from exhaustive lists of attributes (for instance the one included in UNE EN 13816), although some are chosen through other CSS. Some authors like Dell'Olio et al. (2010) recommend identifying the attributes to be included in the CSS independently. Focus-group methodologies are suitable for this objective, although they are costly and require a separate prior study. The heterogeneity of the users and services must also be taken into account when analysing CSS results, as demonstrated by Bordagaray et al. (2014) when modelling bus transit quality in the city of Santander.

Apart from the selection of attributes, the design of a CSS survey depends strongly on the approach used to estimate the relative importance of the attributes to the customers. In conventional CSS designed by companies to obtain a general satisfaction index (CSI), it is necessary to consider both the attribute-performance rating and attribute-importance measures when the operator's priority is to improve or sustain the current overall SQ. This dual target often requires a long questionnaire, although researchers only use the results of the first part (attribute-performance rating), as the attribute importance can be indirectly derived from the attribute-performance rating. This problem has already been debated in the literature, as described below.

Weinstein (2000) was the first author to clearly distinguish two main approaches to estimate attribute importance: stated importance and derived importance. Stated importance involves asking customers to rate each attribute on a scale of importance; this is the more intuitive and direct of the two methods, but requires a significant increase in the length of the questionnaire (which can lower the overall response rate and the accuracy of the survey). It can also sometimes fail to differentiate sufficiently between mean importance ratings; if customers score nearly all the measures near the top of the scale, certain attributes may be rated as important even though they in fact have little influence on overall satisfaction. As this is the more intuitive and direct of the two methods, operating companies have tended to use this type of questionnaire, while the scientific research has focused on more complex methodologies using the derived importance approach.

The derived importance approach is less intuitive and is based on "deriving" a measure of attribute importance by statistically testing the strength of the relationship of individual attributes with overall satisfaction. A simple conventional attribute rating survey is needed to derive importance, and this type of questionnaire is always included in the CSS. Recent literature is now set on seeking other alternatives to the methods commonly used until now to derive importance, namely; (a) bivariate Pearson correlations, (b) factor analysis, and (c) multiple regression analysis. These other alternatives include structural equation models (SEM), based on a multivariate technique combining regression, factor analysis and analysis of variance to estimate interrelated dependence relationships simultaneously. This approach allows a phenomenon to be modelled by considering both the unobserved "latent" constructs and the observed indicators that describe the phenomenon. SEM has also been adopted to measure customer satisfaction in several public transport services such as metropolitan public transport (Lai and Chen, 2011; Shen et al., 2016). More recently, de Oña et al. (2012) have used decision trees to derive attribute importance in public transport quality, and a new methodology of "index numbers" has been developed to monitor the evolution of attribute importance throughout successive CSS (de Oña et al., 2016). However, these last complex methodologies are not based on stated attribute importance from the CSI, but on derived importance. As far as the authors are aware, there are no studies comparing the different methodologies for obtaining attribute importance using the same case study data (or even a comparison between the most commonly used derived importance methodologies).

The possibility of comparing techniques and estimating stated importance has been practically abandoned by academics, but other survey formats could have been tested and studied, such as ranking attributes using hierarchy process together with stated preference techniques. Analytic Hierarchy Process (AHP) is a general theory of measurement used to derive ratio scales from both discrete and continuous paired comparisons (Saaty, 1987), which may be taken from actual measurements or from a basic scale that reflects the relative strength of preferences and feelings. Pairwise comparisons are fundamental in the use of AHP, although this theory can be extrapolated to a three-option choice. Three principles guide problem-solving using the AHP: decomposition, comparative judgments and synthesis of priorities. In our service quality case study, the decomposition is based on the selection of the attributes to be ranked and on the comparative judgments given by the surveys. The AHP priorities are synthesised from the second level down by multiplying local priorities by the priority of their corresponding criterion in the level above, and adding a level for each element according to the criteria it affects. This gives the composite or global priority of that element, which in turn is used to weight the local priorities of the elements in the level below.

Aydin et al. (2015) recently used a type of AHP methodology, FAHP (Fuzzy Analytic Hierarchy Process), to measure the performance of rail transit lines in Istanbul; however the FAHP was applied to fix the weights of the main "criteria" (train comfort, ticketing, information system, accessibility, station comfort, fare and time) based on the unbiased opinions of experts. The weights of the sub-criteria were simultaneously calculated by trapezoidal fuzzy numbers based on customer responses. The FAHP was therefore not directly applied to the CSS itself, as in our case.

In designing the survey questionnaire using an AHP process, some practical ideas have been borrowed from the stated preferences experiments in transportation described by Saako (2001) in order to collect useful data with as little bias as possible. Stated preference surveys have been used in transportation to analyse alternative trip choices (each alternative is composed of various attributes), but we have found no literature that ranks simple quality attributes, although statistically the problem to be solved is fairly similar. Ampt and Meyburg (1995) suggest a maximum of 9–16 options as acceptable in this type of stated preference surveys, with most current designs now adopting the lower end of this range. With a maximum of nine options for the respondent to ponder, this severely limits the number of attributes that can be considered. Our Customer Satisfaction surveys consider over 10–15 attributes, so this limitation must be overcome while allowing the consideration of more attributes and/or more attribute levels. One of the strategies proposed by Pearmain et al. (1991) in stated preference surveys is to separate the options into "blocks", so that the full choice set is completed by groups of respondents, but with each group responding to a different sub-set of options. Each group responds to a full-factorial design within each sub-set of options, and the responses from the different sub-groups can be assumed to be sufficiently homogeneous to provide the full picture when combined.

As part of a research project led by the Madrid Polytechnic University, the authors of this paper had the opportunity to design an ad-hoc CCS, based on this previous literature, in a Spanish case study: the Madrid-Tres Cantos corridor, with four urban bus lines (operated by the company ALSA). A new type of survey questionnaire (to state importance) was tested using a more sophisticated process of hierarchy, separating the options into blocks and reducing the length of the survey questionnaire (not all users were asked for the same attribute ranking). In order to validate this new stated importance method, a conventional survey was also required (designed to derive importance), and the whole campaign was based on face-to-face surveys (293 surveys to state attribute importance and 520 to derive importance). As the face-to-face survey campaign was starting to become very costly, additional research based on Quick Response (QR) code surveys was also implemented in the study. A third type of questionnaire was therefore designed for the QR survey (also derived-importance) and uploaded to the operating company's (ALSA) website. The QR code is a simple way of providing the user with a virtual link to the questionnaire in order to test how to reduce the cost of future SQ survey campaigns using new Intelligent Transport Systems (ITS). The results of the QR research have been published recently (Guirao et al., 2015), and this article shows the results of the main part of the research project, based on stated importance.

3. The case study: CSS in a bus corridor in Madrid

The initiative to conduct surveys in different formats among urban bus users came from a research project led by the Madrid Polytechnic University (UPM). The campaign was carried out in March 2013 in four periurban bus lines along the Madrid-Tres Cantos corridor operated by the company ALSA. This corridor leads towards the north of Madrid, starting from the interchange at Plaza de Castilla and connecting the UAM (Universidad Autónoma de Madrid), El Goloso and Tres Cantos along a length of 20 km, as shown in Fig. 1. Bus lines 712, 713 and 716 connect the Madrid public transport interchange hub in Plaza de Castilla to the city of Tres Cantos along the M-607 corridor (M-607 is a dual carriageway with two lanes in each direction). The last part of the route, now in Tres Cantos, divides into different routes inside the city. Line 714 is a special case, since it connects the interchange hub to a university campus (Universidad Autónoma de Madrid – UAM) a few kilometres outside the city, meaning this bus service is a specialised line for trips for the purpose of study.

Suburban bus services in Spain are usually tendered in route bundles according to factors like line proximity, feasibility of lines or in order to avoid overlapping. Contracts are normally subject to European Regulation 1370:2007 on public passenger transport services by rail and by road, which envisages the concept of Public Service Obligations (PSO) and other national requirements (LOTT, 1987). In the case study, the bundle is a combination of purely metropolitan lines connecting the city of Madrid at one of the mayor interchanges (Plaza de Castilla) with the municipality of Tres Cantos, a city located 18 km north of Madrid with a population of 50,000. Table 1 shows the main characteristics of the bus lines included in this case study.

Two previous groups of questionnaires were designed in order to achieve the objectives of the research project: one to determine the derived attribute importance (group 1), and the other to find the stated importance (group 2). Over 850 face-to face surveys were finally conducted following various parameters of statistical significance and maximum error, from which 813 observations were drawn as valid (520 from group 1 and 293 from group 2). These results allowed the quality analysis to be completed with a sufficient sample size for the planned objectives. The pilot survey was carried out on February 20, 2013 and definitive surveys were made throughout the last two weeks of March from 6 am to 11 am (18.3% of the sample), 11:01 am to 4:40 pm (64.8%) and 4:41–11 pm (16.9%), at the main bus stops (Plaza de Castilla interchange hub, La Paz Hospital, Ramón y Cajal Hospital, Einstein-Rectorado UAM) and on board. Table 2 shows the sample rate for each line for survey group 1 (designated "conventional survey" in our research on QR codes). These sample rates present errors of around 5–7% for high confidence intervals. Bus line 714 has a distinct student dimension, and while the sample rate is low, the results are still considered sufficient for the analysis. All the bus lines have a similar age and gender distribution except for line 714, which – as it is used mainly by students – has a higher percentage of young users; it also carries more women than men. Table 2 shows the number of valid questionnaires per user and trip profile (ticket type, gender, activity, frequency, age and trip purpose) in the conventional survey, with their percentages. An additional questionnaire (group 3) was designed to test the implementation of QR codes in web-based surveys. Only group 1 was used to validate the QR surveys, as the format was comparable.

In the conventional survey carried out at the time and in the place mentioned above, the users were asked to rate the following 15 attributes in addition to the overall level of satisfaction with the

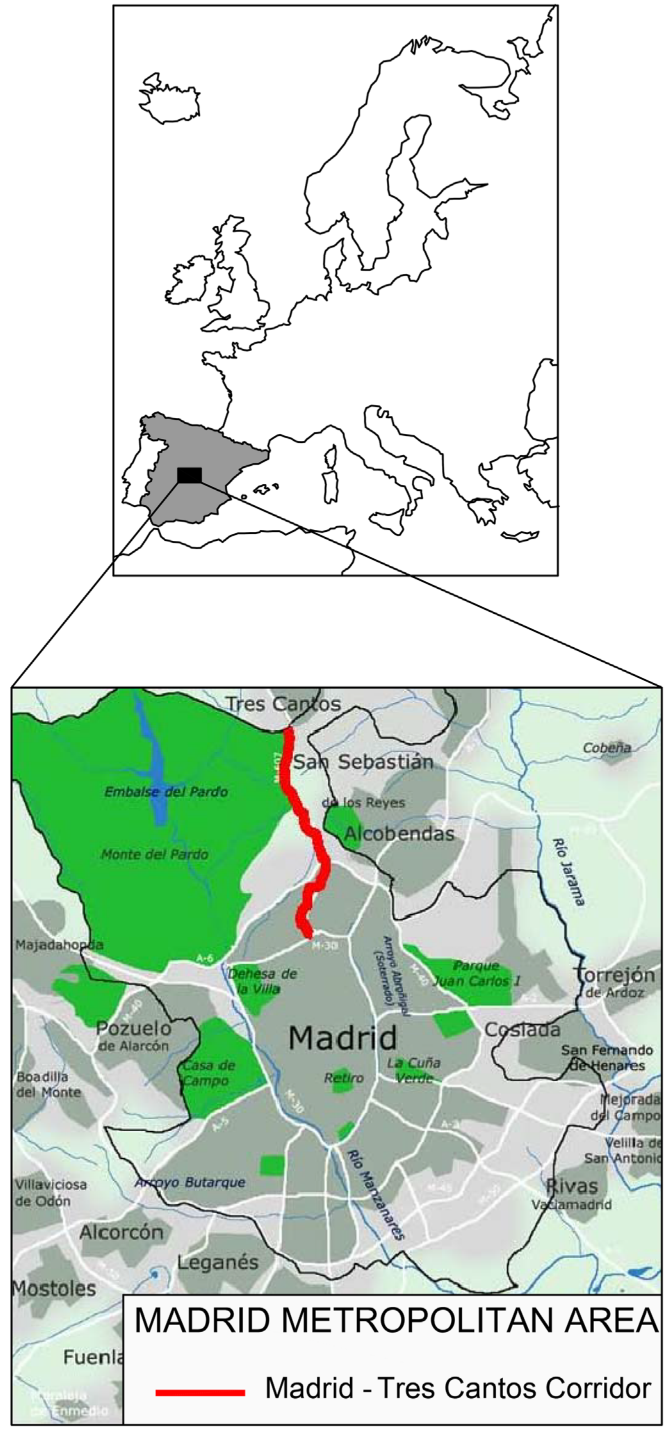

Fig. 1. Location of the Madrid-Tres Cantos corridor (M-607 dual carriageway) in Spain.

Table 1
Main characteristics of the suburban lines (Madrid-Tres Cantos corridor).

Line	Length (km)	Travel time (min per way)	Headway (min)	Yearly passengers (2012)
712	22.3	45	15	1,050,901
713	21.3	45	15	879,525
714	11.5/13.0	35	12	687,099
716	22.8	35	20	651,455

Table 2
Conventional survey collection per bus line. Sample rates and questionnaires collected per user and trip profile.

Sample rate estimation	Bus Line				Total
	712	713	714	716	
Workday demand (trips)	4106	3072	3250	3160	13,588
No. of surveys collected	207	116	91	106	520
Sample rate	5%	3.8%	2.8%	3.4%	3.8%
Number of valid questionnaires per user and trip profile					
User activity					
Working	112 (54.1%)	68 (58.6%)	17 (18.7%)	62 (58.5%)	259 (49.8%)
Unemployed	11 (5.3%)	6 (5.2%)	1 (1.1%)	2 (1.9%)	20 (3.8%)
Retired	26 (12.6%)	9 (7.8%)	6 (6.6%)	6 (5.7%)	47 (9.0%)
Student	43 (20.8%)	26 (22.4%)	67 (73.6%)	29 (27.4%)	165 (31.7%)
Other	15 (7.3%)	7 (6.0%)	0 (0.0%)	7 (6.6%)	29 (5.6%)
Ticket					
Single	10 (4.8%)	6 (5.2%)	0 (0.0%)	7 (6.6%)	23 (4.4%)
10 trips	16 (7.7%)	10 (8.6%)	2 (2.2%)	5 (4.7%)	33 (6.3%)
Season ticket	176 (85.0%)	99 (85.3%)	89 (97.8%)	94 (88.7%)	458 (88.1%)
Other	5 (2.4%)	1 (0.9%)	0 (0.0%)	0 (0.0%)	6 (1.2%)
Frequency of trip					
≥ 5 days	142 (68.6%)	84 (72.4%)	65 (71.4%)	73 (68.9%)	364 (70.0%)
3–4 days	22 (10.6%)	14 (12.1%)	13 (14.3%)	11 (10.4%)	60 (11.5%)
1–2 days	31 (15.0%)	9 (7.8%)	10 (11.0%)	13 (12.3%)	63 (12.1%)
Less than 1 d	12 (5.8%)	9 (7.8%)	3 (3.3%)	9 (8.5%)	33 (6.3%)
Trip purpose					
Work	117 (56.5%)	65 (56.0%)	15 (16.5%)	63 (59.4%)	260 (50.0%)
Study	38 (18.4%)	23 (19.8%)	71 (78.0%)	25 (23.6%)	157 (30.2%)
Medical	11 (5.3%)	8 (6.9%)	0 (0.0%)	4 (3.8%)	23 (4.4%)
Leisure	10 (4.8%)	3 (2.6%)	0 (0.0%)	3 (2.8%)	16 (3.1%)
Other	31 (15.0%)	17 (14.7%)	5 (5.5%)	11 (10.4%)	64 (12.3%)
Age					
≤ to 23	48 (23.2%)	22 (19.0%)	60 (65.9%)	30 (28.3%)	160 (30.7%)
From 23 to 35	59 (28.5%)	33 (28.4%)	19 (20.9%)	24 (22.6%)	135 (25.9%)
From 36 to 50	38 (18.4%)	30 (25.9%)	7 (7.7%)	29 (27.4%)	104 (20.0%)
≥ 50	62 (30.0%)	31 (26.7%)	5 (5.5%)	23 (21.7%)	121 (23.2%)
Gender					
Male	66 (31.9%)	37 (31.9%)	33 (36.3%)	41 (38.7%)	177 (34.0%)
Female	141 (68.1%)	79 (68.1%)	58 (63.7%)	65 (61.3%)	343 (66.0%)
TOTAL	207 (39.8%)	116 (22.3%)	91 (17.5%)	106 (20.4%)	520 (100%)

service:

- Route (bus route).
- Connections (connection with other lines and transport modes).
- Punctuality (on-time performance).
- Frequency (timetable and headway).
- Access (ease of access to the bus stop from origin –home, work, university, etc.).
- Information-incidents (delays, breakdowns, changes in the line, etc.).
- Cleanliness (cleanliness of the bus).
- Information-service (timetables, routes, etc.).
- Journey time (of the route).
- Comfort (air conditioning, seating, etc.).
- Information and communication technologies (ICTs) (internet on board, mobile payment, real-time information screens both on board and at stops).
- Shelters (along the route).
- Bus driving security.
- Customer attention from the bus driver.
- Possibility of sitting during the journey.

Faced with the impossibility of developing a focus group of corridor users, twelve attributes were selected based on a sociological study (on SQ attributes for periurban lines) carried out by the Madrid's Regional Transport Consortium (CRTM, 2005). Finally the last three on the list (bus driving security, customer attention from the bus driver and the possibility of sitting during the journey) were introduced at the request of the operating company and located in a different part of the conventional survey.

The statistical mode and median of the results of the analysis of these abovementioned bus lines show that most of the variables have an average and median with the semantic meaning “Good”. Only the variable “Frequency” has a semantic value “Not Good” for the median, which indicates the importance of this variable and how it is valued by respondents. The statistical analysis by line does not reveal any substantial difference, except in the case of the valuation of ICTs by the users of bus 714, who describe it as “very good”. A preliminary aggregated analysis of the conventional survey is shown in Table 3, with the average rating of each attribute-performance. The three best rated attributes (over 7.0 out of 10.0) are bus cleanliness, access to bus stops and the possibility of sitting during the journey, while the three worst rated are ICTs, information about incidents and frequency. It should be noted that global customer satisfaction on this line is high (7.0 out of a maximum score of 10.0). These data allow the importance of each attribute to be estimated mathematically, although the most intuitive and direct method for operating companies would be to ask the customers directly which attributes they consider more important from a general point of view – not necessarily linked to their trip experience – when answering the survey. Although

Table 3
Average rating of each attribute-performance in Madrid-Tres Cantos corridor.

Rated variables	Rating (over 10)
Cleanliness	7.72
Access	4.57
Possibility of sitting during the journey	7.48
Journey time	7.36
Customer attention from the bus driver	7.28
Comfort	7.04
Connections	7.00
Punctuality	6.96
Bus driving security	6.86
Route	6.85
Information-service	6.81
Shelters	6.75
Frequency	5.64
Information-incidents	4.59
Information and communication technologies (ICTs)	3.28

operating companies have tended to use – and most continue to use – this type of "stated format", the required length of the questionnaire is excessive and can lower the overall response rate and the accuracy of the survey. The following section contains a proposed design for a new type of stated important questionnaire together with its application to the case study.

4. A proposal for a stated importance survey

The stated importance survey was carried out in the Madrid-Tres Cantos corridor in March 2013, but on a different date from the conventional survey, in order to avoid biases or "contamination" between them. The new questionnaire was designed to include the same 15 attributes as in the conventional survey but these were offered to the customers in four different sub-sets of attributes (blocks) according to the literature review and in order to reduce the length of the survey. The customers were asked to identify the three most important attributes in each sub-set, and to rank them in descending order of importance. This solution allowed the number of attributes to be reduced to a smaller ranking, thereby improving the reliability of the survey process. The first questions in the survey concerned user and trip profile, and these were common to all the users surveyed. In contrast, the attribute importance questions were organised in four scale cards and the customers were assigned only one, with no more than eight attributes. One of the main problems with earlier long stated preference surveys was that they sometimes failed to differentiate sufficiently between mean importance ratings if customers rated nearly all the measures near the top of the scale. Certain attributes could therefore be rated as important even though they in fact have little influence on overall satisfaction. To avoid this type of bias, the attributes in each card and their order of appearance were selected according to the following guidelines:

- Each card includes a total of seven or eight attributes (almost half the total attributes).
- Each attribute appears only twice; that is, on only two of the four available cards.
- Each time an attribute appears, attempts were made to change its order of appearance, alternating the top and bottom positions in the cards. To achieve this target, all the attributes meet the requirement that the difference between their two appearances is at least two positions (at the top or bottom of the scale).

Table 4 shows the four scale cards used in the case study (namely A1, A2, A3 and A4). 293 valid surveys were collected with this type of questionnaire: 79 from Card 1 (A1), 79 from Card 2 (A2), 74 from Card 3 (A3), and 61 from Card 4 (A4). The number of cards collected per bus line ranged from 41.98% on Line 712–18.77% on Line 714 (with 19.45% on Line 713% and 19.8% on Line 716). This means that each survey has an error of around 11% for high confidence intervals. However, as each card has a scale with several attributes, we obtained three pairs of discrete choices per user, and the error per survey thus drops to 6% for a confidence interval of 95.5%. Obviously, this simplification would have been unnecessary had we collected a higher number of stated importance surveys, but does not invalidate the results. Moreover, the user profile registered in the stated survey is consistent with the one obtained through the conventional survey and with the information on demand provided by the operating company.

Table 4 shows the structure of each card and a preliminary analysis of the attribute importance results depending on the number of times the attribute is in first, second and third position. Each time an attribute is in first place in a survey, it is assigned a value of 3.0. This value is 2.0 in second place, and 1.0 in third place. Table 4 shows the score given to each attribute for each type of card. The number of valid surveys obtained per card must be taken into account to guarantee statistically robust results. Each card contains seven or eight attributes and it is also necessary to average (or weight) the number of times an attribute appears in the top three positions. For example, in card 1 the score for punctuality has been divided by 474 (the sum of all the scores in this card); this percentage (out of ten) is shown in the last column of Table 4. Once these values have been calculated, the scores are aggregated for each attribute from two different cards, but considering the total range of scores; the highest score corresponds to the "punctuality" attribute on card 1 (4.43) while the lowest corresponds to "ICTs" on card 2. We therefore assigned the value 10.0 to the highest score and 0.0 to the lowest, interpolating the intermediate scores (see the last column in the table). Table 5 shows the final aggregation per attribute, and the ranking of attributes in terms of their importance for users.

Punctuality, frequency and driving security can be seen to be the three most important attributes for customers, while ICTs, bus driver attention and incident information appear at the bottom of the table. According to Table 3, two of the least important attributes for users are also the worst rated in the conventional survey (ICTs and incident information). After defining this pioneer survey tool, we validated and analysed our results using the conventional survey database for the same corridor.

5. Validation of the stated importance survey

The stated importance survey was validated based on the conventional survey analysis, in which the same 15 attributes were rated using a 5-point Likert scale and subsequently normalised with a 0–10 scale during data processing. The number of valid questionnaires collected in the conventional survey (520) shows a uniform error of 4.4% for a confidence interval of 95.5%.

Before deriving the attribute importance mathematically from the conventional survey, the valid surveys were analysed in depth using different statistical techniques. An independence test was first carried out considering the different bus lines to check that the samples were independent and unbiased. As the variables were categorical, this test was done by estimating the Pearson Chi-squared (χ^2). The Chi-square goodness-of-fit test revealed that there is sample independence for most of the variables; that is, the survey answers do not depend on the chosen segment although there are variables that show different behaviours. For example, bus route perception, frequency and information-service depend on the bus line considered. Age affects the perception of the connection to other transport modes, frequency and information-

Table 4
Structure of the four ranking cards in the stated important survey, and number of times an attribute appear in the first, second and third position of the ranking.

Card 1	First position	Second position	Third position	Score 1	Score 2 (over the card)	Score 3 (over all cards)
Punctuality	61	12	3	210.00	4.43	10.00
Information-service	4	26	9	73.00	1.54	3.15
Cleanliness	4	7	10	36.00	0.76	1.30
Shelters	1	4	5	16.00	0.34	0.30
Access	3	11	17	48.00	1.01	1.90
Journey Time	4	12	15	51.00	1.08	2.05
Route	2	7	20	40.00	0.84	1.50
Card 2						
Frequency	47	7	5	160.00	3.38	7.50
Bus seating	12	15	7	73.00	1.54	3.15
Journey time	6	20	3	61.00	1.29	2.55
Comfort	2	11	10	38.00	0.80	1.40
Information-service	7	17	22	77.00	1.62	3.35
Information-incidents	2	1	19	27.00	0.57	0.85
Bus driver attention	3	5	9	28.00	0.59	0.90
ICTs	0	3	4	10.00	0.21	0.00
Card 3						
Bus driving security	34	9	8	128.00	2.88	6.33
Information-incidents	4	5	8	30.00	0.68	1.10
Route	4	7	7	33.00	0.74	1.26
ICTs	4	1	6	20.00	0.45	0.57
Punctuality	22	27	8	128.00	2.88	6.33
Frequency	2	18	14	56.00	1.26	2.49
Connections	2	4	11	25.00	0.56	0.83
Comfort	2	3	12	24.00	0.54	0.78
Card 4						
Shelters	14	0	3	45.00	1.23	2.41
Access	21	12	5	92.00	2.51	5.46
Bus driver attention	1	9	2	23.00	0.63	0.99
Connections	13	21	8	89.00	2.43	5.26
Bus driving security	5	13	23	64.00	1.75	3.64
Cleanliness	2	2	9	19.00	0.52	0.73
Bus seating	5	4	11	34.00	0.93	1.70

Table 5
Final ranking of attributes according to importance.

Variable	Ranking	Points over 100
Punctuality	16.33	20.47
Frequency	9.99	12.52
Bus driving security	9.98	12.50
Access	7.36	9.22
Information-service	6.50	8.15
Connections	6.10	7.64
Bus seating	4.85	6.08
Journey time	4.60	5.76
Route	2.76	3.46
Shelters	2.71	3.40
Comfort	2.18	2.73
Cleanliness	2.03	2.54
Information-incidents	1.95	2.45
Bus driver attention	1.89	2.37
ICTs	0.57	0.71

Table 6
Cluster analysis results according to bus lines. Number of clusters in two stages.

Line	Estimation	Clusters in two stages		Total
		Stage 1	Stage 2	
712	Counting	43	164	207
	% inside the clusters (two stages)	25.4	46.7	39.8
713	Counting	26	90	116
	% inside the clusters (two stages)	15.4	25.6	22.3
714	Counting	70	21	91
	% inside the clusters (two stages)	41.4	6.0	17.5
716	Counting	30	76	106
	% inside the clusters (two stages)	17.8	21.7	20.4
Total bus lines	Counting	169	351	520
	%	100.0	100.0	100.0

service. Gender also influences perception of cleanliness, bus comfort, route and access to stops. Finally, trip purpose has an effect on perception of frequency and the presence of bus shelters along the route.

To complement the Pearson Chi-squared test and detect different behaviours between bus lines, a cluster analysis was applied to the sample (see Table 6), which revealed that the majority of the sample observations from lines 712, 713 and 716 belong to the same group (2), while those obtained from line 714 were assigned to another group (1). This is consistent with the fact that Line 714 is a special case, since it connects the interchange hub to a university campus (Universidad Autónoma de Madrid UAM), meaning this bus service is a specialised line for trips for the purpose of study, and the socioeconomic user profile differs from the rest of the lines.

To exploit the opportunity to test other techniques using this case study sample, we looked for latent SQ attributes or a group of attributes that could best explain user perception. As factor analysis has already been used for this purpose in SQ studies on urban public transportation (D'Ovidio et al., 2014) and recently also in high-speed train HST services (Alpu, 2015), this tool was applied to the sample together with a SEM methodology called MIMIC (Multiple Indicators Multiple Causes). Three of the 15 attributes included in the conventional survey – suggested by the operating company – were excluded from this analysis, as they were in a separate block in the questionnaire and were directly correlated to comfort (driving security, bus driver attention and seating comfort). The main aim of the factorial analysis was to identify the underlying variables or factors that explain the pattern of correlations within a set of observed variables. After several trials, three indicators were identified with high factor loading (>0.5), and in order to explain the variance of 56.1% with a KMO index (Kaiser-Meyer-Olkin) of 0.803, all the 11 attributes were included in the analysis, except “shelters”, due to its loading factor (lower than 0.5). The first factor identified was designated SERVICE, as it describes the quality attributes associated to the characteristics of the service operation such as punctuality, frequency, information-service and information-incidents. The second factor was called INTEGRATION as it captures concepts associated to the inclusion of the bus in the transportation systems, such as access to bus stops, connections to other modes, journey time and route. Finally the third factor, identified as SUPPLEMENTARY FEATURES, includes attributes that are usually secondary to the users such as comfort, cleanliness and ICTs.

Fig. 2 shows the path diagram with the significant parameters and relations ($p<0.1$) from the best MIMIC model estimation obtained with the following modelling fit indexes: root mean square error of approximation (RMSEA=0.079), confirmatory fit index (CFI=0.956) and adjusted goodness-of-fit statistic (AGFI=0.820). The MIMIC estimation was obtained using the AMOS program from the SPSS package (Arbuckle, 2013). This diagram shows how the relations between the observable variables and the three main factors (Service, Integration and Supplementary Features) are weaker (0.3) than the relation between factors and quality attributes. The results show more clearly how gender mainly affects the perception of attributes associated to factors of Integration and Service, while users with work purpose trips are more sensitive to Service and Integration attributes. Finally, the age of the customers conditions the perception of the attributes linked only to the Supplementary Features of the service, and usually means that the older the users, the lower the rating of comfort, cleanliness and ICTs. These results are consistent with those obtained using the frequency analysis of the attribute ratings in the conventional survey, in which the work-based trips are not as sensitive to the comfort attribute, and Supplementary features and Integration attributes are worse rated by women than by men.

The results obtained with the study of latent variables inevitably led to the issue of the disaggregation of attribute importance according to user and trip profile. Two of the most important attributes in the stated importance survey were punctuality and frequency, and both attributes are directly link to work-based trips. This is consistent with the survey's main statistics, as more than 70% of the users surveyed were workers (49.8%) or students (31.7%).

In addition to these supplementary studies, the conventional survey allowed us to derive the attribute importance and compare the results with those obtained from the conventional survey. Multiple regression analysis was used to design a model in which the dependent variable was overall service satisfaction (CSI) – whose values were collected from the last question in the conventional survey – and the dependent variables were the 15 quality attributes. The coefficients for each attribute therefore represented the average weight (or importance) given by the users. Table 7 shows the descriptive statistics of the variables from the multiple regression. Specifically it can be seen that the attributes considered most important are “frequency”, “punctuality”, “route” and “bus driving security”, three of which are also the most important attributes in the stated importance survey, providing a consistent validation of the top positions in the ranking. There is also consistency in the worse positions for the rest of the attributes (“ICTs” and “Information-incidents”) and in some intermediate positions (“journey time” and “bus seating position), although some differences can be seen in the rest of the ranking positions.

It is clear from the stated importance survey that attributes linked to supplementary features (cleanliness, comfort, ICTs, bus driving attention) are worse ranked than those associated to service (punctuality, frequency, bus driving security and information service), which are in the top positions. Attributes included in the concept of integration (connections, access, route, journey time) are in the middle of the ranking list. These results are clearly connected to the MIMIC model designed to analyse the presence of latent variables in this case study, and also consistent with the trip and user profile statistics (more than 49% work trips).

In the derived importance ranking, the main differences in the attribute position correspond to the integration category (“route”, “access”, “connections”) and two in the “supplementary features” category (“comfort” and “bus driver attention”). As “route” was the second most important attribute, we revised the survey format for differences in the wording or interpretation of the route question in the conventional survey and the stated importance survey. We found one small difference in the wording that may have affected the results: while in the stated survey the concept of route was explained in brackets (“itinerary and stops”), in the conventional survey no explanation was given of this category. This means that the interpretation of what the route actually involves (location of stops, distance between stops, adaptation to urban sprawl and

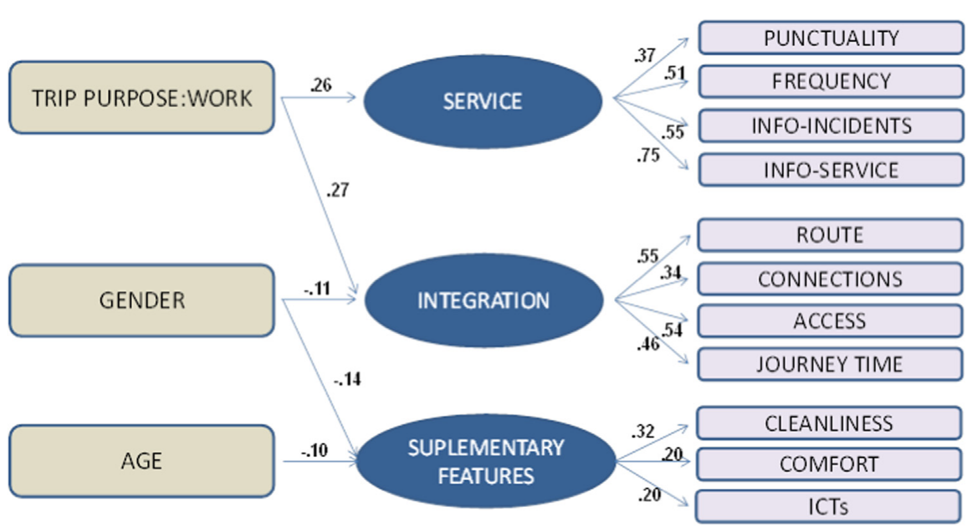

Fig. 2. Path diagram with the significant parameters and relations ($p<0.1$) of the best MIMIC model estimation (RMSEA=0.079. CFI=0.956. AGFI=0.820).

Table 7
Multiple regression model to derive importance. Descriptive statistics of the variables.

Main model indicators			
R	Squared R	Adjusted Squared-R	Standard error of the estimate
0.669	0.447	0.431	1.043

Model coefficients					
Variable	Non-standardised coefficients		Standardised coefficients	T	Sig.
	B	Error	Beta		
Constant	−0.915	0.532		−1.719	0.086
Route	0.317	0.060	0.203	5.266	0.000
Access	0.042	0.053	0.030	0.794	0.428
Punctuality	0.248	0.056	0.163	4.423	0.000
Frequency	0.254	0.053	0.190	4.822	0.000
Access	0.018	0.064	0.010	0.075	0.784
Information-incidents	0.008	0.030	0.010	0.267	0.790
Cleanliness	0.133	0.081	0.060	1.652	0.099
Information-service	0.099	0.058	0.066	1.717	0.087
Journey time	0.113	0.077	0.054	1.457	0.146
Comfort	0.209	0.072	0.090	2.593	0.009
ICTs	0.007	0.025	0.011	0.302	0.763
Shelters	0.006	0.044	0.005	0.130	0.897
Driving security	0.349	0.076	0.169	4.617	0.000
Seating	0.208	0.087	0.085	2.395	0.017
Driver attention	0.188	0.071	0.097	2.649	0.008

even journey time were open to interpretation) was not delimited. We agree that differences in wording should be avoided in any repetition of these comparative surveys in order to strengthen the validation process, but there is another argument that reinforces the consistency of the results from the stated importance survey. The study carried out in 2005 by the CRTM (Consorcio Regional de Transportes de Madrid, 2005) on bus SQ in the same city and for the same type of (periurban) corridors using focus group techniques revealed a ranking of SQ attributes quite similar to our stated importance survey results, where the "route" attribute (when used to define itinerary and stops) was not in the top positions but in intermediate ones.

We can therefore conclude that the stated preference survey clearly identifies the top and bottom positions in the attribute ranking, while the rest of the positions are more consistent (compared to the ranking derived from the attribute perception) with the MIMIC model results, and even with the rankings obtained in other SQ studies in the same study case area/context. We agree that the disaggregation of attribute importance according to user and trip profile is also recommended for the stated preference survey and would give a better understanding of the ranking results; however the number of valid surveys (293) was insufficient for a significant disaggregation, and the user and trip profile were fairly similar in both surveys. The results of this pioneer experience indicate that stated survey techniques (to estimate attribute importance in SQ studies) should not be abandoned by academics, and that transport planners and operators should also continue to use these techniques, guided by this experience and implementing this intuitive and simple survey method in their CSS design.

6. Conclusions and recommendations

This paper proposes a method to estimate attribute importance directly from a stated preference survey of customers. Attribute importance is one of the main indicators of user perception used by operating companies to estimate when an improvement in the SQ is required. We maintain that the simplicity and potential of the method are two of the strengths of our work, since this allows it to be easily applied in CSS designed and conducted by operating companies. Stated importance methods are more intuitive than derived importance methods for these companies, but require a significant increase in the length of the questionnaire, which lowers the overall response rate and the accuracy of the survey. As shown in the literature, new techniques to obtain stated importance through CSS have been practically abandoned by academics, whereas it is becoming increasingly common to develop new methods to derive importance from conventional CSS. There is also a lack of studies on comparative methodologies to obtain attribute importance using the same case study data (and even comparison between the most commonly used derived importance methodologies). This absence is also a problem for operating companies, who also need practical guidelines and recommendations to implement new approaches in their CSS.

The method proposed in this paper calculates attribute importance on the basis of both stated preference and hierarchy process theories, separating the ranking options into blocks and calculating the final score for each attribute. It reduces the length of the questionnaire, which is one of the main drawbacks of the traditional ranking technique used by companies to state attribute importance. This pioneer survey campaign (293 valid questionnaires) was validated using conventional face-to-face surveys (520 valid questionnaires). Using Madrid as a case study, the reliable survey database also offered a good opportunity to test different traditional techniques to study attribute importance (factor analysis, MIMIC models and multiple regression analysis). The validation process revealed that the top and bottom ranking positions are perfectly identified by the stated survey. Some differences between the two ranking results were observed, mainly in intermediate positions, although the derived importance results were less consistent with the MIMIC model analysis and even with SQ studies conducted in the same context (same area as the study case and same type of periurban service).This pioneer experience could be improved by avoiding any differences in wording in a repetition of these comparative surveys (to strengthen the validation process) and increasing the number of valid questionnaires in order to allow a direct disaggregation study (according to trip and user profile).

The authors of this paper acknowledge that the policy recommendations derived from this study are constrained by the number of similar experiences and the sample size used (which only focuses on four periurban bus lines), but the findings cannot be ignored by operating companies. We encourage operating companies to include (in their CSS surveys) both attribute rating questions and attribute importance questions together with the overall level of satisfaction with the service. This part of the survey can be simplified by using a reliable hierarchy card system (like the one used in our experience), thus reducing the length of the survey format. This recommendation would allow operating companies to easily obtain their own results on attribute importance, in addition to using the findings from more complex scientific academic research.

The results of this paper clearly offer transport management companies a simple and useful tool for use in their Customer Satisfaction Surveys (CSS), thereby narrowing the gap between practitioners' needs and scientific research.

Acknowledgments

The conclusions in this article are derived from a research work on service quality in public transport by the authors of this paper as members of the Department of Transport at the Madrid Polytechnic University (UPM) and the Consorcio Regional de Transportes de Madrid (Regional Transport Board – CTRM). The ALSA Group (the bus operator) made significant contributions to this research, particularly with regard to designing the customer satisfaction survey, authorising on-board survey campaigns and making the bus staff available to the research team throughout the study.

References

Alpu, O., 2015. A methodology for evaluating satisfaction with high-speed train services: a case study in Turkey. Transp. Policy 44, 151–157.

Ampt, E., Meyburg, A., 1995. Survey Methods for Transport Planning. Eucalyptus Press, United States.

Arbuckle, J.L., 2013. IBM SPSS Amos 22 User's Guide. Amos Development Corporation, Crawfordville, FL.

Aydin, N., Celik, E., Gumus, A.T., 2015. A hierarchical customer satisfaction framework for evaluating rail transit systems of Istanbul. Transp. Res. Part A: Policy Pract. 77, 61–81.

Bordagaray, M., dell'Olio, L., Ibeas, A., Cecín, P., 2014. Modelling user perception of bus transit quality considering user and service heterogeneity. Transp. A: Transp. Sci. 10 (8), 705–721.

Brady, M.K., Cronin, J.J., 2001. Some new thoughts on conceptualizing perceived service quality: a hierarchical approach. J. Mark. 65 (3), 34–49.

Celik, E., Aydin, N., Gumus, A.T., 2014. A multiattribute customer satisfaction evaluation approach for rail transit network: a real case study for Istanbul, Turkey. Transp. Policy 36, 283–293.

Chau, V.S., Kao, Y.Y., 2009. Bridge over troubled water or long and winding road? Gap-5 in airline service quality performance measures. Manag. Serv. Qual. 19 (1), 106–134.

Chou, C.C., Liu, L.J., Huang, S.F., Yih, J.M., Han, T.C., 2011. An evaluation of airline service quality using the fuzzy weighted SERVQUAL method. Appl. Soft Comput. 11, 2117–2128.

Consorcio Regional de Transportes de Madrid, 2005. Estudio Sociológico sobre Los atributos utilizados En La evaluación DE La calidad del servicio En La Red DE líneas interurbanas. EMIC-ETIC Consultans. Madrid.

D'Ovidio, F.D., Leogrande, D., Mancarella, R., Schinzano, A., Viola, D., 2014. A Multivariate Analysis of the Quality of Public Transport Services. Procedia Econ. Financ. 17, 238–247.

Del Castillo, J.M., Benitez, F.G., 2013. Determining a public transport satisfaction index from user surveys. Transp. A: Transp. Sci. 9 (8), 713–741.

Dell'Olio, L., Ibeas, A., Cecín, P., 2010. Modelling user perception of bus transit quality. Transp. Policy 17, 388–397.

Eboli, L., Mazzulla, G., 2007. Service quality attributes affecting customer satisfaction for bus transit. J. Public Transp. 10 (3), 21–34.

Eboli, L., Mazzulla, G., 2009. A new customer satisfaction index for evaluating transit service quality. J. Public Transp. 12 (3), 21–37.

European Committee for Standardization, 2003. EN 13186. Transportation. Logistics and Services. Public Passenger Transport. Service Quality Definition, Targeting and Measurement.

European Union, 1998. QUATTRO – 4th F.P. project Quality approach in tendering/contracting urban public transport operations. Final Report. Synthesis and recommendations 1998. ⟨http://www.transport-research.info/web/projects/project_details.cfm?id=636⟩, (29.07.14).

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

Guirao, B., García, A., López, M.E., Acha, C., Comendador, J., 2015. New QR survey methodologies to analyze user perception of service quality in public transport: the experience of Madrid. J. Public Transp. 18 (3), 71–88.

Hensher, D.A., Stopper, P., Bullock, P., 2003. Service quality-developing a service quality index in the provision of commercial bus contracts. Transp. Res. Part A 37, 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. Transp. Rev. 30 (2), 195–217.

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

Lehtinen, U., Lehtinen, J., 1982. Service Quality – A Study of Quality Dimensions. Service Management Institute, Helsingfors.

de Oña, J., de Oña, R., Calvo, F.J., 2012. A classification tree approach to identify key factor of transit service quality. Expert Syst. Appl. 39, 11164–11171.

de Oña, J., de Oña, R., Eboli, L., Mazzulla, G., 2016. Index numbers for monitoring transit service quality. Transp. Res. Part A 84, 18–30.

Opricovic, S., Tzeng, G.H., 2004. Compromise solution by MCDM methods:a comparative analysis of VIKOR and TOPSIS. Eur. J. Oper. Res. 156 (2), 445–455.

Parasuraman, A., Zeithaml, V.A., Berry, L.L., 1985. A conceptual model of service quality and its implications for future research. J. Mark. 49, 41–50.

Pearmain, D., Swanson, J., Kroes, E., Bradley, M., 1991. Stated Preference Techniques: A Guide to Practice. Steer Davies and Gleave and Hague Consulting Group, London.

Román, C., Martín, J.C., Espino, R., 2014. Using Stated preferences to analyze the service quality of public transport. Int. J. Sustain. Transp. 8 (1), 28–46.

Sánchez, M., Gázquez, J.C., Marín, G.M., Sánchez, R., 2007. Effects of service quality dimensions on behavioural purchase intentions: a study in public-sector transport. Manag. Serv. Qual. 17 (2), 134–151.

Saako, N., 2001. Guidelines for Stated Preference Experiment Design. Dissertation submitted for publication for the degree of Master of Business Administration. School of International Management Ecole Nationale des Ponts et Chaussées (Class 2000/01).

Saaty, R.W., 1987. The analytic hierarchy process-what it is and how it is used. Math. Model. 9 (3–5), 161–176.

Shen, W., Xiao, W., Wang, X., 2016. Passenger satisfaction evaluation model for Urban rail transit: a structural equation modeling based on partial least squares. Transp. Policy 46, 20–31.

Transportation Research Board (TRB), 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 edition.

Weinstein, A., 2000. Customer satisfaction among transit riders. How customers rank the relative importance of various service attributes. Transp. Res. Rec. 1735, 123–132.