Heterogeneity in desired bus service quality from users' and potential users' perspective

Maria Grazia Bellizzi
University of Calabria
Department of Civil Engineering
P. Bucci, Cubo 46/B
Rende, 87036, Italy
Ph. +39 984 49 4020
mariagrazia.bellizzi@unical.it

Luigi dell'Olio*
University of Cantabria
Department of Transportation and Projects and Processes Technology
Avda. de los Castros, 44
Santander, 39005, Spain
Ph. +34 942 202 262
luigi.dellolio@unican.es

Laura Eboli
University of Calabria
Department of Civil Engineering
P. Bucci, Cubo 46/B
Rende, 87036, Italy
Ph. +39 984 49 6784
laura.eboli@unical.it

Gabriella Mazzulla
University of Calabria
Department of Civil Engineering
P. Bucci, Cubo 46/B
Rende, 87036, Italy
Ph. +39 984 49 6782
gabriella.mazzulla@unical.it

*Corresponding Author

Abstract

Investigating on transit users' opinions represents a useful strategy to understand the level of quality of a service, and consequently pursuing the most convenient interventions for satisfying users and improving the service. On the other hand, an important issue for increasing the use of transit systems is linked to the necessity to attract new users, in addition to customize the current users. For this reason, it becomes fundamental to know also the perceptions of people who do not use transit systems with the aim to discover which are the reasons of their travel choices and particularly which are the most relevant transit service aspects for them. In this work, a Stated Preferences (SP) survey was conducted in order to investigate on both transit current and potential users' preferences. This kind of survey allows potential users opinions to be captured due to its peculiarity of the possibility to propose to interviewees hypothetical services and not only real services, which could not be judged by people who do not use the service. The collected data were adopted to calibrate discrete choice models such as Random Parameters mixed Logit models and Latent Class models. These models allow heterogeneity of users' opinions to be analysed. Calibrating this kind of models using data collected from both users and potential users permitted to capture the differences among users in their preferences about transit service quality, and the differences among potential users in their desired service quality. Model results showed some interesting findings concerning the differences of categories of users and potential users, and especially the diversity of preferences between people who know well the service and people who could become users even if they have not a real perception of it. We also calculated the Willingness-To-Pay (WTP), which allowed to discover that the WTP values obtained for the potential users are very much higher than the values obtained for the users. The proposed models can be also useful for the practitioners and transit operators to identify the best combination of the quality levels to assign to the various service aspects.

Key words: desired service quality; heterogeneity in preferences; users and potential users; Latent Class models; Random Parameter mixed logit models.

1. Introduction

The quality of a service can be measured starting from the customer's point of view through the well-known "customer satisfaction surveys", where the interviewed customers express their opinions about the main service characteristics, traditionally in terms of rating according to an evaluation scale. In turn, this customer subjective evaluation of service quality can be expressed in terms of perceptions, representing what customers receive from the service, and/or in terms of expectations, representing what customers expect of the service (Parasuraman et al., 1985). While perceptions are measured in terms of satisfaction ratings, expectations are not univocally interpreted. As an example, Teas (1993) stated that expectations can be interpreted as predictions of service, as an ideal standard, or as attribute importance. But, Smith (1995) found that importance ratings were rated lower than expectations. On the contrary, Landrum and Prybutok (2004) found that comparing what customers consider important with service performance is the same that comparing what customers expect with performance. Anyway, this judgement of importance or expectations can be evaluated according to an evaluation scale, as well as for the perceptions or satisfaction ratings. In the literature, it has been verified that users who are asked to give a direct rate of importance tend to give much importance to several service attributes (Eboli and Mazzulla, 2010). In addition, asking travellers to express also importance rates increases the length of the questionnaire and could undermine the accuracy of the survey (Hernandez et al., 2016). For this reason, an alternative and more convenient way to capture customer expectations, or importance, is through the Stated Preferences (SP) surveys, which allows to indirectly capture which service attributes are important to customers. As an example, in Prioni and Hensher (2000) and Eboli and Mazzulla (2008) transit users are asked to choose between two or more services, each defined by certain levels of service attributes. Some attributes or levels can be also unavailable (or hypothetical) for the users at the moment of the interview. An important advantage of this kind of experiments is that hypothetical services can be evaluated both by current and potential users, who have the possibility to express an opinion even if they do not use the service. In this way, the analyst can capture what are the opinions of the potential users, and consequently which are the service aspects influencing their decisions of using or not public transport service. This kind of analysis can conveniently conduct to the final aim of improving the services and attracting more new users. These aims are in line with the objectives of promoting more sustainable transport modes in order to solve the problems due to the excessive use of the private car in most urban areas, such as traffic congestion, parking deficiency, air and noise pollution, and so

on. In other words, future investment in transport policies can be oriented to improve the service aspects (e.g. waiting time, reliability, comfort on board, journey time, and so on) which will have positive repercussions on the overall service quality (dell'Olio et al., 2010, Bordagaray et al., 2014).

International literature is rich of studies analysing users' perceptions and expectations on public transport services (e.g. Prasad and Shekhar, 2010, Ravichandran et al., 2010, Too, 2010, Awasthi et al., 2011, Eboli and Mazzulla, 2011, Randheer et al., 2011, Monzon et al., 2013, de Oña and de Oña, 2015, de Oña et al. 2016, Guirao et al., 2016, Alonso et al., 2018, dell'Olio et al., 2018, Eboli et al., 2018).

The studies investigating on public transport service quality have been generally based on the judgements of the only users of the service, while we retain that understanding the opinions of the potential users can be the key that can conduct to a significant improvement of the service quality and a greater utilization of the service. Just in this context, to the largely investigated concept of expectations or importance, the concept of 'desired quality' can be combined, which more conveniently can be used for the potential users, who are able to express what they would desire from the service, instead what they expect. Desired quality is what users want to be fully satisfied with the service; it reflects the maximum level of service quality to which potential users aspire. In the context of the operational management of public transport, desired quality may be defined as a concise set of parameters to be measured which are of relevance for current and potential users (Mahmoud et al., 2011) or in terms of thresholds established for parameters to be considered as acceptable from the users' viewpoint (Liekendael et al., 2006). Desired quality can be conveniently measured through the SP surveys (Hensher, 1994, dell'Olio et al., 2011, Román et al., 2014), although an inappropriate survey design can introduce potential hypothetical bias (Cherchi and Hensher, 2015). Desired quality may be measured also by the Analytical Hierarchy Process (AHP), where a pair-wise comparison on several service quality attributes is made by a group of decision makers (Saaty, 1990a; 1990b; Harker and Vargas, 1990). A ratio scale of relative magnitudes is then derived from each set of comparisons, allowing to identify priorities that contribute to its effectiveness in resources allocation (Saaty, 1990a). AHP may be questionable equally to SP method, as discussed in Dyer (1990).

Starting from the observation of a lack of service quality studies considering both current users and potential users, this paper aims at investigating on the idea of quality they desire. We retain that if the investigation is oriented only to the perceived and expected quality, transit companies can know only the opinions of their customers. On the contrary, if the investigation is oriented also to the desired quality, the companies can acquire information also about the preferences of the potential users, and consequently they can develop more acceptable transit policies reducing the gap between desired and perceived or expected quality. European standards about service quality also pointed out on the importance to evaluate desired quality of current and potential passengers as the first step in the measurement of quality (CEN/TC 320, 2002, CEN/TC 320, 2006). However, there are very few authors in the literature that investigate on the preferences of both current and potential users. In order to investigate on the desired quality, we proposed to current and potential users the choice of some hypothetical services characterized by attributes such as waiting and journey time, level of occupancy, fare. We avoided more qualitative service aspects such as cleanliness or safety because they are less easily measurable in numerical terms, and they are also less easily understandable by potential users who have not a reliable perception of the service like current users. The choice of the various services allows to capture the expected quality for the current users, and the desired quality for the potential users. We decided to differentiate between expected and desired quality, because users have an idea of what they expect from the service, but potential users can only think to what they desire more than expect, because they do not know the service well. Our potential users are people travelling nearly always by private car but having an idea of the public transport services because they used it recently. In other words, we proposed to users and potential users the same hypothetical services, but for the first category we can obtain an expected level of service quality, while for the second category we have to talk about desired level of service quality. The desired quality tends to measure how the potential users change their preferences in relation to the proposal of scenarios which not necessarily exists but built considering realistic attributes. In this way, we want to measure their wishes using attributes that are much realistic as possible.

In line with the expressed concepts, the contribution of this work is twofold. From a methodological perspective, we provide a framework to analyse transit service quality by considering not only the opinion of the current users but also the preferences of the potential ones. From an empirical perspective, we expose and analyse the behavioural choices of current users of a bus service operating in the city of Santander (Spain), and also of potential users who provide precious information useful to capture new customers.

To do it discrete choice models have been estimated to study the preference heterogeneity of different users' categories, and the Willingness to Pay (WTP) estimated to investigate the influence of each variable on desired quality. Differently from dell'Olio et al. (2011), where the desired quality concept was introduced and Multinomial Logit models were estimated to determine the weight of certain attributes, in this paper we present Latent Class and Random Parameter Mixed logit models in order to study the differences in public transport perceptions of users and potential users. These models permit to find the latent classes of users and potential users and to study the variation of the perceptions in the population. In addition, the WTPs of users and potential users associated with the variation of different variables are determined in order to compare them. The obtained models can be applied also by the practitioners to calculate the utility of different services with the aim to identify the best combination of the levels to assign to the various service aspects. In the rest of the paper, we propose a theoretical background where we introduce previous theories at the basis of our methodology; therefore, the methodological framework to analyse transit service quality is described. In the following, we introduce experimental sections concerning the SP survey, the sample characteristics, and the model results. Finally, a discussion section and the conclusions section about the work are reported. In the conclusion section the general findings, policy proposals and practical implication are described.

2. Theoretical background

In the last decades, SP surveys have been extensively used in different fields, such as marketing, transport, health economics, agricultural and environmental economics. This class of methods differs from traditional economic approaches, which are based on Revealed Preference (RP) data obtained by observing individual behaviour in real market (Louviere et al., 2010). Despite the lack of realism generally recognised in SP surveys, a very rich literature exists showing the ability of SP methods to elicit individual's preferences for "alternatives" expressed in a choice context. The issue of realism can be found also in RP surveys, where a number of measurement errors can occur. This well-known issue has been carefully accounted for in several studies and discussed in a deep manner in Cherchi and Hensher (2015). In SP survey, individuals can have limitations in their capacity to process information, and are not always willing to spend an adequate effort in evaluating alternatives; therefore, in order to ensure realism and reduce potential hypothetical bias, analysts may need to build rather complex survey tasks which respondents are asked to process in a short time. Further problems can be found when contacting respondents are not the right persons to interview, the context is not precisely defined, or attributes are not measured correctly. One of most popular way to elicit SP from individuals is the discrete choice experiment method. In a discrete choice experiment, the alternatives are described as sets of different attributes varying on different levels. Specifically, different choice sets, consisting in several alternatives defined by different levels of the same attributes, are presented to the respondents; for each choice set, the individual must select the most preferred option or choose none of those proposed (Louviere et al., 2000).

Discrete choice experiments are based on the well-known theory of choice behaviour proposed by Thurstone in 1927, and called Random Utility Theory (RUT). The basic hypothesis underlying RUT is that there is a latent construct for each choice alternative called "utility", which cannot be observed by researchers. The latent utilities can be summarized by two components: (1) a systematic explainable component; (2) a random unexplainable component. Systematic components comprise attributes explaining differences in choice alternatives and covariates explaining differences in individuals' choices. Random components comprise all unidentified factors affecting choices, and can also include factors reflecting the differences in choices associated with individuals (McFadden, 1974, Louviere et al., 2010).

By making different assumptions on the random components probability distributions, different discrete choice models can be derived, ranging from the first MNL model proposed by McFadden in 1974 to a large amount of advanced discrete choice models where the basic assumption of random components independently and identically distributed is relaxed. The basic assumption greatly facilitates estimation and forecasting, providing a convenient closed form for the underlying choice probabilities without any requirement of multivariate integration (Hausman and McFadden, 1984). At the same time, it is very restrictive because it implies homogeneity among individuals; as a consequence, MNL model allows to take into account only variations of the preferences that change in a systematic way with respect to the observable variables, whereas it is not able to reflects variations in preferences due to unobserved variables.

Mixed Logit (ML) models have been instead proposed in order to consider the heterogeneity among individuals permitting the differences in individuals' perceptions and responses to be considered. ML models can be formulated according to two different structures: "error component structure" and "random coefficients structure" (reported in the literature also as Random Parameter, RP model). In the first structure, some hypotheses of correlation between alternatives are made; in the second one, some hypotheses of unobserved heterogeneity among individuals regarding observed variables are made (Bhat, 2003). RP model has the standard form of an MNL model except that one or more parameters can be considered as random parameters, with the standard deviation estimated together with the mean.

As an alternative specification, the Latent Class (LC) model provides an approach for accommodating individual heterogeneity by considering the population composed by a finite number of group of individuals (Hensher et al., 2015). The classes are different from each other, and the individuals of a same group have a common vector of parameters characterizing the class. The term "latent" is due to the model's unknown way of distributing individuals among the classes. The theory behind LC model states that individual behaviour depends on observable attributes but also on latent heterogeneity varying with unobserved factors.

3. Methodological framework

In this paper, we provide a framework to analyse transit service quality by considering not only the opinion of the current users but also the preferences of the potential ones. The first step of the methodological framework focuses on the conjoint use of SP survey method and discrete choice experiments with the aim to capture the individual preferences of both users and potential users expressed in choice contexts describing real or hypothetical bus services. Discrete choice experiments are generated starting from a certain number of scenarios, each consisting in two alternatives described as sets of different attributes characterizing the services varying on different levels. SP experiments were careful designed and implemented with the aim to propose to respondents such hypothetical scenarios as realistic as possible. We chose the levels by taking into account the performance of the current service. For this reason, we retain that both users and potential users had not confusion about the proposed experiments. When regular users are interviewed, they can express their preferences about the service and also a judgement about the expected quality level. However, when potential users are involved in the survey, they can express only an opinion on the desirable quality level about a service never experimented, or experimented few times.

The design of the SP survey has followed several stages: 1) design of a preliminary survey after an extensive state-of-the-art bibliography review; 2) organisation of a focus group to define the variables that should be part of the final survey; 3) design of a pilot survey to check the goodness of the design, as well as to fix the final sample size; 4) design of the definitive survey based on the necessary modifications of the pilot survey and on the methodology proposed by Rose et al. (2008).

We supposed service quality attributes having parameters with a normal probability distribution revealing preference heterogeneity around the mean of each random parameter. This allows us to consider not only the systematic variations, but also the random variations in the tastes of the users. The presence of these random parameters permits us to account for the panel effects because to each respondent a sequence of choice sets is proposed, and making a choice of an alternative from each choice set is requested. These data, often called as "instantaneous panel", engender correlation among observations common to a respondent (Hensher et al. 2015). We do introduce error component which leads to a utility heteroskedastic interpretation and which makes possible to account for flexible substitution patterns across alternatives. Random Parameter ML models are introduced also for verifying the interactions occurring among individual socio-economic characteristics and service quality attributes included in the SP survey.

Thirdly, the methodological framework provides for LC models calibration, which can help the analyst in defining classes of individuals with common socio-economic characteristics and/or similar behavioural choice. Latent classes of both users and potential users are identified, and heterogeneity among the different classes are highlighted and analysed. The number of classes to be used is chosen according to estimated criteria of goodness of fit, like Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC), as reported in Hensher et al. (2015).

4. SP survey

The data were collected by a survey conducted in the port town of Santander, capital of Cantabria, located on the north coast of Spain. The town relies more than 180,000 habitants distributed unevenly. The highest density of population is concentrated in the downtown. In Santander, 47.5% of journeys are made on foot, 42.1% by car, 9% by bus and 0.5% by bike and 0.9% by other modes. This modal split was determined by a study carried out in 2015 within the PARK-INFO research project (TRA2013-48116-R) funded by the Ministry of Economy and Competitiveness of the Government of Spain. The public transport of this city is composed by 18 bus lines, operating daily, and by 3 night lines that operate on Fridays and Saturdays. This set of lines covers almost the entire territory of the city, and the entire population of Santander has a bus stop available at a maximum of 300 meters. Every year the urban transport service of Santander relies 20 million of trips.

The questionnaire is made up of three sections. The first section aims to draw the socio-economic profile of users and potential users (gender, age, income level, driving licence, owing a car). The second section has the objective to collect information about the trip (trip origin/destination, frequency of the journey, trip purpose, access and egress time, ticket payment mode). In this section, a question about the perception of the public transport overall service quality was added to both users and potential users. Specifically, respondents were asked to express a rate according to a 5-point scale from "very bad" to "very good". Finally, in the third section of the survey, the SP scenarios were presented.

The survey was addressed to users and potential users of the public transport service of the city. Even if in both cases the survey was the same, the objectives are different. Specifically, the survey addressed to the users seeks to know what they expect from the public transport service, according to the experience they have of it. Instead, the survey carried out to potential users seeks to know what they desire from the public transport service, even if they never experienced it. Through the SP survey, it is possible to capture which are the most important variables for both users and potential users. A focus group was organized in order to obtain information about participants' points-of-view, expectations and knowledge on public transport in the city of Santander by following a workshop-style methodology. The participants to focus group were selected to represent the users of public transport in Santander as closely as possible. Specifically, ten participants spreading from four young people, three adults and three mature adults were selected. The focus group allowed us to define the variables used in the pilot SP survey (Powe et al., 2005).

An unlabelled survey was designed. Each questionnaire consists of eight scenarios where interviewees had to choose between two hypothetical but realistic alternatives, defined by five variables, namely: waiting time, time on board, access time, level of occupancy, and fare (Table 1). These variables were chosen among the relevant aspects emerged from the focus group, by selecting the aspects that were resulted the most understandable during the pilot survey, even by people who have never experienced the transport service. Moreover, the data collected through the pilot survey formed the basis for the efficient design of the definitive survey, carried out using the D-error (Rose et al., 2008). The efficient design allows to estimate parameters with minimum standard error by calculating the variance–covariance matrix on the basis of previous experimentation.

Each scenario was transformed into cards, as shown in Figure 1, and the chosen ranges of variation of the variables are shown in Table 1.

For the variable "Waiting Time", the average value of the line chosen by the user was considered as a reference value, and this value was estimated as half of the average interval considering that the arrivals are distributed uniformly over time. For the potential users, fixed levels were set, in line with the most recurrent waiting time values of Santander's public transport lines. Concerning the variable "Journey Time", a similar criterion was used, making the average value of the journey time for the declared route vary as indicated in table 1. For potential users, average time on board is calculated by estimating the average journey time by bus for the same trip on foot, or by car, or another mode that the user is carrying out when intercepted for the survey. Regarding the variable "Access Time", the average value for the line chosen by the user was considered as reference value. For the potential users, fixed levels were set, in line with the most recurrent access time values of Santander's public transport lines. For the variable "Level of Occupancy", two levels were used: average occupation (50%) and high occupation (80%). Finally, four levels of variation were used for the variable "Fare", using the same values for users and potential users. More specifically, the levels were defined by considering that the fare during the period of the survey was 1.30€/trip without smart card and 0.66€/trip using the smart card.

Table 1 – Levels of variation of the variables.

Variable	Levels of variation for Users	Levels of variation for Potential Users

Waiting Time	1	current average waiting time of the line	2.5 minutes
	2	+100 % of the current average waiting time of the line	5 minutes
	3	+ 50 % of the current average waiting time of the line	7.5 minutes
	4	− 50 % of the current average waiting time of the line	10 minutes
Journey Time	1	current average journey time of the line	declared average time on board
	2	+100 % of the current average journey time of the line	+100 % of the declared average journey time
	3	+ 50 % of the current average journey time of the line	+ 50 % of declared average journey time
	4	− 50 % of the current average journey time of the line	− 50 % declared average time journey
Access Time	1	+ 30 % of the current average access time of the line	4 minutes
	2	− 30 % of the current average access time of the line	7 minutes
Level of Occupancy	1	50 % of the bus capacity, corresponding to the medium level of occupancy	50 % of the bus capacity, corresponding to the medium level of occupancy
	2	80 % of the bus capacity, corresponding to the high level of occupancy	80 % of the bus capacity, corresponding to the high level of occupancy
Fare	1	0.50 €	0.50 €
	2	1.00 €	1.00 €
	3	1.20 €	1.20 €
	4	1.50 €	1.50 €

The survey was asked to both users and potential users at rush hour on working days, and during the survey, to give to the interviewees a better understanding of the variable related to the level of occupancy, support material (e.g. photographs) was shown to them (dell'Olio et al., 2011). The interviews addressed to the users have been carried out at bus stops as well as on board of all the public transport lines of the city. Instead, the interviews addressed to the potential users have been conducted in the street, while they are parking their private vehicle or at the time they were traveling on foot (the two most significant modes of transport in the city, as previously mentioned) or in any other mode of transport different from the bus. Once their attention has been captured, they are asked if they are currently bus users or if they consider the bus as a possible alternative for their journeys. If the answer is negative, the SP survey continues. If the answer is positive, they are thanked for their participation and the survey is interrupted. The same time slot has been used as for user surveys on public transport, that is, from 12:30 to 14:30, which corresponds to rush hours. The surveys were carried out in the month of October 2015. The average time each interviewee spent for answering to the questionnaire has been around 15 minutes. 810 users and 158 potential users were interviewed. The low number of contacted potential users is due to the difficulty to find people who habitually travel with private transport modes, but occasionally also by public transport modes. Since each interviewee evaluated 8 scenarios, 6,480 observations for users and 1,264 for potential users were collected.

SCENARIO 1			
Waiting Time		5 minutes	8 minutes
Journey Time		17 minutes	9 minutes
Access Time		7 minutes	4 minutes
Level of Occupancy		Medium	High
Fare		1.2 €	1.5 €
LINE L	SCENARIO 1	CARD 1	CARD 2

Figure 1 – Example of experiment proposed to the interviewees.

5. Sample characteristics

Before the SP experiment, the questionnaire contains a part aiming to collect people socio-economic characteristics and information about the development of their trip. Table 2 shows the socio-economic characteristics of both public transport users and potential users.
In both the samples female interviewees are predominant, and in the group of users there are twice as many women compared to men. As regard age composition, the highest rates refer to people younger than 24: more than 30 % for users, and more than 50 % for potential users. The rest of users is evenly distributed between the age ranges from 25 to 54 years old (rates close to 20 %), and over 55 years old (rates close to 10 %). The income level fully characterizes the two groups: the major part of users has a medium household income; on the contrary, most of potential users are people with a high income. According to the frequency of use of the bus, the sample of users is divided in: very frequent users (daily use of the bus), frequent users (no more than 5 times in a week), sporadic or circumstantial users (use of public transport under certain conditions). The percentage of users who use the bus more than 5 times per week is much higher than the rest of the

categories. For the group of potential users, the frequency of journey is referred to the frequency they made the trip they had just finished or were making at the time they were asked the survey. In this case, there is no a great sub-category predominance. The driving licence is not determinant in the composition of the sample of users. On the other side, more than the 80 % of potential users have a driving licence. The results regarding the owning of a car are interesting because it seems that the public transport system of the city is not competitive in relation to the private transport. In fact, only a little more than 50 % of potential users own the private car. Concerning the trip purpose, users move to go at work or to go home. Here we must keep in mind that the survey has been conducted around the peak time of public transport in Santander, and most of the trips are made by students to go home after school. Regarding the sample of potential users, approximately 60 % of trips are made to go to work or study. These trips are subject to very specific schedules and, probably for this reason, conduct people to prefer the private car. The potential users intercepted in this survey are distributed by mode as follow: on foot (28.9%), by bike (8.2%), by car as passenger (19.5%), by car as driver (36.5%), by taxi (1.9%), and by motorbike (5%).

Table 2 – Socio-economic characterisation of the sample.

Category	Sub-category	Users (%)	Potential Users (%)
Gender	Male	33.09	44.30
	Female	66.91	55.70
Age	Younger than 24	32.84	56.33
	Between 25 and 34 years old	16.54	12.66
	Between 35 and 44 years old	17.53	7.59
	Between 45 and 54 years old	16.17	10.76
	Between 55 and 64 years old	9.26	6.96
	Older than 65	7.65	4.43
	-	0.00	1.27
Income level	< 900 €	11.73	8.23
	900 – 1500 €	13.46	17.72
	1500 – 2500 €	66.91	13.29
	> 2500 €	7.90	60.76
Frequency of journey	Very Frequent	56.67	37.97
	Frequent	25.06	27.22
	Sporadic or circumstantial	18.27	34.81
Driving licence	Yes	53.33	80.38
	No	46.67	19.62
Owning of a car	Yes	33.58	55.70
	No	66.42	44.30
Trip purpose	Home	29.14	13.92
	Work	29.75	41.77
	Study	19.38	17.72
	Other	21.73	26.58

The socio-economic characterisation ends with a question on the evaluation of the service (Table 3). The results show how the major part of users expressed high ratings for the public transport service, with an average score of 3.59 out of 5. On the other side, the potential users are more critical about the offered service, even if almost 20% of the interviewees expressed a high rate (4 or 5). This last result could be not very reliable because they have not enough knowledge of the service to be able to evaluate it. On the other hand, users could have expressed higher rates because of their habits to use the service.

Table 3 – Overall evaluation of the service.

	% Distribution by evaluation of the service						
	1	2	3	4	5	Non-response	Average score
Users	2.47	8.77	26.91	36.42	14.94	10.49	3.59
Potential Users	4.43	23.42	35.44	13.29	4.43	18.99	2.88

6. Modelling results

Random Parameter ML and LC models were estimated separately for the users' and potential users' sample. All the models were calibrated by using the software N-Logit (Greene, 2016). The definition of the variables included in the models is reported in Table 4, together with their abbreviation, units or values.

Table 4 – Definition of the variables used in modelling.

Variable	Abbreviation	Definition	Unit/Values
Gender	Sex	Female	1=yes, 0=no
Age	A24	Younger than 24	1=yes, 0=no
	A2534	Between 25 and 34 years old	1=yes, 0=no
	A3544	Between 35 and 44 years old	1=yes, 0=no
	A4554	Between 45 and 54 years old	1=yes, 0=no
	A5564	Between 55 and 64 years old	1=yes, 0=no
	A65	Older than 65	1=yes, 0=no
	YA	Younger than 34	1=yes, 0=no
	MA	Between 35 and 54 years old	1=yes, 0=no
	AA	Older than 55	1=yes, 0=no
Income level	I900	< 900 €	1=yes, 0=no
	I9015	900 – 1500 €	1=yes, 0=no
	I1500	< 1500 €	1=yes, 0=no
	I1525	1500 – 2500 €	1=yes, 0=no
	I2500	> 2500 €	1=yes, 0=no
Frequency of journey	VFrequent	Very Frequent	1=yes, 0=no
	Frequent	Frequent	1=yes, 0=no
	Sporadic	Sporadic or circumstantial	1=yes, 0=no
Driving licence	Licence	Owning a driving licence	1=yes, 0=no
Owning of a car	Car	Owning a car	1=yes, 0=no
Trip purpose	Home	Home	1=yes, 0=no
	Work	Work	1=yes, 0=no
	Study	Study	1=yes, 0=no
	Other	Other	1=yes, 0=no
Waiting Time	T_W	Average waiting time	minutes
Journey Time	T_J	Average journey time	minutes
Access Time	T_A	Average access time	minutes
Level of Occupancy	LO	Medium level of bus occupancy	%
Fare	FARE	One-way ticket fare	€

In Table 5 and 6, results of the ML and the LC models are reported for users and potential users respectively.

Table 5 – Results of the ML and the LC models referred to users' sample.

| | ML Model | | LC Model | | | | | |
| | | | Class 1 | | Class 2 | | Class 3 | |
	Param.	z	Coeff.	z	Coeff.	z	Coeff.	z
T_W	-0.113 (***)	-11.93	-0.235 (***)	-6.09	-0.076 (***)	-9.18	-0.121 (***)	-3.47
sigma(T_W)	0.079 (***)	7.35						
T_J	-0.147 (***)	-13.96	-0.234 (***)	-7.52	-0.059 (***)	-8.59	-0.089 (***)	-3.76
sigma(T_J)	0.099 (***)	14.76						
T_A	-0.100 (***)	-8.04	-0.178 (***)	-3.80	-0.068 (***)	-5.03	-0.149 (***)	-2.38
LO	-0.028 (***)	-14.53	-0.012 (***)	-2.99	-0.029 (***)	-14.12	-0.022 (***)	-2.79
sigma(LO)	0.025 (***)	10.59						
FARE	-2.231 (***)	-11.83	-1.801 (***)	-6.13	-0.943 (***)	-8.39	-8.462 (***)	-4.09
sigma(FARE)	1.745 (***)	5.44						
T_W*I1500	-0.040 (**)	-2.34						
T_W*I2500	-0.064 (**)	-2.30						
T_J*Sex	0.023 (**)	2.18						
T_J*A65	0.049 (***)	2.67						
T_J*I1500	-0.042 (***)	-3.21						
T_J*I2500	-0.061 (***)	-3.23						
LO*Study	-0.009 (**)	-2.19						
FARE*Sex	0.498 (**)	2.46						
FARE*I1500	-0.826 (***)	-3.64						
Class Probabilities			0.286 (***)	4.96	0.590 (***)	9.23	0.125 (***)	6.70

N. of observations	6480			6480		
Degree of freedom	24			17		
Log Likelihood	-3447.920			-3496.776		
LL Constant Only	-4491.594			-4485.013		

(***) Significance at 1% level, (**) at 5%, (*) at 10%

Random Parameter ML models have been estimated to consider the differences that there may be in the preferences of interviewees. These differences are considered through random variables, defined by the value of their mean and the value of their standard deviation, which will depend on the distribution function chosen for them. In the proposed models, the distribution function best representing the random variables is the Normal function.

Table 6 – Results of the ML and the LC models referred to potential users' sample.

	ML Model		LC Model			
			Class 1		Class 2	
	Param.	z	Coeff.	z	Coeff.	z
T_W	-0.051	-1.29				
sigma(T_W)	0.173 (***)	4.26				
T_J	-0.135 (***)	-4.27	-0.480 (***)	-7.34	-0.036 (***)	-3.08
sigma(T_J)	0.189 (***)	6.60				
T_A	-0.055	-1.37				
LO	-0.045 (***)	-7.27	-0.099 (***)	-5.55	-0.023 (***)	-5.92
sigma(LO)	0.046 (**)	2.28				
FARE	-4.636 (***)	-6.44	-0.620 (*)	-1.85	-2.146 (***)	-9.67
sigma(FARE)	3.556 (***)	4.69				
T_W*Car	-0.165 (***)	-2.83				
T_J*I9015	-0.098 (*)	-1.87				
T_J*A4554	-0.155 (*)	-1.83				
T_J*Car	-0.110 (***)	-2.80				
FARE*I9015	-2.195 (***)	-2.10				
FARE*Car	2.073 (***)	2.79				
Class Probabilities			0.470 (***)	9.82	0.530 (***)	11.08
N. of observations	1264			1264		
Degree of freedom	21			7		
Log Likelihood	-578.103			-645.620		
LL Constant Only	-876.138			-870.046		

(***) Significance at 1% level, (**) at 5%, (*) at 10%

In these models, it was taken into consideration also the possible heterogeneity around the mean of the random variables due to the socio-economic characteristics, like sex, income and age. Before obtaining the presented ML models, several specifications were investigated including the quadratic effect of the fare and the ratio of fare to income. As reported also in Jara-Diaz and Videla (1989), these are the most common ways of considering the income effect in the modelling. According to the authors, three properties have to be met: (i) the coefficient associated with the variable related to fare has to be negative and decrease with income; (ii) the coefficient associated with the quadratic effect of fare has to be positive and decrease with income; (iii) the marginal variation of the systematic utility with respect to income has to decrease with income.
In our case, for potential users the parameter associated to the quadratic effect of fare was positive and the one related to fare negative, but the quadratic effect was not statistically significant; this implies that the presence of income effects is not detected for potential users. On the other hand, for users the parameter associated to the quadratic effect of fare was negative and the one associated to fare positive; however, the quadratic effect was significant, whereas the direct effect was non-significant. We investigated the conjoint effect of fare and income by introducing in the model specifications different kinds of interaction between the two variables. Definitively, the variables and their interactions introduced in the presented models give the best results, whereas the specifications including the ratio of fare to income turned out to be less explanatory. As can be seen in tables 5 and 6, the interactions of the income levels with other relevant variables are included too, allowing us to take into account the effect of income on the perceptions of users and potential users.
LC models have been introduced to know how the individuals of each sample are implicitly sorted into a set of classes. Three latent classes and two latent classes were selected for the users and potential users, respectively. The number of classes was chosen after some modelling carried out with different number of classes. The LC models are those with the best fit. The variables T_W and T_A were not statistically significant in the LC model for potential users; therefore, we omitted them from the model. We consider that it makes

no sense to estimate the WTP of variables that are not statistically significant, because including them could bias the calculation of the other WTPs. As we have multiple observations per individual, we accounted for the panel effects both for RP and LC models.

Evidently, in such a study heterogeneity of the individuals cannot be ignored.

For comparing the models to each other, using the likelihood ratio test is not appropriate because mixed logit models and latent class models are not nested (Greene and Hensher, 2003). On the contrary, interesting results emerged from the average values of WTP (Table 7 and Table 8). As is well known, the WTP measurements is a manner to evaluate in terms of monetary cost the willingness for saving time or for improving the offered service (Ortúzar and Willumsen, 2011). Specifically, for the calculation of the WTPs from the ML and LC models, we have used the sample enumeration method as in the following: a) we have generated the individual parameters for the variables of interest; b) we have proceeded to the calculation of the WTP including the corrections due to the interactions of the variables of interest with the socio-economic variables considered; c) having the WTPs of each individual, we have applied the corresponding filters in order to obtain the average values for the categories reported in tables 7 and 8.

The final WTP values for ML models were obtained by excluding 99 observations from users' sample, and 66 observations from potential users' one, in order to avoid negative WTPs or very high extreme values (outliers).

The values of WTP are expressed in euro (€): those regarding times represent the WTP to save an hour; those regarding the level of occupancy represent the WTP for reducing the congestion on board by 50%. Table 7 shows the average values of WTP obtained from the ML and LC models. In Table 8 the average values of WTP derive from the ML models and were calculated considering the heterogeneity of the individuals.

Table 7 – Average values of WTP from ML and LC models.

| | Users | | | | Potential Users | | |
| | | LC | | | | LC | |
	ML	Class 1	Class 2	Class 3	ML	Class 1	Class 2
T_W	7.17	7.84	4.84	0.86	15.84		
T_J	6.97	7.80	3.77	0.63	20.59	46.44	1.00
T_A	4.90	5.94	4.30	1.06	3.36		
LO	1.37	0.50	1.50	0.15	1.68	8.00	0.50

Table 8 – Average values of WTP from ML models, considering the heterogeneity.

| | | Users | | | | Potential Users | | | |
		T_W	T_J	T_A	LO	T_W	T_J	T_A	LO
Sex	Male	5.63	6.02	3.81	1.05	17.90	23.70	3.91	3.65
	Female	8.05	7.52	5.53	1.55	14.40	18.39	2.98	2.75
Age	Younger than 35	5.63	5.92	4.05	1.20	17.40	22.81	3.84	3.85
	Between 35 and 55 years old	9.29	9.04	6.31	1.70	17.14	23.21	3.12	1.80
	Older than 55	6.66	4.06	3.70	0.95	4.90	3.44	0.88	0.50
Income level	< 1500€	4.92	4.27	2.46	0.65	8.34	9.41	1.24	0.65
	> 1500€	8.05	8.03	5.85	1.65	20.04	26.84	4.55	4.50
Frequency of journey	Very frequent	7.19	7.00	4.67	1.30	25.09	38.25	6.05	6.40
	Frequent	7.02	6.97	5.16	1.50	5.26	5.39	1.07	0.70
	Sporadic or circumstantial	7.28	6.92	5.23	1.45	12.11	10.60	1.85	1.00
Driving licence	Yes	8.37	8.29	5.62	1.55	20.36	26.41	4.24	4.00
	No	5.80	5.48	4.08	1.20	1.46	2.07	0.58	0.25
Owning of a car	Yes	9.61	9.33	6.33	1.70	26.97	35.98	5.43	5.50
	No	5.93	5.79	4.17	1.20	3.16	3.05	1.02	0.35
Trip purpose	Home	7.67	7.43	5.28	1.35	3.21	3.99	0.66	0.40
	Work	6.88	6.67	4.41	1.10	24.30	32.42	5.38	5.55
	Study	6.32	6.47	4.49	1.75	14.60	19.31	2.71	1.60
	Others	7.63	7.23	5.43	1.50	5.49	5.31	1.08	0.60

7. Discussion

From the model results, interesting discussions emerge, especially from the interactions of the five service quality attributes adopted in the SP experiment with the socio-economic users' and potential users' characteristics.

7.1 Mixed Logit Models for users' and potential users'

By observing the results of the ML model in Table 5, we can make some considerations concerning the users of the service. Characteristics such as gender and age represent elements that differentiate bus users and their preferences about service quality. More specifically, time on board seems to be less important for women than for men, as well as fare, maybe because they are more sensible to other service aspects such as comfort or safety, and so on (Allen et al., 2018). On the other hand, users over 65 give less importance to time on board than others. Also, income is a characteristic that differentiates the sample preferences. Particularly, users having high income seem to give more importance to the journey times and waiting time, and users having a low income, give much importance to fare, as expected. Also travel habits represent elements that differentiate users. Specifically, users travelling for study seem to give more importance to the level of occupation.
By comparing findings obtained for users with the results reported in dell'Olio et al. (2011), we can observe that the signs of the variables are in the same direction as well as the sign of some variables like sex, interacting with journey time.
By observing the results of ML model in Table 6, we can make other interesting considerations concerning the potential users of the service. More specifically, we found that potential users aging between 45 and 54 years give more importance to time on board. Finally, potential users owning a car give more importance to time on board, waiting time and less importance to fare, as regards potential users without a private car. As can be seen for potential users, both the waiting time and the access time are statistically insignificant and have a negative sign. This probably happens because the potential users, who do not use public transport or are not very familiar with it, do not perceive this variable. For potential users in dell'Olio et al. (2011) no interaction is considered but the signs of the corresponding variables are in the same direction.
In addition, from the analysis of the ML models we can observe that, if we look at the interactions, in the case of users, the variation of the parameters in most of the interactions is minimal, although statistically significant. However, concerning waiting time, in the case of potential users, significant variations in interactions are observable with the variable representing car owning, while concerning journey time, there are significant variations with income level, age, and car owning. We can observe that the parameters reach the same order of magnitude of the parameters without interactions. This means that the parameters on average can be considered similar, but on an individual level the differences in perception can produce values of the waiting time parameters that can double the journey time ones. If we consider also that these parameters follow a normal probability distribution, their variation at the individual level may become greater.
Finally, from Table 5 and 6 it can be observed that ML models considers random variations in tastes that are statistically significant. This permits to consider the heterogeneity among the tastes of the users, and to explain this heterogeneity where there is a significant interaction with some socio-economic variables.

7.2 Latent Class Models for users' and potential users'

Interesting findings emerges also from the LC models of both the groups (users and potential users). Concerning bus users, it seems that there is a latent class of users who give more importance to the journey times (including almost 30% of users), another class that gives more importance to comfort (to which almost 60% of users belong), and a third class most directed to the fare (including little more than 10% of users). Regarding potential users, LC model identifies two latent classes: the first group would seem to give more importance to time, and the second to fare. The results show that there is a similar class belonging probability among the two classes (47% and 53%, respectively).
Concerning the users, from Table 5 it can be seen how LC model has a Log-Likelihood slightly more negative than ML model. This means that the three latent classes provide the same information explained by ML model considering systematic and random variations in tastes. Concerning the potential users (Table 6), the difference between the Log-Likelihoods is not so small. Definitively, ML model can be considered as the best model in both cases.

Furthermore, if we observe the standard deviations of the random parameters of both models (Table 5 and 6) we realize that these values are very big compared with the average values of the parameters and this effect is greater for the potential users. This means that the perceptions of the users vary significantly in the population. Something similar happens with the LC, whose classes allow differentiating users more.

7.3 Willingness to pay analysis

More interesting considerations can be made by observing the WTP values reported in Tables 7 and 8. Waiting time and time on board are the most relevant service aspects both for users and potential users, even if potential users would be willing to pay about twice of users to save an hour of waiting time or threefold for one hour of time on board (table 7). Much less importance is given to comfort considering that both users and potential users would be willing to pay a few odd coins to improve the comfort level of 50%.
The WTP values obtained from the LC models confirm the above made consideration that there is a class of users who give much importance to journey times, a class that gives a discrete importance to comfort as regards the other two classes, and a third class of users for whom fare is the most relevant aspects, in fact they would be willing to pay very little for saving in terms of time. Also for potential users the WTP values confirm the considerations emerged from the analysis of the model coefficients.
From Table 8, we can observe that female users are willing to pay more than male users for improving all the considered service aspects. From the analysis of the model coefficients seemed that for women both journey times and fare are less important than for men; considering also that women give less relevance to fare than journey times with respect to men, we can conclude that women are willing to pay more than man to reduce journey time. The medium age class of users would be willing to pay more than the other age classes. Also in this case, WTP results are more informative than model coefficients. On the other hand, the WTP values concerning the differences in terms of income confirm the observation emerged from the analysis of the model coefficient, that is users having low incomes would be willing to pay less than the other users.
Concerning travel purposes, WTP values suggest evidences that were not shown from the analysis of the model coefficients. As an example, it emerges that the highest values of WTP concern the trips made for backing at home, while the lowest values regards the trips for going to work or study purpose. This is a singular finding, which suggests that users would be willing to pay more to return in brief times at home rather than to reach the work place. On the other hand, it is confirmed that users travelling for study give more importance to comfort as regards users travelling for other purposes. Another final interesting result is that very frequent users would be willing to pay more than the other users for saving in terms of journey times, maybe because they know very well the service and perceive better the delay of the bus with respect the scheduled time or the time on board. The WTP regarding comfort shows evidences enough expectable. These are the categories of people more willing to pay for improving comfort: women, users from 35 to 55 years, users with medium and high income, and frequent users, in addition to users travelling for studying and owing a car.
Very different and much less expectable are the WTP values for the potential users. First of all, almost all the values of WTP of potential users are much higher than the values obtained for users; in some case, they are five times the values obtained for users. The differences are less relevant for the access time and comfort, as we can see in Tables 7 and 8; particularly access time is a fixed coefficient, and this fact justifies this evidence. On the contrary, for waiting time and time on board the differences are very important, especially for some classes of users. From Table 8, concerning the gender, we can observe the opposite evidence as regards WTP of users; in fact, in the case of potential users, males are willing to pay more than females for improving all the considered service aspects. While, also for potential users the younger and medium age people would be willing to pay more than the older people. Also for this sample, people having low incomes would be willing to pay less than the others, but the differences are very relevant (more than 3 times less). Concerning travel purposes, the results are very different from the WTP values obtained for the sample of users, in fact, it emerges that the highest values of WTP concern the trips for going to work or study, which is a less strange result as regards the findings discovered for the users. While the characteristics linked to the owning of driving licence and car did not show relevant findings concerning users, in the case of potential users we can observe that WTP values of people who have driving licence and car are much higher than people who have not them, maybe because this last categories of people have not a real perception of the trip characteristics.
Also Sillano and Ortuzar (2005) and Hensher (2001a; 2001b; 2001c) found very high mean WTP values for heteroscedastic and autoregressive specifications, which could indicate that mixed-logit models tend to overestimate WTP values. For potential users we obtain in general very high values. However, we can observe that the values of WTP for low-income users and people without car or without a drive license are lower than values of WTP calculated through the ML model for the users. This explains that the relevant

difference between users and potential users' average WTPs is due to the potential users' high income and to the car ownership prevalent profile, as we have commented above. In other words, potential high-income users who own a car or have a driver's license perceive the relevant variables in a very different way from the users.

8. Conclusion

8.1. General findings

The main objective of this paper was to try to consider potential users of the service as important elements to be considered for defining the strategy for improving the transit services, in addition to the evaluation of the perceptions of people who know the service because they use it. We retain that the right policies for improving transit services and attracting new users have to be defined and determined by taking into account opinions of both users and potential users. Thanks to the SP survey and the calibration of the discrete choice models we had the possibility to discover the differences between users' and potential users' desired quality, and also the differences among different groups of users according to the socio-economic characteristics and travel habits. The calculation of WTP provided very interesting findings and suggestions.

The willingness to pay is a well-studied subject in the literature but, to the best of the authors' knowledge, no studies reported WTP calculation from a model referred to potential users, and there is not any study where WTP is calculated for the level of occupancy. Differently, results obtained for WTP calculated for the users are in line with other studies like Wardman (2004).

The most evident finding is that the WTP values obtained for the potential users are very much higher than the values obtained for the users. Some values are five times the values obtained for users. This finding could seem strange; it could suggest that potential users did not understand well the service attributes and that their vision of the service could be quite distorted, because they do not know the public transport system. But, we retain that it is a justifiable result, because it shows that potential users are more exigent of current users; therefore, they would be willing to pay much more for saving in journey times and improving comfort levels. The two types of users are different in their attitudes towards the use of public transport, due to the fact that the first category used it and the second category prefers other transportation modes. We have observed also other interesting differences concerning socio-economics characteristics. As an example, among users women are the most exigent as regards the used service, while among potential users men would be willing to pay more than women for improving the service.

Another interesting finding concerns the owning of a private car. This characteristic is not very relevant when considering the sample of users; in fact, we found no important differences of WTP between the group of users owning a car and the users who have not a private car for travelling. In general, bus users choose the bus with the consciousness that bus is less expensive than private car, while potential users have completely different perceptions and preferences. In this case in fact, people owning a car would be willing to pay an exorbitant price for saving in terms of journey times.

By considering this finding, the strategy of the transit operator should be to improve the service in terms of journey time for capturing potential users owning a car.

Definitively, transit planners and operators could conveniently consider the opinions of potential users, and not only users, in order to satisfy the expectations of the users, but above all to accommodate the desired levels of quality of the potential users.

8.2. Policy proposals

As a result of this analysis, we can draw some policy proposals allowing to retain loyal users and attract potential users. To keep loyal users, we observe that the most important variables are Waiting time (T_w) and Journey Time (T_J). Users want an improvement in these variables that can be carried out by increasing service frequency and commercial speed. It is also necessary to consider that 60% of the users give great importance to the level of occupancy (Class2). This implies that acting on the LO we can obtain an improvement in the desired quality by the users. The use of articulated buses, the increase in service frequency and commercial speed, can solve this problem. Users who travel for study purposes are those who are willing to pay more to reduce the LO (Table 8); this is because the rush hour match with the time of entry to the schools and university. In these cases, additional services can be provided in the most critical sections of the network. To attract potential users, the LC model tell us that we must act on times and fare service attributes. There is a big difference between potential high-income and low-income users. Capturing high-income users is more complicated. These users tend to travel more than the other users, and they are more

demanding with the waiting and journey time especially; therefore, these users are willing to pay much than the other users for improving certain attributes. More, users who own a car or a driver's license have a WTP that is very different from those who do not have them. Finally, the reason for traveling home is less important to potential users than users. From the findings emerged from our analyses, we can conclude that users and potential users are completely different. Therefore, we should offer more flexible, reliable and comfortable public transport services, taking into account that potential users are willing to pay more than public transport users to benefit from a better service. As an example, public transport services with shared autonomous vehicles can be a solution to attract users to public transport by offering an alternative solution that fits their requirements.

8.3. Practical implications

Public transport sector has been going through a reform oriented to provide efficient and effective services, characterized also by high quality levels. In this manner, the different needs of the actors involved in the overall public transport system should be satisfied: maximizing revenues for transport operators, reducing the subsidies paid to the operators for transport authorities, and having high-quality services for passengers. In order to achieve this goal, transport operators and authorities' relationship changed from concession-based practices to contracting one, with a payment model based on incentives depending on performance outcomes. Specifically, in European countries competitive tendering is promoted for public transport service procurement by EU Regulation 1370/2007 (Van de Velde, 2008). Definitively, in a performance-based contracting model the payment depends on service quality provided by the operators, and an objective evaluation of the quality levels cannot prescind from an evaluation of the expectations of both users' and potential users' about each service quality aspects. By using SP experiments, the proposed methodology can support decision-makings in defining the quality standard in performance-based contracting models.

Acknowledgements

References

Allen, J., Eboli, L., Mazzulla, G., Ortúzar, JdO., 2018. Effect of critical incidents on public transport satisfaction and loyalty: An Ordinal Probit SEM-MIMIC approach. Transportation. https://doi.org/10.1007/s11116-018-9921-4

Alonso, B., Barreda, R., dell'Olio, L., Ibeas, A., 2018. Modelling user perception of taxi service quality. Transport Policy 63, 157–164.

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

Bhat, C.R., 2003. Random Utility-Based Discrete Choice Models for Travel Demand analysis. In: Goulias, K.G. (ed.), Transportation systems planning, 10(1)-10(29), CRC Press, Washington D.C., USA.

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

CEN/TC 320 (2006) Transportation – Logistics and services, European Standard EN 15140: Public passenger transport – Basic requirements and recommendation for systems that measure delivered service quality, European Committee for Standardization, Brussels. Technical Report.

CEN/TC 320 (2002) Transportation – Logistics and services. European Standard EN 13816: Public passenger transport – Service quality definition, targeting and measurement. European Committee for Standardization, Brussels. Technical Report

Cherchi, E., Hensher, D.A., 2015. Workshop synthesis: Stated preference surveys and experimental design, an audit of the journey so far and future research perspectives. Transportation Research Procedia 11, 154 – 64.

de Oña, J., de Oña, R., 2015. Quality of Service in Public Transport Based on Customer Satisfaction Surveys: A Review and Assessment of Methodological Approaches. Transportation Science 49 (3), 605–622.

de Oña J., de Oña R., Eboli L., Mazzulla G., 2016. Index numbers for monitoring Transit Service Quality. Transportation Research Part A 84, 18-30.

dell'Olio, L., Ibeas, A., Cecín, P., 2011. The quality of service desired by public transport users. Transport Policy 18, 217-227.

dell'Olio, L., Ibeas, Á., de Oña, J., de Oña, R., 2018. Public transportation quality of service: factors, models, and applications. First edition. Amsterdam, Oxford, Cambridge: Elsevier.

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

Dyer, J. S., 1990. Remarks on the Analytic Hierarchy Process. Management Science 36(3), 249-258.

Eboli L., Forciniti, C., Mazzulla G., 2018. Spatial variation of the perceived transit service quality at rail stations. Transportation Research Part A 114, 67-83.

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

Eboli L., Mazzulla G., 2010. How to capture the passengers' point of view on a transit service through rating and choice options. Transport Reviews 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(1), 172-181.

Greene, W.H., 2016. NLOGIT reference guide: version 6.0. Plainview, NY: Econometric Software, Inc.

Greene W.H., Hensher D.A., 2003. A latent class model for discrete choice analysis: contrasts with mixed logit. Transportation Research Part B 37, 681-698.

Guirao, B., García-Pastor, A., López-Lambas, M.E., 2016. The importance of service quality attributes in public transportation: Narrowing the gap between scientific research and practitioners' needs. Transportation Policy 49, 68–77.

Harker, P. T., Vargas, L. G., 1990. Reply to "Remarks on the Analytic Hierarchy Process". Management Science 36(3), 269-273.

Hausman, J., McFadden, D., 1984. Specification tests for the multinomial logit model. Econometric 52(5), 1219 – 40.

Hensher, D.A., 1994. Stated preference analysis of travel choices: the state of the practice. Transportation 21 (2), 107-33.

Hensher D.A., 2001a. The sensitivity of the valuation of travel time savings to the specification of unobserved effects" Transportation Research 37E, 129 -142.

Hensher D.A., 2001b. The valuation of commuter travel time savings for car drivers: evaluating alternative model specifications" Transportation 28, 101 – 118.

Hensher D.A., 2001c. Measurement of the valuation of travel time savings" Journal of Transport Economics and Policy 35, 71 – 98.

Hensher, D.A., Rose, J.M., Greene, W.H., 2015. Applied choice analysis. Second Edition. Cambridge University Press.

Hernandez S., Monzon A., de Oña R., 2016. Urban transport interchanges: A methodology for evaluating perceived quality. Transportation Research Part A 84, 31-43.

Jara-Díaz, S. R., & Videla, J. (1989). Detection of income effect in mode choice: theory and application. Transportation Research Part B: Methodological, 23(6), 393-400.

Landrum, H., and Prybutok, V.R., 2004. A service quality and success model for the information service industry. European Journal of Operational Research 156, 628-642.

Liekendael, JC, Furth, PG, Muller, THJ., 2006. Service quality certification in Brussels, Belgium. A quality process with teeth. Transportation Research Record 1955, 88-95.

Louviere, J.J., Flynn, T.N., Carson, R.T., 2010. Discrete choice experiments are not conjoint analysis. Journal of Choice Modelling 3(3), 57 – 72.

Louviere, J.J., Hensher, D.A., Swait, J.D., 2000. Stated Choice Methods. Analysis and Application. Cambridge University Press, United Kingdom.

Mahmoud, M., Hine, J., Kashyap, A., 2011. Bus transit service quality monitoring in UK: A methodological framework. Proceedings of the Irish Transport Research Network (ITRN) Conference, 31 August - 1 September, University College, Cork, Ireland.

McFadden, D., 1974. Conditional logit analysis of qualitative choice behaviour. In Zarembka, P. (Eds.), Frontiers in Econometrics, 105 – 42, New York Academic Press, USA.

Monzon A., Hernandez S., Cascajo R., 2013. Quality of bus service performance: Benefits of real time passenger information system. Transport and Telecommunication 14 (2), 155-166.

Ortúzar, J. de D., Willumsen, L. G., 2011. Modelling transport. Wiley-Blackwell.

Parasuraman, A., Zeithaml, V.A and Berry, L.L., 1985. A conceptual model of service quality and its implication for future research. Journal of Marketing 49, 41-50.

Powe, N.A., Garrod, G.D., McMahon, P.L., 2005. Mixing methods within stated preference environmental valuation: choice experiments and post-questionnaire qualitative analysis. Ecological Economics 52(4), 513-526.

Prasad M. D., Shekhar, B. R., 2010. Development of RAILQUAL: A Service quality scale for measuring Indian railway passenger services. Management Science and Engineering 4(3), 87-94.

Prioni, P., and Hensher, D.A., 2000. Measuring Service Quality in Scheduled Bus Services. Journal of Public Transport 3, 51-74.

Randheer, K., Al-Motawa, A. A., Vijay, P. J., 2011. Measuring commuters' perception on service quality using SERVQUAL in public transportation. International Journal of Marketing Studies 3(1), 21-34.

Ravichandran,K., Mani, B. T., Tamilnadu, D., Kumar, S.A., Prabhakaran, S., 2010. Influence of service quality on customer satisfaction: Application of SERVQUAL model. International Journal of Business and Management 5(4), 117-124.

Román, C., Martín, J.C., Espino, R., 2014. Using Stated Preferences to Analyse the Service Quality of Public Transport. International Journal of Sustainable Transportation 8, 28–46.

Rose, J.M., Bliemer, M., Hensher, D., Collins, A., 2008. Designing efficient stated choice experiments in the presence of reference alternatives. Transportation Research Part B 42, 395 – 406.

Saaty, T. L., 1990a. How to make a decision: the Analytic Hierarchy Process. European Journal of Operational Research 48(1), 9-26.

Saaty, T. L., 1990b. An exposition of the AHP in reply to the paper "Remarks on the Analytic Hierarchy Process". Management Science 36(3), 259-268.

Sillano, M., Ortúzar, J.de.D., 2005. Willingness-to-pay estimation with mixed logit models: some new evidence. Environment and Planning 37A, 525–550.

Smith, A.M., 1995. The consumer's evaluation of service quality: An examination of the SERVQUAL methodology. Doctoral dissertation, University of Manchester, 1995. British Thesis Service, D189377.

Teas, R.K., 1993. Expectations, performance evaluation and consumer's perception of quality. Journal of Marketing 57(4), 18-34.

Thurstone, L.L., 1927. A law of comparative judgment. Psychological review 34, 273-286.

Too, L., Earl, G., 2010. Public transport service quality and sustainable development: a community stakeholder perspective. Sustainable Development 18 (1), 51-61.

Van de Velde, D., 2008. A new regulation for the European public transport. Research in Transportation Economics 59, 123-132.

Wardman, M. 2004. Public transport values of time. Transport policy 11(4), 363-377.