Perceived service quality, customer satisfaction, and behavioral intentions structural equation model for the Metro of Seville, Spain

By: Rocío de Oña, José Luis Machado & Juan de Oña

Rocío de Oña, José Luis Machado & Juan de Oña (2015) Perceived service quality, customer satisfaction, and behavioral intentions structural equation model for the Metro of Seville, Spain. **Transportation Research Record**, 2538, 76-85 DOI: 10.3141/2538-09

Direct access to the published version: http://dx.doi.org/10.3141/2538-09

Perceived Service Quality, Customer Satisfaction and Behavioral Intentions: A Structural Equation Model for the Metro of Seville, Spain.

Dr. Rocío de Oña
TRYSE Research Group, University of Granada, ETSI Caminos, Canales y Puertos,
c/ Severo Ochoa, s/n, 18071 Granada (Spain)
Tel.: +34-958-249-455; Email: rociodona@ugr.es

Mr. José Luis Machado
TRYSE Research Group, University of Granada, ETSI Caminos, Canales y Puertos,
c/ Severo Ochoa, s/n, 18071 Granada (Spain)
Tel: +34-665-695-207; Email: jlmachadoleon@ugr.es

Dr. Juan de Oña
TRYSE Research Group, University of Granada, ETSI Caminos, Canales y Puertos,
c/ Severo Ochoa, s/n, 18071 Granada (Spain)
Tel: +34-958-249-979; Email: jdona@ugr.es

Word count: 5,493 words text + 6 tables/figures x 250 words (each) = 6,993

13th March 2015

ABSTRACT
This paper investigates the relationship between perceived service quality, customer satisfaction and behavioral intentions for the Metro of Seville (Spain), a metropolitan partially underground light rail transit (LRT). 3,211 users participated in the case study by completing a questionnaire with 4 different parts: users' attitudes with regards to the public transit service, users' perceptions of service quality, travel habits and socioeconomic characteristics of the participants. A 7-step analytic process was applied to the questionnaire instrument in order to adapt it to the specific characteristics of the service, to purify the scale and reduce the number of items describing the service into less number of underlying dimensions. Then, the above-mentioned relationships between perceptions, satisfaction and intentions of users in regards to the LRT service were estimated by using a Structural Equation Model (SEM) approach. Three SEM were calibrated: one for all the users, another one for captive users and the last one for non-captive users of the service. The outcomes of the models provided interesting findings and identified some differences in the attitudes of these two groups of passengers towards the LRT service.

Keywords: Service quality; customer satisfaction; behavioral intentions; light-rail transit; structural equation model

INTRODUCTION

A major concern in public transit today is to keep improving the level of quality and make it more appealing to customers (*1*). Therefore, customers' expectations need to be met by the public transit provided, which can be achieved with a framework that allows managers to monitor passengers´ perceptions about the service. Customer Satisfaction Survey (CSS) is the means most frequently used to capture customers' perceptions. Operating companies use CSS to monitor the perceptions of the users about the service they provide every year or at a 6-month frequency (*2*). These surveys are usually conducted through traditional methodologies (face-to-face surveys), although currently, internet-based surveys are increasing in popularity due to the advantages they offer, such as greater cost effectiveness and a higher degree of anonymity (*3-5*).

Several authors have considered service quality to be the vehicle to users' satisfaction (*6; 7*). However, despite the fact that service quality and customer satisfaction are distinct concepts, they have been used interchangeably in much of the literature (*8; 9*). Oliver (*9)* defined service quality as a cognitive judgment (thinking/judging), while customer satisfaction is an affective judgment (liking/pleasure), purely experiential. Therefore, the cognitive perception of service quality precedes the affective judgment of satisfaction. In recent years, much has been said about the "Service Quality (SQ)–Customer Satisfaction (CS)– Loyalty/Behavioral Intentions (BI)" paradigm (*10*), which has been analyzed in the case of public transit services such as light-rail, heavy-rail and bus transit services (*11*). This paradigm suggests that satisfaction is the link between service quality and customer loyalty (behavioral intentions) (*10; 12-14*). Additionally, some authors have found evidences that SQ may also have a direct effect on BI in the case of transit services such as high-speed rail (*15*) and in bus and heavy rail (*16*). However, reported evidence that this relationship exists in the case of a light-rail transit (LRT) service such as the Metro of Seville, Spain, is scarce or non-existent.

In the last decade, different approaches have emerged in the public transport industry to investigate the quality of a transit service, such as logit or probit models (*17*), decision trees (*18-20*) or artificial neural networks (*1; 21*). Nonetheless, Structural Equation Models (SEM) represent the most appropriate methodology when a whole phenomenon is occurring at once and needs to be modeled. In fact, SEM is generally considered to be one of the best integrated strategic methods for measuring latent factors and assessing the structural relationships among these factors (*13*). Numerous authors have used SEM methodology for examining the overall framework related to users' behavioral intentions in regards to use of public transportation (*10; 15; 22*).

Therefore, based on the previous literature, the main purpose of this study is to test the SQ-CS-BI paradigm in a LRT system, but also to ascertain if a positive direct relationship between service quality and behavioral intentions exists in that LRT service sector. Additionally, two different groups of passengers coexist in a transit service, captive versus not captive passengers, whose attitudes towards the service differ because captive users only have the LRT as transportation means. Therefore, a second objective is to discover if the overall framework describing the relationships between SQ-CS-BI is different among the two groups of passengers. Data from an ad-hoc questionnaire conducted on the Metro of Seville (Spain) was used in this research.

This paper is structured as follows: Section 2 describes the methodology used in this research (the SEM approach and the step-by-step analytic procedure applied to the questionnaire instrument); Section 3 introduces the data collected and respondents' profile; Section 4 displays the outcomes obtained from the SEM; and, finally, Section 5 describes the main conclusions drawn from the study.

METHODOLOGY

Structural Equation Modeling (SEM)

SEM allows explaining the relationships among multiple variables by examining the structure of interrelationships expressed through a series of equations. Distinct from other multivariate techniques, SEM examines more than one relationship at a time; therefore, it is a technique that tests a set of hypotheses and considers all possible information (*23*). SEM have been used in different fields of study and currently this analytical tool is firmly established and frequently used in the field of transportation (*2; 25-27*).

SEM consists of two components, a measurement model that assesses unobserved latent variables as linear functions of observed variables and a structural model that shows the direction and strengths of the relationships of the latent variables. Additionally, latent variables are classified as endogenous (dependent) or exogenous (independent) variables.

The structural model can be defined with the following basic equation (*28*):

$$\eta = B\eta + \Gamma\xi + \zeta \quad (1)$$

where

$\eta = m \times 1$ vector of the latent variables;
$\xi = n \times 1$ vector of the latent exogenous variables;
$B = m \times m$ matrix of the coefficients associated with the latent endogenous variables;
$\Gamma = m \times n$ matrix of the coefficients associated with the latent exogenous variables;
$\zeta = m \times 1$ vector of error terms associated with the endogenous variables;

The basic equations of the measurement model can be expressed as:

$$x = \Lambda_x\xi + \delta \quad (2)$$

$$y = \Lambda_y\eta + \varepsilon \quad (3)$$

where

x = column q-vector related to the observed exogenous variables;
δ = column q-vector related to the observed exogenous errors
$\Lambda_x = q \times n$ structural coefficient matrix for the effects of the latent exogenous variables on the observed ones;
y = column p-vector related to the observed endogenous variables;
ε = column p-vector related to the observed endogenous errors;
$\Lambda_y = p \times m$ structural coefficient matrix for the effects of the latent endogenous variables on the observed ones;

The model´s parameters are estimated by using a covariance analysis method that aims to obtain the model´s parameters that minimize the differences between the predicted variance-covariance matrix of the variables in the model and the observed one, while respecting the constraints of the model. There are different covariance methods, such as the Maximum likelihood, Generalized Least Squares and the Weighted Least Squares. The estimation method may be chosen considering different assumptions about the probability distribution, the scale properties of the variables, the complexity of the SEM, and the sample size (*24*). Moreover, SEM is confirmatory rather than exploratory since the researcher constructs the model by defining unidirectional effects between variables (*24*). Therefore, in this paper, SEM is used to test the proposed relationships between SQ, CS and BI in the case of the Metro of Seville.

7-step analytical procedure

This research followed a step-by-step methodology to best purify the questionnaire instrument and to test the causal relationships between Service Quality, Customer Satisfaction and Behavioral Intentions. This analytical procedure consists of 7 steps (Table 1) related to the survey design (Steps 1 to 3), the data collection process (Step 4), the reduction of the number of items into underlying dimensions (Step 5), a confirmatory analysis (Step 6) and finally, testing the structural model (Step 7).

Table 1

In Step 1, an ad-hoc CSS was designed for the Seville's LRT service, which aimed to capture customers' service quality perceptions of different aspects of the LRT system, and to ascertain customers' attitudes towards the service. The development of the questionnaire instrument was based on a detailed literature review, which consisted of running through numerous research works grounded in transit service quality (*12*), as well as examining the questionnaires used in other transit systems around the world. Moreover, the European standard *EN 13816 (29)* was used for the selection of the attributes in the survey. Then, seven experts (service operators, transport managers and researchers) checked the first version of the questionnaire instrument to ensure that the items were designed properly. Significant modifications were performed based on the experts' recommendations, which were considered in the pilot survey.

The first data collection was conducted in Step 2. Six trained interviewers collected the data of ~200 complete questionnaires through face-to-face interviews, which took place on board the LRT service (Route 1 in Seville).

The statistical method in Step 3 and Step 5 consisted of an exploratory study that was applied to better assess respondent´s scores on Service Quality. This is an iterative process based on Principal Component Analysis (PCA) and reliability analysis with the coefficient Cronbach alpha (*30*) that was used as a scale purification method, similar to the one described by Hu and Jen (*31)*. The exploratory study aimed to empirically define a lesser number of underlying factors or dimensions by grouping different attributes of the service according to the respondents' service quality perceptions, which were recorded by 37 items in the survey questionnaire. At the same time, these empirically underlying dimensions may fit the originally proposed dimensions of service quality that were initially considered. PCA is a statistical approach that can be used to analyze interrelationships among a large number of variables and to explain these variables in terms of their common underlying dimensions (*23*). PCA has been previously used in the development of customer satisfaction scales in the transportation field (*15; 27; 32*) since it allows the researcher to better understand customer's ratings of service quality and empirically analyze the dimensions that are conceptualized. Additionally, this technique allows to make estimates of the factors themselves (factor scores), which then replace the original variables in the subsequent analysis (*23*).

Therefore, in Step 3 we purified the scale developed in Step 1 by applying this iterative process to the data gathered with the pilot survey (Step 2) but also by considering the feedback provided by the interviewers who collected this data. Then, the soundness of the questionnaire was checked, and some modifications were performed: removing inappropriate questions, changing the order of the sections, reformulating the way some attributes were introduced, and so on. Finally, the definitive questionnaire was designed.

Subsequently, the second data collection process was conducted in Step 4. The survey implementation and collection of data was carried out online, via a web-based platform to conduct surveys. For the distribution process, a card marked with a code was handed out to users. This included a brief description of the survey objectives, a link to the survey website, and information on a prize raffle in order to capture users' attention. The survey code provided each respondent with an individual access to the online survey, which was accessible on computers, smartphone, tablets, etc. This data collection approach combined new technologies (internet-based surveys) with traditional methodologies (a face-to-face distribution process).

In Step 5, a second exploratory study was conducted in order to identify the dimensions that best defined service quality perceptions collected in Step 4. Additionally, factor weights were estimated in order to compute respondents' factor scores on these dimensions. Moreover, these factor scores were used as substitutes in later analysis, i.e., measures of the latent variable Service Quality in the measurement model.

A Confirmatory Factor Analysis (CFA) was developed in Step 6 to assess the construct validity of the measurement model. The constructs that composed the measurement model were SQ, CS and BI. SEM were calibrated with the maximum likelihood method by using AMOS 4.0 package. The construct validity of the model was assessed by analyzing four components: convergent validity, average variance extracted, construct reliability and discriminant validity (*23*). Convergent validity indicates that the items related to a construct converge or share a high proportion of variance in common. The amount of convergent validity can be assumed to be satisfactory if the factor loadings of the items that are related to a construct are statistically significant and ideally higher than 0.7. Moreover, the average variance extracted (AVE) is calculated as the mean variance extracted for the items loading on a construct and its recommended value is 0.5 or higher. Reliability can be assessed by Cronbach alpha and construct reliability (CRE), which indicate a good reliability if they show values 0.7 or higher. Last, discriminant validity refers to the fact that a construct is unique and captures some phenomena that another measure does not explain. We can assess discriminant validity if the average variance extracted for any two constructs is higher than the squared intercorrelation between these two constructs.

In Step 7, the structural model was tested using SEM to analyze the significance of structural coefficients and fit indices (*15*). The goodness-of-fit of the structural models was analyzed following the reported guidelines in the literature (*33*). Three types of fit indices were used: absolute, incremental and parsimony indices. Absolute fit indices determine how well a certain model fits the sample data and allow the researcher to choose the model with the superior fit. Different from other indices, absolute indices do not rely on comparison with a baseline model. Absolute fit indices include the chi-squared test, the goodness of fit index (GFI), the adjusted goodness of fit index (AGFI), the root mean square error of approximation (RMSEA) and the root mean square residual (RMR) (*33*). Furthermore, the normalized fit index (NFI) and the comparative fit index (CFI) can be classified as Incremental fit indices, which are a type of indices that compare the chi-square value to a baseline model for rejecting the null hypothesis that all variables are uncorrelated. Last, the parsimony goodness-of-fit index (PGFI) and the parsimonious normed fit index (PNFI) were used, which are respectively based upon the GFI and NFI by adjusting for loss of degree freedom.

Moreover, in order to deal with missing data, we followed the 4-step process proposed by Hair et al. (*23)*. After analyzing its randomness, the missing data was classified as not missing completely as random (non-MCAR). The robustness of different imputation means was tested and the EM imputation method was found to be the most robust, which is a method that gives reasonably consistent estimates (*23; 27*).

Finally, after the analytical process described in Table 1, we obtained a valid and reliable measurement and structural model that allowed us to test the proposed theoretical framework of SQ, CS and BI. This framework was first tested for the whole sample of users (SEM_ALL). A diagram of SEM_ALL and model's standardized factor loadings are shown in Figure 1. Subsequently, the proposed framework was tested for captive users (SEM_C) and for non-captive users (SEM_NC), in order to find differences between both groups of passengers. Captive users were identified as those who stated in the survey that they were travelling by LRT in that occasion mainly because "This was their only alternative".

DATA

This study was carried out in the Metro of Seville (Spain), a partially underground LRT. This LRT came into operation in 2009 and currently consists of a sole line, *Route 1*, with a length of 18 kilometers and 21 stations that connect four of the main municipalities in the metropolitan area of Seville. Seville municipality registers a population of about 700,000 inhabitants, but taking the other 3 boroughs into account, this number increases to 850,000 people. In 2013, the LRT carried more than 13.7 million passengers.

The survey instrument consisted of four different parts: Part A, which aims to know users' attitudes towards the service; Part B, comprised users' perceptions about the level of quality of different attributes of the service; Part C, regarding users' travel habits; and Part D, which collects users' socioeconomic information. Parts A and B gathered data based on respondents´ overall ridership and in Part C respondents were asked for information in regards to the trip that they were taking when they got the card from an interviewer.

Part B contained 37 questions related to various aspects of the LRT service, such as availability of the service, accessibility, information, timeliness, customer service, comfort, safety and environmental pollution. The perceived level of quality of each of the 37 attributes was asked with an 11-numeric scale (0-lowest quality and 10-highest quality). Additionally, respondents also rated their overall perceived level of quality of the LRT service on a same scale from 0 to 10. Likewise, questions related to users' attitudes towards the LRT service (Part A) were measured on a 11-numeric scale defined as 0-totally disagree and 10-totally agree, except for an overall satisfaction question, which was rated on a 5-point Likert scale (1-lowest level of satisfaction, 5-highest level of satisfaction).

Following the data collection process explained in Step 4 (Table 1), 19,863 cards were administered to users by four trained interviewers during a card delivery period of two weeks (May-June 2014), on weekdays, Saturdays and Sundays. Users who were invited to participate in the study had three weeks for completing the online survey. Afterwards, 3,365 responses were registered (response rate value of 17.09%), from which 3,211 were valid for subsequent analysis.

Table 2 shows the respondents' profile regarding their travel habits and socioeconomic characteristics. The sample was made up of more of females (53.3%). Most of the passengers were 18-25 years old (41.6%), followed by people in the age range of 26 to 40 years (28.8%) and of 41 to 65 years (25.5%). The major portion of participants had graduated from high school or had completed professional education or a bachelor' degree at university (41.9% and 48.5% respectively). Respondents in the sample were mainly students (41.5%) and employees (43.7%). Almost one fifth of the sample (17.8%) did not declared their income, while people who declared their monthly income were approximately evenly distributed among the four levels of incomes (Table 2). Almost three quarters (74.3%) of passengers in the sample traveled by LRT mainly for one of the following two reasons, to go to work (35.5%) or to get to school (38.8%). Half (52.1%) of the sample travelled by LRT every day. Most frequently, people stated that one of the main reasons why they were traveling on the LRT was because of the speed (66.5%), followed by the comfort (50.0%) and the lack of parking available (32.2%). 13.5% of respondents stated that the LRT was the only mode of transportation they had to make that trip, thus they were captive of this transport means. Walking was the most preferred mode of transportation to go from the origin of their trip to the LRT station and from the LRT station to the destination of their trip (62.6% and 86.3% respectively).

Table 2

RESULTS AND DISCUSSION

The analytical process followed (Table 1) resulted in a reliable and valid measurement and structural models that allowed to test the SQ, CS and BI paradigm, as well as to examine if a positive relationship exists between SQ and BI for the case study of the Metro of Seville. For the sake of clarity and brevity only PCA, CFA and SEM results of Steps 5- 7 in the analytical process are included in this paper.

The perceptions about service quality, which were recorded with 37 items, were reduced to eight dimensions by using a PCA. We carefully considered the assumptions underpinning PCA and our data met the recommendations in regards to sample size and number of variables (recommended ratio of over 5 observations per variable, which in our case is 86 observations per variable) (*23*) but also the following parameters provided support for the satisfying PCA of our sample. These eight dimensions had eigenvalues greater than one and explained a satisfactory proportion of the variation 66.13% (*23*). Furthermore, the Bartlett test had a value of 65,675 and was significant ($p<0.001$), which assesses the overall significance of the correlation matrix. The factorability of the overall set of variables and individual variables was assessed with the measure of sampling adequacy (MSA), which showed an acceptable value of 0.962 (above 0.5). Moreover, VARIMAX procedure was used to rotate the eight factors to an orthogonal simple structure. Table 3 shows the resulting dimensions and the items per dimension, in addition to factor loadings and factor weights corresponding to the factor that contains the item, for each item respectively. The item "availability of Internet and phone service in stations and on vehicle" was excluded from the analysis after testing its reliability. Based on PCA results, items that showed a factor loading of 0.4 or higher on the same factor were grouped together (*32*). Only the item "proximity of stops to origin and/or destination" (item 19) showed a factor loading lower than 0.4, however, this value can be considered as significant since the sample size is greater than 350 (*23*). Additionally, we considered this item because it showed the greatest factor loading on the dimension "Availability of the Service" and it also conceptually fitted into this dimension. After items were grouped, the dimensions (factors) were named as: "Tangible service equipment", "Accessibility", "Availability of the service", "Customer Service", "Security", “Information”, “Environmental Pollution” and “Individual space”. Moreover, factor weights were estimated with the regression method. Factor weights represent the weight each variable should get in order to compute the factor score, that is, the standardized composite measure created for each observation on each factor. Therefore, factor scores were estimated as the weighted average of the scores for the questions it contained (*32*). The estimated factor scores substituted the eight dimensions that describe SQ as items in the measurement model.

Table 3

We conducted the CFA by following the steps proposed by Hair et al. (23). We defined a measurement model with three exogenous variables: SQ, CS and BI, and we hypothesized several competing factor structures that were estimated and evaluated. The items and constructs of the resulting measurement model were verified based on CFA results, which are summarized in Table 4. In addition to the 8 dimensions derived with PCA, SQ included an item that recorded users’ scores on “Overall service quality of the service”. Moreover, CS consisted of 4 items that recorded users’ level of satisfaction or agreement in regards to “Overall satisfaction with the service of the LRT”, “The service of the LRT is good”, “I feel comfortable travelling by LRT” and “The service of the LRT meets my expectations”. Third, BI of users described their level of agreement with the statements “I will travel by LRT again under the same conditions (money, time and comfort)” and “Surely, I will use the LRT service again”.

Table 4

The construct validity of the measurement model was assessed based on CFA results (Table 4). Convergent validity is confirmed because all of the standardized loading estimates are 0.5 or higher; and most of them are above the ideal threshold of 0.7. Only the construct “Environmental pollution” is on the limit of 0.5. Furthermore, the three constructs showed a satisfactory value of AVE (0.5 or higher). In regards to construct reliability, all latent variables showed satisfactory results with a value of 0.79 or higher. Additionally, all constructs showed values of Cronbach alpha of 0.77 or higher.

In order to assess discriminant validity, AVE for any two constructs was compared with the squared intercorrelation among these constructs. The squared intercorrelation among SQ and CS was 0.71, which is higher than AVE for both constructs (0.51 and 0.53 respectively). Therefore, we conducted a chi-squared difference test to further analyse the discriminant validity of SQ and CS. Two SEM models were calibrated, one allowed SQ and CS constructs to be correlated, and the second one considered both constructs to be the same one and form one construct. The differences between both models were significant at a 0.01% level, which indicates that discriminant validity might exist between these two constructs. Additionally, the squared intercorrelation had a value of 0.38 between SQ and BI, and a value of 0.41 between CS and BI, which are lower than the AVE of the constructs respectively.

Since we found enough evidence to assess the construct validity of the measurement model, we conducted a structural equation model (SEM_ALL) to analyse the model fit to the data and to test the proposed framework of SQ, CS and BI. SEM_ALL was derived by using a competing model strategy that consisted of several SEMs calibrated in order to find the superior model based on goodness-of-fit parameters and the significance of item-construct and construct-construct relationships. Figure 1 shows SEM_ALL results (factor loadings and fit indices).

Absolute fit indices GFI and AGFI had values near the recommended value of 0.9 (GFI=0.888; AGFI= 0.846). Chi-squared test was significant (2529.8 and 87 degrees of freedom), which might indicate significant differences between the covariance matrices. However, the chi-squared test is less meaningful as the model becomes more complex and sample size becomes larger, which in our case is 3,211 respondents, therefore we carefully consider this result (*23*). On the other hand, RMSEA had a satisfactory value of 0.094 that is in the recommended range from 0.08 to 0.1. Additionally, RMR had a value of 0.139, while a value of 0 indicates a perfect fit. Furthermore, NFI and CFI values closer to 1 indicate a better fit; in our case these coefficients had a value of 0.903 and 0.906 respectively.

Figure 1

Looking into relationships between latent variables (Figure 1), SEM_ALL results proved the effect of SQ and CS with a value of 0.844 ($p<0.001$). Moreover, CS significantly affected BI of the LRT users in the sample (estimate= 0.403, $p<0.001$). Additionally, SQ had a significant positive effect on BI (estimate = 0.276, $p<0.001$). The latest result is consistent with the existing literature that reported the relationship between these constructs in other transportation services (*15; 16*). In regards to the measurement of SQ, "Overall perceived service quality", "Tangible service equipment" and "Accessibility" had the highest factor loads on this construct (estimates 0.783, 0.782, 0.775 respectively, all with $p<0.001$). On the other hand, "Environmental pollution" showed the lowest factor load on SQ (estimate = 0.492, $p<0.001$).

Additionally, we wanted to compare users who had the Metro of Seville as the only means to make that trip (captive users) and respondents who had other transport modes available in addition to the LRT (non-captive users). Table 5 shows the results of the models SEM_C and SEM_NC, which were calibrated with a group of captive users (434 respondents) and a group of non-captive users (2,777 respondents) respectively. Both models showed a significant and positive relationship between SQ and CS (estimate =0.840 and $p<0.001$ for both types of users). Moreover, the effect of CS on BI was significant and positive ($p<0.001$) for both types of users as well (0.535 for captive users and 0.373 for non-captive users).

Additionally, non-captive users showed a significant effect of SQ on BI (estimate = 0.310, $p<0.001$), whereas we could not find evidences that this relationship existed in the case of captive users (estimate = 0.105, $p=0.365$), which is a result that has been previously reported for captive transit users (*34*). Furthermore, Zhao, Webb and Shah (*34)* reported that the effect of the perceived value of the service on the intention to repurchase was significant for non-captive users but not for captive users, probably due to captive users do not have to do the cost-benefit analysis that defines the value perceived of the service. In a similar manner, our results may indicate that in the case of non-captive users, who actually chose between alternative modes of transport, SQ could be important in retaining these costumers. On the other hand, when users are captive of a transit service their intentions to re-purchase the service are not significantly affected by their SQ perceptions. However, it has been reported in the existing literature that service quality perceptions of captive users may significantly affect other dimensions of their customer loyalty, such as likelihood of recommending to others (*34*).

Table 5

CONCLUSIONS

In this study, a 7-step analytical process was conducted to develop an ad-hoc customer satisfaction questionnaire that measured quality perceptions, satisfaction and behavioral intentions in the case study of the Metro of Seville, Spain. Furthermore, a Structural Equation Model was developed to prove that the Service Quality, Customer Satisfaction and Behavioral Intentions paradigm exists in this transportation mode, as well as the existence of a positive direct relationship between Service Quality and Behavioral Intentions. Moreover, differences among users, such as captive and non-captive users, were found in regards to relationships between the three constructs of the model.

Eight dimensions were identified that measured Service Quality perceptions of users. According to their ranking of importance, they are: Tangible service equipment, Accessibility, Availability of the service, Information, Security, Customer service, Individual space and Environmental pollution. For instance, Tangible service equipment, which was the most determinant factor on Service Quality, consisted of items related to the cleanliness and lightning of stations and on vehicles, temperature and ventilation, and appropriate driving. These dimensions proved to be valid, reliable and were significantly related to Service Quality. Furthermore, similar dimensions have been reported in the existing literature (*2; 15; 27; 31; 32*).

Moreover, the structural relationship proposed between Service Quality, Customer Satisfaction and Behavioral Intentions was proved to be significant in this LRT. That is, for the whole sample, Service Quality significantly affected Customer Satisfaction, and Customer Satisfaction had a significant effect on Behavioral Intentions. These results might indicate that considerable improvements in Customer Satisfaction might occur if special attention is paid to Service Quality and that users could be more eager to use the LRT if actions to increase their satisfaction were conducted. Additionally, Service Quality affected Behavioral intention, which indicates that improvements in some aspects of Service Quality could not only make users more satisfied but also directly increase their intention to travel by LRT. This direct relationship between Service Quality and Behavioral Intention has been previously proved in high-speed rail services (*15*) and in bus and heavy rail transit (*16*). Additionally, this article provides evidence of this relationship in a LRT.

However, we should be cautious about the possibility of generalizing this relationship to other transport modes or other contexts. In fact, Chiou and Chen (*13*) could not empirically verify that this relationship existed in a Chinese airline service. Additionally, we have found differences in the structural model among captive and non-captive users. Non-captive users, which were the greater part of the sample (2,777 versus 434 respondents), showed significant relationships between Service Quality-Customer Satisfaction, Customer Satisfaction-Behavioral Intentions, and Service Quality-Behavioral Intention. In a different manner, captive users did not have a significant direct relationship between Service Quality and Behavioral Intentions. Despite SQ perceptions may not directly affect re-purchase intentions of captive users, we believe that future research should still pay special attention to the effect of SQ perceptions of captive users on other aspects of their customer loyalty. For instance, service quality improvements tailored to captive users could make them more willing to provide recommendations.

Therefore, no generalization should be performed about the causal relationship between Service Quality, Customer Satisfaction and Behavioral Intentions, not only among passengers of a particular service, but also among different public transport services. This relationship should be verified on a case-by-case basis. This paper provides practitioners with a validated 7-step analytical procedure to design an instrument to measure and evaluate the quality of the transit service from customers´ perspective and customer satisfaction, and their effect on behavioral intentions.

ACKNOWLEDGMENTS
The authors would like to thank the FEDER of European Union for financial support via project "Mejora de la calidad del TP para fomenter la movilidad sostenible: Metro de Sevilla" of the "Programa Operativo FEDER de Andalucía 2007-2013". We also thank all Agency of Public Works and the Ministry of Furtherance and Housing of Andalusia Regional Government staff and researchers for their dedication and professionalism. We thank the experts who participated in the questionnaire development and the team of trained interviewers that collected the data.

REFERENCES

1. Shen, J., and W. Li. Discrete hopfield neural networks for evaluating service quality of public transit. *International Journal of Multimedia and Ubiquitous Engineering,* Vol. 9, No. 2, 2014, pp. 331-340.
2. de Oña, J., R. de Oña, L. Eboli, and G. Mazzulla. Perceived service quality in bus transit service: A structural equation approach. *Transport Policy,* Vol. 29, No. 0, 2013, pp. 219-226.
3. Duffy, B., K. Smith, G. Terhanian, and J. Bremer. Comparing data from online and face-to-face surveys. *International Journal of Market Research,* Vol. 47, No. 6, 2005, pp. 615-630.
4. Al-Subaihi, A. A. Comparison of web and telephone survey response rates in Saudi Arabia. *Electronic Journal of Business Research Methods,* Vol. 6, No. 2, 2008, pp. 123-132.
5. Cardamone, A. S., L. Eboli, and G. Mazzulla. Drivers' road accident risk perception. A comparison between face-to-face interview and web-based survey. *Advances in Transportation Studies,* Vol. Section B 33, 2014, pp. 59-72.
6. Chen, C. F. Investigating structural relationships between service quality, perceived value, satisfaction, and behavioral intentions for air passengers: Evidence from Taiwan. *Transportation Research Part A: Policy and Practice,* Vol. 42, No. 4, 2008, pp. 709-717.
7. Chou, J. S., and C. Kim. A structural equation analysis of the QSL relationship with passenger riding experience on high speed rail: An empirical study of Taiwan and Korea. *Expert Systems with Applications,* Vol. 36, No. 3 PART 2, 2009, pp. 6945-6955.
8. Cavana, R. Y., L. M. Corbett, and Y. L. Lo. Developing zones of tolerance for managing passenger rail service quality. *International Journal of Quality and Reliability Management,* Vol. 24, No. 1, 2007, pp. 7-31.
9. Oliver, R. L. *Satisfaction: A Behavioral Perspective on the Consumer*. M.E. Sharpe, 2010.
10. Jen, W., R. Tu, and T. Lu. Managing passenger behavioral intention: An integrated framework for service quality, satisfaction, perceived value, and switching barriers. *Transportation,* Vol. 38, No. 2, 2011, pp. 321-342.
11. Oregon, T.-C. M. T. D. o., and T. C. R. Program. *Customer Satisfaction Index for the Mass Transit Industry*. IDEA Program, Transportation Research Board, National Research Council, 1995.
12. De Oña, J., and R. De Oña. Quality of service in public transport based on customer satisfaction surveys: A review and assessment of methodological approaches. *Transportation Science,* Vol. In press, n.d.
13. Chiou, Y. C., and Y. H. Chen. Service quality effects on air passenger intentions: A service chain perspective. *Transportmetrica,* Vol. 8, No. 6, 2012, pp. 406-426.

14. Dabholkar, P. A., C. D. Shepherd, and D. I. Thorpe. A comprehensive framework for service quality: an investigation of critical conceptual and measurement issues through a longitudinal study. *Journal of Retailing,* Vol. 76, No. 2, 2000, pp. 139-173.
15. Chou, P.-F., C.-S. Lu, and Y.-H. Chang. Effects of service quality and customer satisfaction on customer loyalty in high-speed rail services in Taiwan. *Transportmetrica A: Transport Science*, 2014, pp. 1-29.
16. Minser, J., and V. Webb. Quantifying the Benefits. *Transportation Research Record: Journal of the Transportation Research Board,* Vol. 2144, No. -1, 2010, pp. 111-120.
17. Dell'Olio, L., A. Ibeas, and P. Cecin. The quality of service desired by public transport users. *Transport Policy,* Vol. 18, No. 1, 2011, pp. 217-227.
18. De Oña, J., R. De Oña, and F. J. Calvo. A classification tree approach to identify key factors of transit service quality. *Expert Systems with Applications,* Vol. 39, No. 12, 2012, pp. 11164-11171.
19. de Oña, J., R. de Oña, L. Eboli, and G. Mazzulla. Heterogeneity in Perceptions of Service Quality Among Groups of Railway Passengers. *International Journal of Sustainable Transportation*, 2014.
20. De Oña, R., L. Eboli, and G. Mazzulla. Key factors affecting rail service quality in the Northern Italy: A decision tree approach. *Transport,* Vol. 29, No. 1, 2014, pp. 75-83.
21. Garrido, C., R. De Oña, and J. De Oña. Neural networks for analyzing service quality in public transportation. *Expert Systems with Applications,* Vol. 41, No. 15, 2014, pp. 6830-6838.
22. Park, J. W., R. Robertson, and C. L. Wu. Modelling the impact of airline service quality and marketing variables on passengers' future behavioural intentions. *Transportation Planning and Technology,* Vol. 29, No. 5, 2006, pp. 359-381.
23. Hair, J. F., W. C. Black, B. J. Babin, and R. E. Anderson. *Multivariate Data Analysis: A Global Perspective*. Prentice Hall, New Jersey, 2010.
24. Golob, T. F. Structural equation modeling for travel behavior research. *Transportation Research Part B: Methodological,* Vol. 37, No. 1, 2003, pp. 1-25.
25. Stuart, K., M. Mednick, and J. Bockman. Structural Equation Model of Customer Satisfaction for the New York City Subway System. *Transportation Research Record: Journal of the Transportation Research Board,* Vol. 1735, No. 1, 2000, pp. 133-137.
26. Eboli, L. M. G. Service Quality Attributes Affecting Customer Satisfaction for Bus Transit. *JOURNAL OF PUBLIC TRANSPORTATION,* Vol. 10, No. 3, 2007, pp. 21-34.
27. Carreira, R., L. Patrício, R. Natal Jorge, and C. Magee. Understanding the travel experience and its impact on attitudes, emotions and loyalty towards the transportation provider–A quantitative study with mid-distance bus trips. *Transport Policy,* Vol. 31, No. 0, 2014, pp. 35-46.
28. Bollen, K. A. *Structural Equations with Latent Variables*. Wiley, 1989.
29. UNE-EN13816. *Transportation - Logistics and services - Public passenger transport - Service quality definition, targeting and measurement.* B S I Standards, 2003.
30. Churchill, G. A., Jr. A Paradigm for Developing Better Measures of Marketing Constructs. *Journal of Marketing Research,* Vol. 16, No. 1, 1979, pp. 64-73.
31. Hu, K. C., and W. Jen. Passengers' Perceived Service Quality of City Buses in Taipei: Scale Development and Measurement. *Transport Reviews,* Vol. 26, No. 5, 2006, pp. 645-662.

32. Brons, M., M. Givoni, and P. Rietveld. Access to railway stations and its potential in increasing rail use. *Transportation Research Part A: Policy and Practice,* Vol. 43, No. 2, 2009, pp. 136-149.
33. Hooper, D., J. Coughlan, and M. R. Mullen. Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods,* Vol. 6, No. 1, 2008, pp. 53-60.
34. Zhao, J., V. Webb, and P. Shah. Customer Loyalty Differences Between Captive and Choice Transit Riders. *Transportation Research Record: Journal of the Transportation Research Board,* Vol. 2415, No. -1, 2014, pp. 80-88.

LIST OF TABLES

LIST OF FIGURES

TABLE 1 Questionnaire Instrument Design Process

Steps	Content
Step 1: Instrument development	• Literature Review • First Questionnaire • Questionnaire revised by experts
Step 2: First data collection	• Pilot Survey with face-to-face interviews
Step 3: Modified Questionnaire	• Feedback of interviewers after doing the pilot survey • Exploratory study of the pilot survey ✦ Principal Component Analysis ✦ Reliability: Cronbach´s alpha > 0.7
Step 4: Second data collection	• Web-based survey
Step 5: Exploratory study	• Exploratory study of the sample ✦ Principal Component Analysis ✦ Reliability: Cronbach´s alpha > 0.7
Step 6: Confirmatory analysis	• Construct validity ✦ Convergent validity o Factor loadings statistically significant (0.1%) o Standardized loading estimates > 0.5 (>0.7 ideally) ✦ Average Variance Extracted > 0.5 ✦ Reliability o Cronbach´s alpha >0.7 o Construct Reliability >0.7 ✦ Discriminant validity o Average variance extracted for any two constructs > squared intercorrelation among the two constructs
Step 7: Test structural model	• Fit indices ✦ Chi-squared test, GFI, AGFI, RMR, RMSEA ✦ NFI, CFI ✦ PGFI, PNFI • Significance of structural coefficients

TABLE 2 Sample Characteristics

Characteristics	Statistics
1.Gender	Male (46.6%), female (53.3%), no answer (0.1%)
2.Age	< 18 (2.8%), 18-25 (41.6%), 26-40 (28.8%), 41-65 (25.5%), >65 years old (1.0%), no answer (0.2%)
3.Qualification	No studies (0.7%), degree of secondary school (8.3%), degree of high school or professional education (41.9%), Bachelor´s degree at university or higher (48.5%), no answer (0.7%)
4. Employment status	Self-employed (7.4%), employee (36.3%), unemployed (8.6%), student (41.5%), housewife (1.7%), other (1.9%), retired (2.6%), no answer (0.2%)
5.Income level	<= 1,200 (28.7%), 1,201-1,800 (21.1%), 1,801-2,400 (16.5%), > = 2.401 (16.0%), no answer (17.8%)
6.Trip purpose	Work (35.5%), studies (38.8%), leisure (15.3%), others (10.3%)
7.Frequency of journey	Daily (52.1%), 3-4 times a week (17.9%), 1-2 times a week (13.6%), sporadically (16.4%)
8. Reason for taking the LRT (multiple response)	Price (10.2%), comfort (50.0%), speed (66.5%), frequency (28.9%), environmental reasons (16.6%), do not have driving license (14.5%), do not have vehicle (23.1%), it is my unique alternative (13.5%), lack of parking (32.2%), traffic congestion (24.8%), you cannot use your vehicle for any reason (6.0%), other (6.7%)
9. Mode of transport from origin to LRT station	Walking (62.6%), bicycle (3.4%), urban bus (5.9%), interurban bus (3.8%), car (22.2%), motorcycle (0.4%), other (1.8%)
10. Mode of transport from LRT station to destination	Walking (86.3%), bicycle (2.2%), urban bus (4.8%), interurban bus (1.2%), car (4.1%), motorcycle (0.3%), other (1.1%)
11.Type of ticket	One-way ticket (5.5%), return ticket (3.0%), 1 day pass (0.1%), bonometro (24.1%), bonoplus 45 (7.4%), transportation agency’s card (58.7%), other (1.2%)

TABLE 3 Principal Component Analysis (PCA) of Service Quality Attributes

		PCA Factor loadings	PCA Factor scores weights
Tangible service equipment			
1	Cleanliness of the stations	0.729	0.327
2	Lightning in stations	0.709	0.301
3	Lightning on vehicle	0.696	0.293
4	Cleanliness of the vehicle	0.661	0.281
5	Temperature and ventilation system on vehicle and in stations	0.504	0.216
6	Appropriate driving	0.401	0.123
Accessibility			
7	Easy access of persons with reduced mobility	0.703	0.353
8	Easy access to satiations and platforms from the street	0.697	0.324
9	Operation of elevators, escalators, etc.	0.671	0.319
10	Operation of ticket validators at the entrance and exit of stations	0.634	0.309
11	Easy use of ticket vending machines	0.617	0.301
12	Easy connection with other transportation modes such as bike rental, taxis, buses, etc.	0.561	0.253
Availability of the Service			
13	Number of trains per day (frequency of the service)	0.757	0.380
14	Waiting time on the platform	0.730	0.359
15	Speed of the trip	0.594	0.261
16	Operating hours of the service	0.575	0.271
17	Regularity of the service (absence of interruptions caused by breakdown or incidents)	0.552	0.267
18	Punctuality	0.505	0.193
19	Proximity of stops to origin and/or destination	0.334	0.131
Customer Service			
20	Effectiveness and speed of employees to respond, give information and deal with user´s daily problems	0.807	0.410
21	Courtesy of the employees	0.792	0.398
22	Performance of the Customer Service (offices, web site, contact by phone, deal with complaints, etc.)	0.735	0.356
23	Appearance of employees	0.693	0.322
Security			
24	Sense of security against theft and aggression in stations and on vehicles	0.745	0.460
25	Sense of security against accidents while traveling (crash/vehicle derailment)	0.738	0.433
26	Sense of security against slipping, falling and accidents at vehicle doors and escalators.	0.722	0.423
27	Signage of emergency exit and extinguishers	0.577	0.298
Information			
28	Updated, precise and reliable information in stations (price. operating hours. stops. service interruptions. etc.)	0.719	0.438
29	Updated, precise and reliable information on vehicles (operating hours, stops, service interruptions, etc.)	0.718	0.446
30	Clear and simple notice boards with information and directions in stations	0.634	0.360
31	Information available through other communication technologies	0.631	0.407

	(internet, phone, mobile applications, etc.)		
Environmental Pollution			
32	Noise level on the vehicle	0.875	0.421
33	Vibration level on the vehicle	0.852	0.402
34	Noise level in stations	0.822	0.393
Individual Space			
35	Seat availability in stations and on platforms	0.731	0.494
36	Level of comfort on vehicle (seat availability or enough room while standing up)	0.696	0.445

TABLE 4 Confirmatory Factor Analysis Results

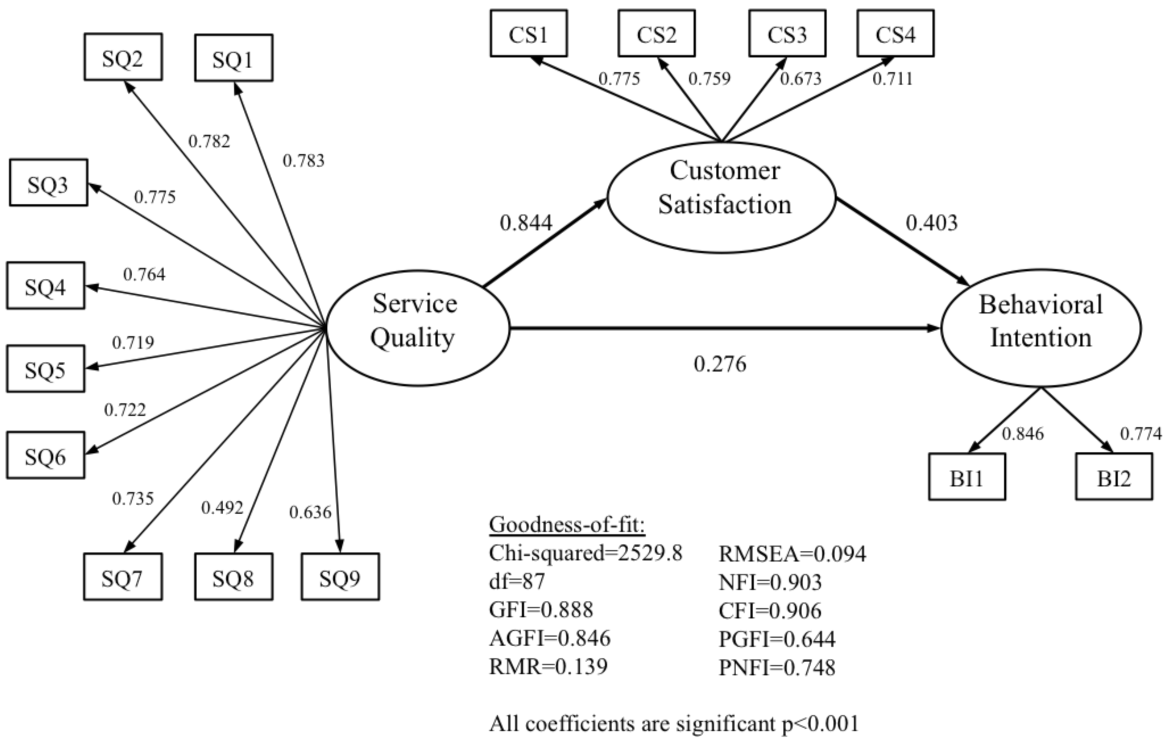
		CFA Factor Loadings	S.E.
Service Quality			
(CRE:0.90; AVE: 0.51; Cronbach Alpha: 0.892)			
SQ1	*Overall Service Quality*	*0,783*	-
SQ2	Tangible service equipment	0.782	0.019
SQ3	Accessibility	0.775	0.021
SQ4	Availability of the Service	0.764	0.021
SQ5	Customer service	0.719	0.025
SQ6	Security	0.722	0.025
SQ7	Information	0.735	0.023
SQ8	Environmental pollution	0.492	0.032
SQ9	Individual Space	0.636	0.032
Customer Satisfaction			
(CRE: 0.82; AVE: 0.53; Cronbach Alpha: 0.797)			
CS1	*Overall Satisfaction with the service of the LRT*	*0.775*	-
CS2	*The service of LRT is good*	*0.759*	*0.049*
CS3	*I feel comfortable traveling by LRT*	*0.673*	*0.051*
CS4	*The service of LRT meets my expectations*	*0.711*	*0.064*
Behavioral Intention			
(CRE: 0.79; AVE: 0.66; Cronbach Alpha: 0.77)			
BI1	*I will travel by LRT again under the same conditions (money, time and comfort)*	*0.846*	-
BI2	*Surely. I will use the LRT service again*	*0.774*	*0.021*
Goodness-of-fit statistics			
	Chi-squared	2,529.824	
	Degrees of freedom	87	
	GFI	0.888	
	AGFI	0.846	
	RMR	0.139	
	RMSEA	0.094	
	NFI	0.903	
	CFI	0.906	
	PGFI	0.644	
	PNFI	0.748	

Note: CRE: construct reliability, AVE: Average variance extracted. Factor loadings are Standardized regression weights using AMOS terminology. S.E.: standard error of the respective not standardized regression weight. All Factor loadings are significant (p<0.001). Items (not dimensions obtained with PCA) are in cursive.

TABLE 5 Factor Loadings (Standardized) and Fit Indices of Models SEM_C and SEM_NC

	Captive (SEM_C)			Non-Captive (SEM_NC)		
Factor Loadings	**Factor Loading**	**S.E.**	**p-value**	**Factor Loading**	**S.E.**	**p-value**
Among constructs						
CS <--- SQ	0.840	0.028	***	0.840	0.012	***
BI <--- SQ	0.105	0.164	0.365	0.310	0.066	***
BI <--- CS	0.535	0.353	***	0.373	0.132	***
Among items and constructs						
Overall SQ (SQ1) <--- SQ	0.799	-	-	0.773	-	-
Tangible service equipment (SQ2) <--- SQ	0.787	0.048	***	0.784	0.021	***
Accessibility (SQ3) <--- SQ	0.745	0.051	***	0.784	0.023	***
Availability of the Service (SQ4) <--- SQ	0.729	0.054	***	0.765	0.023	***
Customer Service (SQ5) <--- SQ	0.753	0.067	***	0.711	0.028	***
Security (SQ6) <--- SQ	0.700	0.068	***	0.726	0.028	***
Information (SQ7) <--- SQ	0.671	0.062	***	0.745	0.025	***
Environmental Pollution (SQ8) <--- SQ	0.484	0.082	***	0.488	0.035	***
Individual Space (SQ9) <--- SQ	0.611	0.079	***	0.635	0.036	***
CS1 <--- CS	0.775	-	-	0.768	-	-
CS2 <--- CS	0.789	0.13	***	0.751	0.055	***
CS3 <--- CS	0.707	0.138	***	0.66	0.056	***
CS4 <--- CS	0.653	0.17	***	0.717	0.071	***
BI1 <--- BI	0.858	-	-	0.840	-	-
BI2 <--- BI	0.767	0.06	***	0.777	0.022	***
Goodness-of-fit statistics						
Sample Size	434			2777		
Chi-squared	447.139			2157.543		
Degrees of freedom	87			87		
GFI	0.865			0.891		
AGFI	0.814			0.849		
RMR	0.199			0.133		
RMSEA	0.098			0.093		
NFI	0.872			0.904		
CFI	0.872			0.907		
PGFI	0.627			0.646		
PNFI	0.723			0.749		

*Note: CS: Customer Satisfaction; SQ: Service Quality; BI: Behavioral Intention. *** ($p<0.001$). Factor loadings are Standardized regression weights. S.E.: standard error of the respective not standardized regression weight.*



Note: error terms of measured items are not included for the sake of clarity.

FIGURE 1 SEM_ALL results. Factor loadings (standardized) and fit indices