Bayesian Networks and Structural Equation Modelling to develop service quality models: Metro of Seville case study

By: Francisco Diez-Mesa, Rocío de Oña, Juan de Oña

This document is a **postprint version** (ie final draft post peer review) of the following paper:

Francisco Diez-Mesa, Rocío de Oña and Juan de Oña (2018)
Bayesian Networks and Structural Equation Modelling to develop service quality models: Metro of Seville case study.
Transportation Research Part A: Policy and Practice, 118, 1-13.

Direct access to the published version: http://dx.doi.org/10.1016/j.tra.2018.08.012

Bayesian Networks and Structural Equation Modelling to develop service quality models: Metro of Seville case study

ABSTRACT

For many years it has been desirable to improve Service Quality (SQ) in Public Transport (PT) because of its strong influence on user satisfaction and its importance in attracting new passengers. Structural Equation Modelling (SEM) is one of the most widely used techniques for analysing SQ due to its ability to address different kinds of variables and to model a whole phenomenon occurring at one time. Nevertheless, its confirmative nature requires the presence of previous knowledge, a hurdle that can be overcome by applying Bayesian Networks (BN) as a technique that learns directly from data without any pre-assumptions. The aim of this paper is to apply a novel methodological approach in the field of SQ, based on a two-step process combining the techniques of BN and SEM, to model SQ in the Metropolitan Light Rail Transit (LRT) Service of Seville (Spain). The methodological approach proposed in this paper has been applied to discover and confirm the possible relationships between the LRT service characteristics and how they are related with passengers' overall perception of SQ directly from data and without the need to make assumptions. A BN was automatically learnt from the data and allowed to establish relationships between various SQ dimensions describing the service. SEM then checked the SQ model and the relationships between the dimensions extracted from the BN. The SEM model fit parameters and its consistency with the real life expected scenario supported and validated the proposed SQ model. Furthermore, the different relationships between the dimensions extracted from the BN were found to support the usefulness and potential of this methodological process for the development and confirmation of new theories and models in any field of knowledge based on data and expert supervision.

Keywords: Structural Equation Modelling, Bayesian Networks, Service Quality, Public Transport.

1. Introduction

Over recent years the study of Service Quality (SQ) and its improvement has become a relevant and crucial factor in many fields. Companies, operators and governments have focused on providing high levels of quality in their services in order to improve the current level of customer satisfaction and to attract new clients. It is important to highlight that this relationship between SQ and Satisfaction has been tested in a variety of fields, such as Marketing (Grönroos, 1984), Tourism (Shonk and Chelladurai, 2008), and On-line (Ho and Lee, 2007), etc.

In the specific field of Public Transport (PT), governments and operators have paid great attention to the study and analysis of SQ from the perspective of passengers (Andreassen, 1995; de Oña et al., 2015; Dell´Olio et al., 2011 a; Nathanail, 2008; Woods and Masthoff, 2017; etc.). Their main purpose, apart from addressing passenger satisfaction, has been to improve the ability of PT to compete with the private car by finding alternative modes of transport able to combat problems of mobility, traffic jams and pollution, etc. (Beirao and Cabral, 2007; de Oña and de Oña, 2015; Linda, 2003).

A wide range of techniques applied to the study of SQ in PT can be found in the literature (e.g., Celik et al,. 2013; Islam et al., 2016; Kuo, 2011; etc.). Their main application has been to investigate the influence different service quality factors have on overall SQ. Such knowledge allows administrations and operators to concentrate their efforts and investment in specific areas in need of improvement or of greatest importance to the passengers.

Examples of such techniques are: Importance-Performance Analysis in Weinstein (2000); A Composite Index in de Oña et al. (2016a); Multinomial Logit models in Eboli and Mazzulla (2008); the VIKOR method in Kuo and Liang (2011); etc.Structural Equation Modelling (SEM) has, nevertheless, been one of the more widely used techniques and its application in this field has grown over recent years (e.g. Amin and Isa, 2008; Chen, 2008; Hapsari et al., 2017; Yang et al., 2012; Yilmaz and Ari, 2017;etc.). The main reason behind the increased interest in SEM is its ability to easily address large numbers of variables, both endogenous and exogenous, as well as latent variables (not observed variables) explained as a linear combination of observed variables (Golob,2003). Indeed, SEM is generally considered to be one of the best integrated methods for measuring latent variables and assessing their structural relationships (Chiou and Chen, 2012; de Oña et al., 2015).

These characteristics are all crucial in the study of SQ as it has been defined as such a complex, fuzzy and abstract concept (Carman, 1990; Parasuraman et al., 1985). SQ also depends on a series of underlying observed and unobserved variables. These unobserved variables are commonly denominated as dimensions and are used to provide a better understanding of how customers perceive various service attributes (de Oña et al., 2013).

The suitability of using SEM to study SQ is, therefore, more than justified. Nonetheless, due to the confirmative nature of this technique (de Oña and de Oña., 2015; Golob, 2003), its use requires previous knowledge about how the different dimensions of the SQ models are related. This means that the bibliography or expert knowledge is usually required to develop the models (e.g.,Bagozzi, 1994; Eboli and Mazzulla, 2012; Fillone et al., 2005; Kamaruddin et al., 2012; etc.) and, depending on the context, this could cause users to miss certain important relationships in the explanation of SQ.

The above has not usually been considered in the specific case of studying SQ in PT (de Oña et al., 2017), where the scientific community agrees that SQ is directly related to and influenced by all its dimensions. Nevertheless, there is evidence in the bibliography to suggest that SQ dimensions could be interrelated and influence each other, and that their relationship with overall SQ or user satisfaction is not always direct but sometimes indirect. Evidence of this can be found in different ways (i.e., directly or indirectly) of relating the dimensions in the PT SQ models found in the literature (e.g., Chou and Kim, 2009; de Oña et al., 2017; Eboli and Mazzulla, 2012; Rahman et al., 2016; etc.). For example, de Oña et al. (2017) grouped service attributes into two latent dimensions: primary attributes (transport service factors), and secondary attributes (comfort and convenience factors). They showed that the secondary factor dimension exerted an effect on the primary attribute dimension, and its relationship with user satisfaction was not direct, but rather indirect through the primary factors dimension.

To overcome this handicap, this paper proposes a two-step procedure based on the combined use of Bayesian Networks (BN) and SEM in a single methodology, as yet unrecorded in the field of SQ. The combined use of BN and SEM has already been applied in the health sector with outstanding results (e.g., Duarte et al., 2011; Scheines et al., 1999; Trentini et al., 2015). Moreover, Kenett and Salini (2011) have presented Bayesian networks and latent variable models as suitable methodologies for analysing customer surveys.

The first step in this approach is to use BN to find certain hypotheses about the relationships between the dimensions which define SQ, followed by a second step which applies SEM to validate these relationships and the model as a whole.

Both techniques can be considered as being complementary due to the different characteristics inherent to each one. On the one hand, BN is an exploratory technique which learns directly from data without the need for pre-assumptions (Heckerman, 1998). Thus, using BN solves the problem of requiring previous knowledge about dimensional relationships. On the other hand, SEM is a confirmatory technique which allows the modelling of a phenomenon in which a set of unidirectional effects or relationships between observed and unobserved variables are established by researchers (de Oña et al,. 2015; Golob, 2003). Moreover, this technique examines more than one relationship and tests a set of hypotheses considering a large amount of information at the same time (de Oña et al., 2013; Hair et al., 2010). Therefore, SEM allows researchers to check and validate the relationships of the SQ framework extracted from the BN.

Thus, the proposed two-step methodological approach is appropriate for studying SQ. This is particularly relevant in the field of PT, due to the probable existence of several relationships between the SQ dimensions which might not be discovered and tested by other means.

The proposed research presented here is oriented towards the study of SQ in the Metropolitan Light Rail Transit Service (LRT) of Seville (Spain) by applying this two-step methodology. The main goal is to achieve a model which explains the SQ of this PT service from the point of view of passengers by using data from a Customer Satisfaction Survey (CSS) carried out on 2014.

This paper is structured as follows: Section 2 presents the PT case study, the data collection procedure and the sample characteristics. Section 3 describes the BN and SEM approaches and the two-step methodological process. The results obtained are explained and discussed in Section 4. Finally, the conclusions are reported in Section 5.

2. Methodology

This section describes all the processes that have been applied in this study. First, BN and SEM techniques are defined, followed by a detailed analysis of the two-step methodological process.

2.1. Bayesian Network (BN)

A BN is a framework for reasoning under uncertainty, and is widely used for representing uncertain knowledge (Pearl, 1988). This is a data mining technique with a wide range of advantages, one of the most important of which is that it makes complex problem analysis easy to understand as the interrelationships and dependencies of the model parameters become visible (Hänninen, 2008).

BN can be described according to two different terms. Firstly, in terms of a quantitative component consisting of a joint probability distribution that factorizes into a set of conditional probability distributions, governed by the structure of a Directed Acyclic Graph (DAG):

Let $U = \{x_1, \ldots, x_n\}, n \geq 1$be a set of variables. A BN over a set of variables U is a network structure, which is a DAG over U and a set of probability tables (1):

$$B_p = \{p(x_i|pa(x_i), x_i \in U)\} \qquad (1)$$

where $pa(x_i)$ is the set of the antecedents of x_i n BN and $i = 1, 2, 3, \dots, n$.

A BN represents joint probability distributions (2):

$$P(U) = \prod_{x_i \in U} p(x_i | pa(x_i)) \quad (2)$$

Secondly, BN can also be described in terms of a qualitative component, consisting of a DAG. In other words, BN can be defined as graphic models of the interactions between a set of variables, where the variables are represented as the nodes of a graph and the interactions (direct dependences) as directed links (also known as arcs and edges) between the nodes.

In the model, the nodes can be connected, which shows direct dependence between them, and, in the opposite case, the nodes can be not connected, which shows independence among them. A particular nomenclature for these situations of dependence/independence between nodes is: nodes with arrows directed into them are called "child", while the nodes from which the arrows depart are called "parent".

Each node contains the states of the random variable and it represents a conditional probability table. The conditional probability table of a node contains the probability of the node being in a specific state, given the states of its parents.

For a detailed description of BN the reader is directed to Kjaerulff and Madsen(2008).

2.1.1. Bayesian Networks learning and the scoring metric

The first step in a BN analysis is to learn both the structure, which has to be validated, and the parameters. There are two main approaches to learning structures in BN:

- Automatic Learning Approach: The structure of the BN is provided by algorithms which learn the structure using only one database. There are three main approaches to structural learning in BN (De Oña et al., 2011):
  a) Constraint based: Tests of conditional independence are performed on the data, and a search is made for a network that is consistent with the observed dependencies and independencies.
  b) Scored based: A score that evaluates how well the dependencies or independencies in a structure match the data is defined and a search is made for a structure that maximizes the score.
  c) Hybrid: Combines aspects of both constraint-based and score-based algorithms, as they use conditional independence tests and network scores at the same time.

  Note that before learning the BN by applying any of the previous approaches, mandatory relationships or independence can be imposed between the considered variables.

- Manual Learning Approach: The structure of a BN is provided by human expertise. This can be a highly labour intensive task, requiring a great deal of skill and creativity as well as close communication with problem-domain experts (Kjaerulff and Madsen, 2008).

In this paper, the methodology of automatic learning is used due to its multiple advantages, an example being the possibility of combining previous knowledge with automatic learning from data or finding new and unknown relationships between the variables. In addition, the purpose of using BN in this research is to achieve a model (i.e., structure) directly from data.

To apply this technique and to get the most robust BN, we have relied on the methodology explained in Cugnata et al. (2016). It lies in the application of a wide range of different learning algorithms and in measuring the arc strength using re-sampling techniques. The most robust structure reappears with specific arcs in the most learned networks and can then be chosen.

2.2. Structural Equation Modelling (SEM)

SEM can be defined as a multivariate technique combining regression, factor analysis, and variance analysis to simultaneously estimate interrelated dependence relationships. If the latent variables are classified as endogenous (dependent) and/or exogenous (independent) variables, the two components of SEM can be defined. The first is called the structural model and describes the relationships between endogenous and exogenous variables, showing the direction and strengths of the relationships between the latent variables. The second is called a measurement model and assesses the relationships between latent and observed variables.

The basic SEM equation can be defined as in Bollen (1989).

$$\eta = B\eta + \Gamma\xi + \zeta \qquad (4)$$

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

The basic equations of the measurement model can be expressed as (5) and (6).

$$x = \Lambda_x \xi + \delta \qquad (5)$$

$$y = \Lambda_y \eta + \varepsilon \qquad (6)$$

In which 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 variables; y = column p-vector related to the observed endogenous variables; ε = column p-vector related to the observed endogenous errors and $\Lambda y = p \times m$ is a structural coefficient matrix for the effects of the latent endogenous variables on the observed ones.

Different methods can be used to estimate the parameters of the model (e.g., maximum likelihood, weighted and un-weighted least squares, generalized least squares, etc.). All of them have similar goals to minimize the differences between the predicted variance-covariance matrix of the variables in the model and the observed variable, while respecting the constraints of the model. Therefore, in order to select the most suitable, each case must be analysed by focusing on the probability distribution, the scale properties of the variables, the complexity of the SEM, and the sample size (Golob, 2003). Moreover, SEM is a confirmatory rather than an exploratory technique because the researcher constructs the model by defining unidirectional effects between variables (Golob, 2003), and with this purpose in mind it was used in this study.

2.3. Methodological process: step by step

The methodological process proposed in this paper is based on two crucial steps: 1) knowledge extraction from data; and 2) relationship validation.

Step 1: Knowledge extraction from data

The BN is calibrated to find the possible relationships between the dimensions explaining SQ.

This research applied the process described in Cugnata et al. (2016) to achieve the most robust BN structure. For this purpose, an adaptation of the R script provided in Cugnata et al. (2016) has been applied to our case of study.

This step begins when different network structures are learned. With this end in mind, the data is analysed using 17 algorithms implemented in the R package *bnlearn* (i.e., Constraint-based, Scored-based and Hybrid learning algorithm) and encode in the R script of Cugnata et al. (2016). It is important to highlight that this number of algorithms was considered as being sufficient for the present case study in this research, however it could be very different depending on the criteria of the researcher.

The most robust network is then selected. The occurrence of an arc between two nodes at each learned BN is reported and scored following the process described in Cugnata et al. (2016). This process measures the occurrence of each arc in the learned BN which could be interpreted as the arc's strength. Therefore, the most robust BN has the highest number of arcs with the highest number of occurrences. A threshold of 11 was established for the occurrence score, which corresponds to almost 2/3 of occurrence in the total learned BN. An arc is considered to have a high level of occurrence if it has appeared at least 11 times in the algorithms that were implemented. Where two robust BN have the same number of arcs with high occurrence, the one with the lowest misclassification rate is selected.

Where different numbers of algorithms are considered, this threshold will be different depending on where it represents an almost 2/3 occurrence in the total learned BN.

Finally, a bootstrap re-sampling procedure is performed on the initial dataset to analyse the robustness of the chosen network and the extracted relationships. This consists of learning 1,000 BN using the same algorithm (which achieved the most robust BN in the previous step) on 1,000 randomly generated subsets with 1,000 random observations. The proportion of occurrence of each arc in the bootstrap replicates is obtained indicating the robustness of the dimension relationships extracted from the BN.

Step 2: Relationship validation

In this step the relationships between the dimensions identified with the BN are modelled and validated by SEM (*Amos Graphics v.22* software was used). For this purpose, the goodness-of-fit parameters of the structural model were analysed following Hooper et al. (2008). They showed different kinds of indices to determine the model fit (i.e., absolute, incremental and parsimony), and suggested acceptable threshold levels for each one.

The following goodness-of-fit indices were used for this research:

- Absolute fit indices: the chi-squared test, the goodness of fit index (GFI) and the adjusted goodness of fit index (AGFI), the root mean square error of approximation (RMSEA) and the root mean square residual (RMR).
- Incremental fit indices: The normed fit index (NFI) and the comparative fit index (CFI).
- Parsimony fit indices: The parsimony goodness-of-fit index (PGFI) and the parsimonious normed fit index (PNFI).

Moreover, all the relationships in the model are tested in order to be significant at a 0.05 level of confidence. If not, they are removed.

Table 1 shows a summary of the process.

Table 1.Summary of Methodology.

Steps	Content
Step 1: Knowledge extraction from data	- Following Cugnata et al. (2016): a) Learning of different network structures. b) Selection of the most robust network. c) Analysis of the robustness of the chosen network and extraction of the relationships between dimensions.
Step 2: Relationships validation	-Test Structural Model: ➢ Fit indices: - Chi-squared test, GFI, AGFI, RMR, RMSEA. - NFI, CFI. - PGFI, PNFI. ➢ Significance of structural coefficients.

3. Data Collection

The data of this study was gathered from a CSS asked about the LRT Service of Seville (Spain). This service consists of one 18 Km long line and 22 stations which are distributed throughout the city of Seville (690,566 inhabitants in 2016). The survey implementation and data collection were carried out online, via a web-based platform for conducting surveys. Passengers had three weeks (May-June 2014) to complete the online survey providing 3,365 registered responses, of which 3,198 were found to be valid for subsequent analysis.

Different kinds of subjects were covered in the survey (e.g., attitude of the passengers towards the LRT service, travel habits, socioeconomic characteristics, etc.). Nevertheless, this study was mainly focused on the perceptions of passengers about the service characteristics (i.e., the availability of the service, accessibility, information, timeliness, customer service, comfort, safety and environmental pollution). Moreover, overall passenger satisfaction with the service and their global score for SQ were also considered.

The main socioeconomic characteristics and travel habits of the sample showed a distribution between females and males of 53.30% and 46.70%, respectively. Around half of them were aged between 18-25 years (41.70%), followed by 26-40 (28.90%) and 41-65 years (25.60%). Note that in this sample the age groups younger than 18 and older than 65 were under-represented (2.80% and 1.00%, respectively). The main reasons for travelling were studies (38.80%) and work (35.50%), followed by leisure and other reasons (15.30% and 10.30%, respectively). Their frequency of use of this service is daily (52.10%) and, generally, they have a high-school diploma (41.90%) or are university graduates (48.50%). A smaller group is also represented in the sample, those who only have secondary compulsory education (8.40%). Most of the respondents have a low household monthly family income (lower than 1,800 Euros) and, there is not much difference between the percentage of passengers who had a private vehicle available to make the trip and those who did not have one (54.78% and 45.22%, respectively). The answers given for the question “Overall Service Quality” show that the passengers in this sample perceived a suitable level of SQ (average rate of 7.6 and standard deviation of 1.5).

4. Results and discussion

Many variables were considered for the case of the Seville Metro which could have influenced the results of the BN. Therefore, the data base needed to be pre-processed before taking the steps described in the proposed methodology.

4.1. Data pre-processing

The following questions were used as variables in the analysis: Thirty seven questions/SQ attributes which are related to various aspects of the LRT service, such as Availability of the service, Accessibility, Safety, etc.; one question about the overall perceived level of quality with the LRT (i.e., “Overall Service Quality”); and, finally, another question about the overall satisfaction of the passengers (i.e., “Overall Service Satisfaction”) (see Table 2).The perceived level of quality of each of the 37 attributes and the “Overall Service Quality” were asked with an 11-point Likert scale (0-lowest quality and 10-highest quality). The “Overall Service Satisfaction” was scored on a 5-point Likert scale (1-lowest satisfaction and 5-highest satisfaction).

The data pre-processing began with a re-categorization of the attributes (i.e., 37 SQ attributes and “Overall Service Quality”). The accuracy of the BN model, its complexity and the probability of each class are influenced by the number of categories of different variables (Kashani and Mohaymany, 2011). Therefore, because of the wide range of categories (11 categories) and the number of variables, the authors decided to reduce them by a re-categorization of the perceived scores, moving from an 11-point scale to a 5-point scale. The “Overall Service Satisfaction” was performed on a 5-point scale and no re-categorization was necessary.

This was followed by performing a Principal Component Analysis (PCA) on the 37 SQ attributes in order to determine the SQ latent dimensions. Another PCA was performed on the “Overall Service Quality” and “Overall Service Satisfaction” to determine the Quality of Service (QoS) latent dimension. The use of a PCA meant the number of dimensions was reduced to 9 (Table 2 shows the attributes and the dimensions obtained with the PCA). The resulting dimensions represent the variables of this study and are as follows: Accessibility (ACCESS), Availability of the Service (AVAIL), Customer Service (CUST_SER), Tangible Service Equipment (TANG_E), Security (SEC), Individual Space (IND_S), Information (INF), Environmental Pollution (ENV_POL), Quality of Service (QoS).

It is important to highlight that, based on the PCA results (Table 2), the attributes that showed a factor loading of 0.4 or higher in the same dimension were grouped together (Brons et al., 2009). Finally, the attributes B30, B1 and B3 were removed and the dimensions were extracted according to 36 attributes.

Table 2. Principal Component Analysis (PCA) results.

		PCA Factor loadings	PCA Factor weight scores
Accessibility (ACCESS)			
B8	Easy access of persons with reduced mobility	0.696	0.348
B6	Easy access to stations and platforms from the street	0.691	0.327
B7	Operation of elevators, escalators, etc.	0.671	0.326

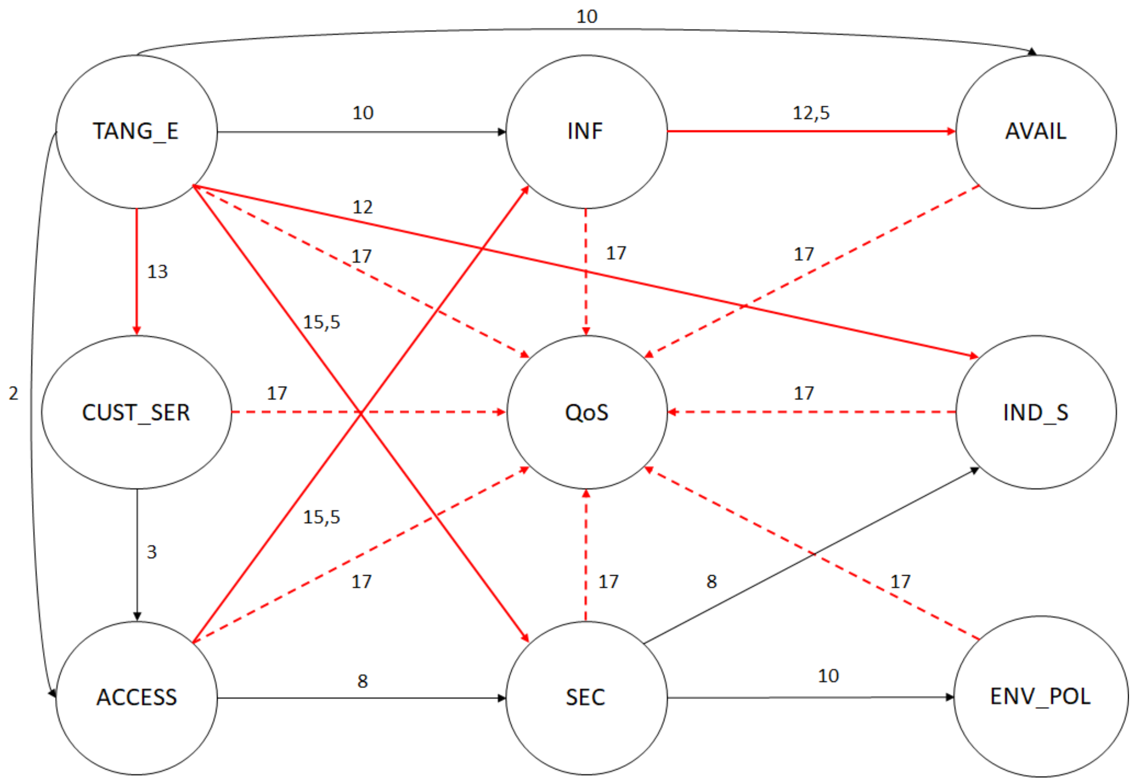
B9	Operation of ticket validators at the entrance and exit of stations	0.635	0.311
B10	Easy use of ticket vending machines	0.613	0.301
B5	Easy connection with other transportation modes such as bike rental, taxis, buses, etc.	0.548	0.242
Availability of the Service (AVAIL)			
B2	Number of trains per day (frequency of the service)	0.769	0.434
B17	Waiting time on the platform	0.728	0.414
B16	Speed of the trip	0.632	0.315
B4	Regularity of the service (absence of interruptions caused by breakdown or incidents)	0.565	0.296
B15	Punctuality	0.554	0.243
Customer Service (CUST_SER)			
B20	Effectiveness and speed of employees to respond, give information and deal with users' daily problems	0.806	0.412
B19	Courtesy of the employees	0.794	0.400
B21	Performance of the Customer Service (offices, web site, contact by phone, dealing with complaints, etc.)	0.733	0.359
B18	Appearance of employees	0.693	0.323
Environmental Pollution (ENV_POL)			
B36	Noise level on the vehicle	0.868	0.425
B37	Vibration level on the vehicle	0.847	0.407
B35	Noise level in stations	0.819	0.401
Individual Space (IND_S)			
B26	Seat availability in stations and on platforms	0.739	0.514
B27	Level of comfort on vehicle (seat availability or enough room while standing up)	0.728	0.486
Information (INF)			
B12	Updated, precise and reliable information in stations (price. operating hours. stops. service interruptions. etc.)	0.733	0.451
B11	Updated, precise and reliable information on vehicles (operating hours, stops, service interruptions, etc.)	0.733	0.442
B14	Clear and simple notice boards with information and directions in stations	0.648	0.368
B13	Information available through other communication technologies (internet, phone, mobile applications, etc.)	0.596	0.356

Security (SEC)			
B32	Sense of security against theft and aggression in stations and on vehicles	0.740	0.458
B31	Sense of security against accidents while traveling (crash/vehicle derailment)	0.735	0.431
B33	Sense of security against slipping, falling and accidents at vehicle doors and escalators.	0.715	0.416
B34	Signage of emergency exits and extinguishers	0.592	0.315
Tangible service equipment (TANG_E)			
B22	Cleanliness of the stations	0.732	0.370
B24	Lighting in stations	0.707	0.341
B25	Lighting on vehicle	0.689	0.325
B23	Cleanliness of the vehicle	0.643	0.298
B28	Temperature and ventilation system on vehicle and in stations	0.418	0.160
B29	Appropriate driving	0.340	0.085
Quality of Service (QoS)			
SQ1	Overall Service Quality	0.921	0.543
SQ2	Overall Service Satisfaction	0.921	0.543

4.2. Bayesian Networks results

The 9 dimensions extracted from the PCA and re-categorized into a 5 point – Likert scale were used for learning the BN. Before calibration, the condition that QoS was directly linked to all the SQ dimensions was imposed as mandatory for the learning structures. Although the literature provides examples where the dimensions indirectly influence SQ (de Oña et al., 2017), most researches show a direct influence (e.g., de Oña et al., 2013; Eboli and Mazzulla, 2008; Redman et al., 2013; etc.).

Figure 1 shows the most robust network. This is learned using the TABU algorithm with mBDeu score. The total score of all 17 algorithms is reported for each arc. The red arcs have a score equal to or higher than 11. Note that the arcs between QoS and the rest of the variables have a score of 17 because these relationships were set as mandatory in all the algorithms. Other red arcs showing strong relationships between SQ dimensions are TANG_E with CUST_SER, SEC and IND_S; ACCESS with INF and INF with AVAIL. There are also other relationships with scores close to the thresholds, for example, TANG_E with INF and AVAIL or SEC with ENV_POL.

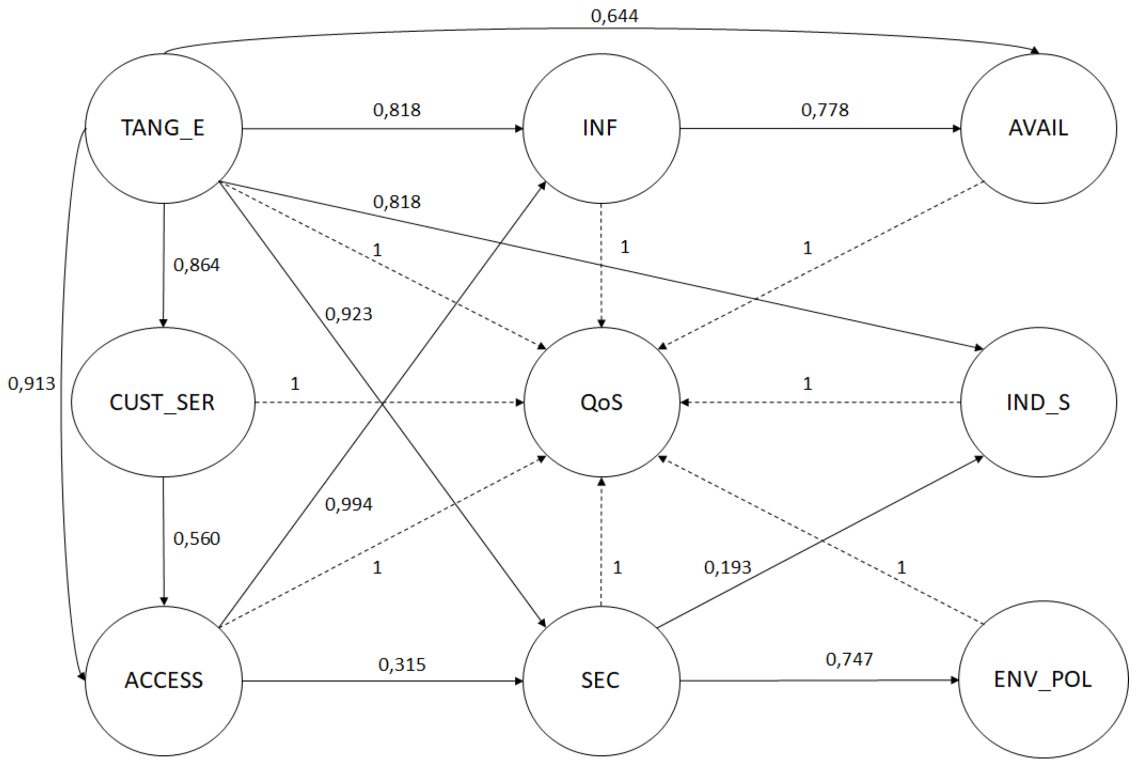


Note: the dotted arcs are mandatory relationships.

Figure 1. Learned Bayesian Network.

Figure 2 shows the occurrence proportion of each arc in the bootstrap replicates. The arcs with a proportion close to 1 are considered to be robust while those close to 0 need further analysis.

Note that almost all the red arcs have a value close to 1, which means that the relationships in the present BN are robust. Only the relationships between ACCESS and SEC, and SEC and IND_S present values under 0.5, so, their significance and relevance were analysed using SEM in order to discover whether or not they should be removed. The relationships between the 8 SQ dimensions and QoS show a value of 1 because these arcs were imposed as mandatory in accordance with the literature.



Note: the dotted arcs are mandatory relationships.

Figure 2. BN with the proportion of occurrence of each arc in bootstrap replicates.

4.3. Structural Equation Model results

The SEM was built using the dimension relationships obtained from the previous BN model. The latent variables representing the 8 SQ dimensions were explained using the 36 SQ attributes (from B2 to B37, without B1, B3 and B30) based on the PCA (Table 2) whereas the latent variable QoS, was explained with the two observed variables: “Overall Service Quality” (SQ1) and “Overall Service Satisfaction” (SQ2)

Table 3. Regression weights of Measurement Relationships.

Measurement Relationships			Unst	S.E.	St	p-value
B5	<---	ACCESS	1	-	0.625	
B6	<---	ACCESS	1.021	0.030	0.760	***
B7	<---	ACCESS	0.986	0.030	0.712	***
B8	<---	ACCESS	1.088	0.033	0.717	***
B9	<---	ACCESS	1.054	0.034	0.659	***
B10	<---	ACCESS	1.076	0.035	0.640	***
B2	<---	AVAIL	1	-	0.639	
B4	<---	AVAIL	0.835	0.031	0.561	***
B15	<---	AVAIL	0.827	0.025	0.733	***
B16	<---	AVAIL	0.966	0.029	0.739	***
B17	<---	AVAIL	1.101	0.033	0.715	***
B18	<---	CUST_SER	1	-	0.795	
B19	<---	CUST_SER	1.200	0.023	0.853	***
B20	<---	CUST_SER	1.293	0.024	0.852	***
B21	<---	CUST_SER	1.207	0.025	0.797	***
B35	<---	ENV_POL	1	-	0.775	
B36	<---	ENV_POL	1.095	0.022	0.866	***
B37	<---	ENV_POL	1.069	0.022	0.853	***
B26	<---	IND_S	1	-	0.673	
B27	<---	IND_S	1.087	0.036	0.791	***
B11	<---	INF	1	-	0.782	
B12	<---	INF	1.019	0.022	0.797	***
B13	<---	INF	0.914	0.028	0.588	***
B14	<---	INF	0.906	0.021	0.757	***
B31	<---	SEC	1	-	0.764	
B32	<---	SEC	0.925	0.024	0.711	***
B33	<---	SEC	1.013	0.025	0.741	***
B34	<---	SEC	0.858	0.021	0.733	***
B22	<---	TANG_E	1	-	0.740	
B23	<---	TANG_E	1.171	0.028	0.741	***
B24	<---	TANG_E	1.060	0.024	0.782	***
B25	<---	TANG_E	1.082	0.025	0.782	***
B28	<---	TANG_E	1.118	0.037	0.558	***
B29	<---	TANG_E	1.192	0.036	0.605	***
SQ1	<---	QoS	1	-	0.891	
SQ2	<---	QoS	0.872	0.020	0.773	***

Note: Unst., unstandardized; St., Standardized; S.E., estimate of the standard error of the covariance. *** (p<0.001).

Table 4. Regression weights of Structural Relationships.

Structural Relationships			Unst	S.E.	St	p-value
ACCESS	<---	CUST_SER	0.278	0.022	0.306	***
ACCESS	<---	TANG_E	0.608	0.031	0.534	***
AVAIL	<---	INF	0.381	0.024	0.418	***
AVAIL	<---	TANG_E	0.525	0.033	0.427	***
CUST_SER	<---	TANG_E	0.862	0.027	0.690	***
ENV_POL	<---	SEC	0.581	0.024	0.518	***
IND_S	<---	SEC	0.437	0.036	0.414	***
IND_S	<---	TANG_E	0.537	0.052	0.340	***
INF	<---	ACCESS	0.661	0.036	0.557	***
INF	<---	TANG_E	0.371	0.037	0.275	***
SEC	<---	ACCESS	0.402	0.037	0.305	***
SEC	<---	TANG_E	0.757	0.044	0.506	***
QoS	<---	AVAIL	0.530	0.030	0.507	***
QoS	<---	CUST_SER	0.139	0.023	0.135	***
QoS	<---	ENV_POL	0.058	0.014	0.075	***
QoS	<---	IND_S	0.284	0.024	0.349	***
QoS	<---	SEC	0.078	0.028	0.090	0.006
QoS	<---	TANG_E	-0.213	0.050	-0.165	***

Note: Unst., unstandardized; St., Standardized; S.E., estimate of the standard error of the covariance. *** ($p<0.001$).

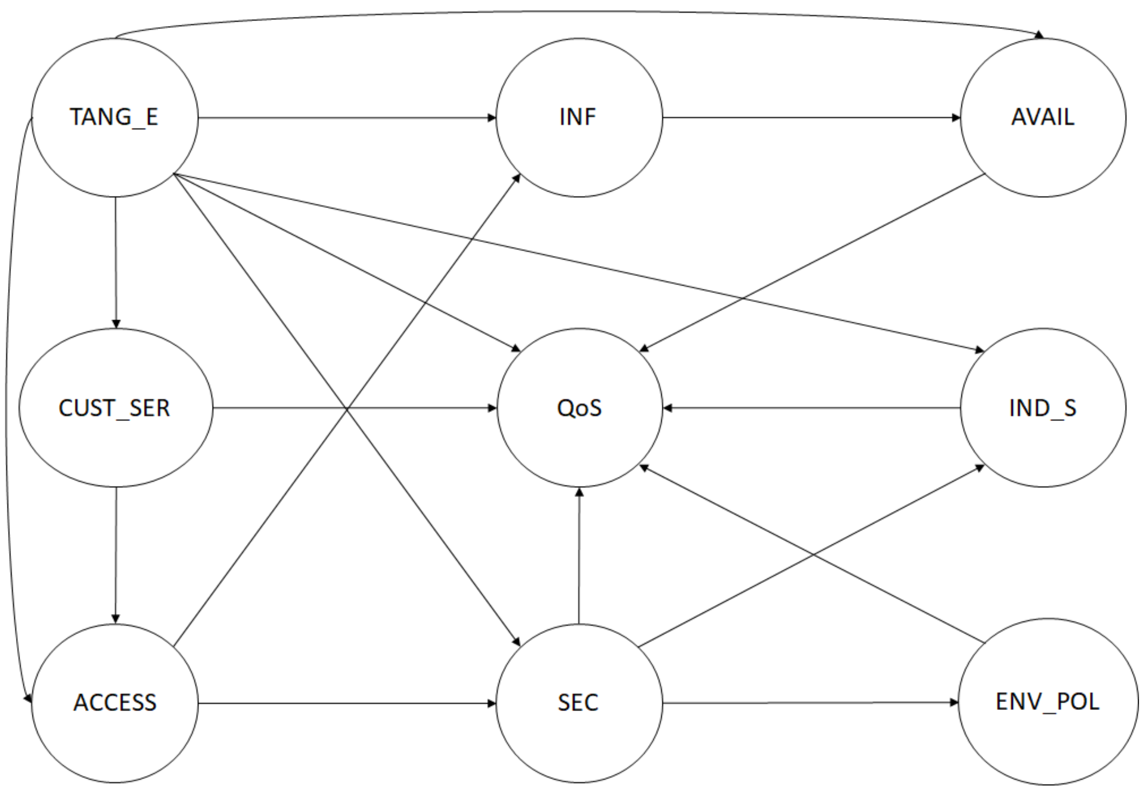


Figure 3. Structural model of SEM with significant relationships.

The results of the SEM are displayed in Table 3 and Table 4. The measurement model shows the relationships among the latent and observed variables (i.e.,the 37 SQ attributes, SQ1 and SQ2 (Table 3), while the structural model is formed by the 9 dimensions and their relationships (Table 4). Concerning the measurement model, all relationships were significant (at a significance level of 0.001), and all the standard regression weights were considerably high (St. > 0.5). SQ1 and SQ2 explained the QoS dimension very well with a standard regression weight of 0.891and 0.773, respectively.

For the case of the structural model, most of the relationships were significant at a significance level of 0.001, and all of them at a level of 0.01. Even the two relationships where the proportion of occurrence in the final BN structure (Figure 2) was under 0.5 (i.e., INF with AVAIL and SEC with ENV_POL) were significant in the SEM model. Similarly, all the attributes show significant and positive relationships with their respective dimensions at a significance level of 0.001. Almost all the dimensions show significant and positive relationships with QoS. Only TANG_E shows a significant but negative relationship with QoS. However, this fact must be analysed through the overall effect the variables have on QoS.

Therefore, Table5 shows that all the dimensions had a direct effect and ACCESS, INF, SEC and TANG_E had an indirect effect on QoS. The case of TANG_E can be seen to have had a slightly negative direct effect which can be considered to be not as important as its indirect effect.

Any analysis or study of the SQ model needs to take into account the overall effect, because if the research only considered the direct effects then TANG_E would have a negative influence on QoS, meaning the pre-assumption made in this study would not have been logical and validated and, therefore, the model would have been incorrect. However, the existence of relationships between the SQ dimensions allows the indirect effects on QoS to be included and these effects have an overall positive effect which supports our hypothesis about the different relationships between SQ and its dimensions (i.e., direct or indirect).

Considering the Standardized total effects shown in Table 5, the TANG_E and AVAIL dimensions are shown to be the most relevant aspects influencing passenger perception of QoS, whereas ENV_POL has the lowest total influence on QoS. These results agree with other analyses which used the same data base (i.e., De Oña et al., 2016b; De Oña et al., 2015; Machado-Leon et al., 2016; etc.)

Table 3. Direct, indirect and total effect on the "Quality of Service" latent variable.

		ACCESS	AVAIL	CUST_SER	ENV_POL	IND_S	INF	SEC	TANG_E
Direct Effects	**St.**	0.000	0.507	0.135	0.075	0.349	0.000	0.090	-0.165
Indirect Effects	**St.**	0.202	0.000	0.062	0.000	0.000	0.212	0.184	0.776
Total Effects	**St.**	0.202	0.507	0.197	0.075	0.349	0.212	0.274	0.611

The model's fit parameters can be seen in Table6. The fit parameters of a basic model in which the SQ dimensions have direct relationships with QoS but without any interrelationships between them are also shown (Third column). A comparison between the fit parameters of both models clearly highlights that the proposed two-step process model described in this study achieves better results than the alternative used and thereby supports the existence of relationships between the various dimensions forming SQ.

The two-step process model proposed in this study achieves the following fit parameters: Model chi-square indicates that the magnitude of discrepancy between the sample and fitted covariance matrix is insignificant at a level of 0.05, which is the threshold value suggested by several authors (e.g., Golob, 2003; Hooper et al., 2008; Mulaik et al., 1989, etc.). Moreover, the same authors indicate that the sample size should be greater than 200 for an acceptable model, as is the case here. Absolute fit indices like GFI and AGFI have values similar to the recommended value, which is 0.90 (GFI=0.921, AGFI=0.909); RMSEA (0.045) is also under 0.08 for a good fit. Incremental fit indices have comparable values (NFI=0.929and CFI=0.938); in this case a value closer to 1 indicates a good fit. The parsimony fit indices, PGFI and PNFI, have values of around 0.796 and 0.850, respectively, consistent with the statement expressed by Mulaik et al. (1989). Moreover, RMR (0.044) is below 0.05 which indicates well-fitting models. Therefore, the level of goodness-of-fit of the model was considered to be fine, even better than in other studies in the literature which have applied SEM in the field of PT and others (e.g., De Oña et al., 2013; Dimitrov, 2006; Philip et al., 2003, etc.).

Table 4. Goodness of fit measures of SEM.

	Model with Relationships between SQ dimensions	**Model without Relationships between SQ dimensions**
Fit Indices		
Chi-square	4,273.87	15,112.00
Degrees of freedom	576	586
Probability level	0.000	0.000
Number of distinct parameters to be estimated:	90	80
Absolute fit indices		
GFI	0.921	0.687
AGFI	0.909	0.643
RMSEA	0.045	0.088
RMR	0.044	0.281
Incremental fit indices		
NFI	0.929	0.750
CFI	0.938	0.757
Parsimony fit indices		
PGFI	0.796	0.604
PNFI	0.850	0.698

4.4. Discussion about the SQ Model

Figure 1 shows the relationships that are learnt using BN, and Figure 3 shows the definitive relationships of the SQ model which have been validated by SEM. Note that the relationships between INF and ACCESS with QoS have been removed. However, their influence on QoS is made indirectly through other dimensions. Therefore, it can be stated that all the different SQ dimensions have a direct or indirect influence on overall SQ.

TANG_E was shown to be the most influential dimension on SQ (total effect=0.611). Note that the important influence it has is made in an indirect way which means that the influence of TANG_E comes through other dimensions (e.g., AVAIL, ACCESS, CUST_SER, IND_S, INF or SEC). This fact can be explained by analysing the attributes grouped into TANG_E (i.e., temperature, cleanliness, lighting, appropriate driving). They might be the "base" of a good performance for these dimensions. For example, ACCESS will be perceived in a better way if the cleanliness and the lighting work perfectly, which in turn, will have a positive influence on passenger perception of QoS.

AVAIL was shown to be the second most influential dimension on SQ (total effect=0.507). This dimension has been shown in the literature to represent a basic and fundamental aspect of any public transport service (Eboli and Mazzulla, 2008; UNE-EN 13186, Tripp and Drea, 2002, etc.).

Note the cases of ACCESS and INF which did not have a direct influence on QoS. The passengers considered that the different attributes covered by ACCESS and INF did not have a direct influence on their global perception of QoS, nevertheless, they can still have a bearing on other dimensions (e.g., AVAIL or SEC) and therefore have an indirect influence on QoS. For example, if elevators and escalators operate appropriately, the sense of security against falls will improve and have a positive direct influence on passenger perception of QoS.

Finally, the ENV_POL was shown to be the dimension that had the least influence on QoS. This can be explained by the fact that the LRT Seville service is a modern environmentally friendly electric vehicle and the only pollution being produced is acoustic (noise and vibration). The passengers are fully aware of this.

Certain relevant relationships between SQ dimensions can be highlighted and could not have been identified if the proposed two-step process had not been applied.

A relevant case is t-he TANG_E dimension which influences the remaining dimensions apart from ENV_POL. It covers all the aspects relating to lighting, cleanliness (both vehicles and stations), temperature, comfort, ventilation, or even appropriate driving. A positive or negative perception of this dimension by the passengers is going to influence their perception of the remaining dimensions:

a) Suitable temperature, cleanliness and ventilation allow passengers to perceive a good level of on-vehicle comfort (Relationships with IND_S).
b) Good illumination of elevators, escalators, etc., allow passengers to easily access the stations and platforms from the street. Moreover, a good level of cleanliness for these facilities means they achieve a good score. Thus, the relationship with Access is more than justified (Relationship with ACCESS).
c) In the case of security, lighting has a direct influence on the numbers of robberies or assaults and the number of physical accidents due to bad lighting (i.e. falling, crash, slips, etc.), proving the relationship (Relationship with SEC).
d) Bad driving is going to influence the passenger perceptions about the performance of this service, from the point of view of frequency, punctuality, etc. These aspects are covered by AVAIL and support their relationships.
e) Adequate lighting in stations also helps passengers to identify the different information boards in order to guide them around the stations or in the train. (Relationships with INF).

f) Finally, passengers usually demand an adequate performance from the LRT, no excessive speeds, a good level of comfort, cleanliness, etc. In other words, these are all daily requirements or problems, the resolution of which will impact the perception passengers have about the effectiveness and speed of employees in dealing with them. Therefore, TANG_E has an influence on the CUST_SER dimension. (Relationships with CUST_SER)

The ACCESS dimension is connected to the INF dimension because well operating ticket validating or vending machines allow passengers to be reliably informed which improves their satisfaction with aspects associated to information.

The INF dimension is related to the AVAIL dimension. Updated, precise and reliable information means passengers can plan and organize their trips in advance. This fact reduces the waiting time on the platform, improves the sense of punctuality and, if this information is about the same service in other cities, it can improve their current level of satisfaction over the operating hours of the service, its frequency and regularity, etc.

Finally, the SEC dimension is linked up with the ENV_POL dimension. When passengers feel safe it entails a lower level of vibration and noise perception.

To sum up, these examples show that the relationships extracted from the BN are suitable for modelling SQ in this LRT and agree with real life expected scenarios. Bayesian Networks have also been shown to be a suitable appropriate technique for finding potential relationships which may exist between the variables used in the study. But, when combined with SEM it is possible to determine if these relationships are significant or not and, in consequence, to check and validate the SQ model framework obtained from the BN. Therefore, this study supports the use of the proposed two-step methodology and demonstrates its usefulness and relevance in developing new theories to identify new relationships between variables and to confirm their validity in this field and others.

5. Conclusions

This research demonstrates that the combined use of BN and SEM in a two-step methodological process represents a powerful tool which can be used to develop new theories or frameworks in any field of knowledge. In this study, the authors have applied this methodological approach for analysing SQ in the LRT Service of Seville (Spain). The research has demonstrated the usefulness of this methodology in identifying and validating hidden relationships between the SQ dimensions of the LRT service.

The authors have used the BN as an exploratory technique due to its ease of application, it learns directly from the data without the requirement of any pre-assumptions and its results can be easily interpreted. Furthermore, its characteristics overcome the limitations of SEM as a confirmatory technique. BN results are represented as a DAG from which the relationships between the model’s variables can be extracted and analysed. This technique allows information to be introduced to condition the links between the variables (e.g., compulsion of link, prohibition of link, etc.) where researchers have partial expert knowledge about the phenomena being studied. Similarly, SEM has been used as a confirmatory technique. It has been used to analyse the achieved SQ model due to its suitability as a technique for describing complex phenomena. SEM can be considered as a similar but more advanced technique to regression modelling; in fact, it allows researchers to introduce latent constructs which really appear in such phenomenon where latent dimensions are present. Therefore, the limitations associated with learning directly from data or the requirement for pre-existing knowledge are overcome by the characteristics of the two-step methodological procedure applied in this research.

In this study, the authors have followed the learning BN methodology shown in Cugnata et al., (2016). It consists of applying a wide range of different learning algorithms and measuring the arc strength using re-sampling techniques. In this paper, the most robust BN is learned by Hill Climbing with mBDeu score from the data of a CSS of the LRT Service of Seville (Spain). The robustness of this BN was checked by bootstraps. The final model was formed with 9 latent variables/dimensions (i.e., ACCESS, AVAIL, INF, TANG_E, CUST_SER, SEC, IND_S, ENV_POLCUST_SERENV_POL and QoS) and various observed variables (i.e. 37 SQ attributes, “Overall service quality” and “Overall Service Satisfaction”). The goodness-of-fit parameters of the SEM model provided excellent results, better than those from a model without relationships between the SQ dimensions. Moreover, combining both techniques in a two-step methodological process represents an interesting approach to model SQ in PT, as a more complex phenomenon SQ dimensions not only influence service quality, but also other SQ dimensions.

Furthermore, for this specific case, the resulting model structure provides valuable information for understanding what aspects of the service have the greater influence on passengers when they are deciding to use the service. This information can help transport managers to prepare new strategies and investment plans in order to continually improve the quality perceived by their passengers, and consequently the use of the system. Transit operators can also use these findings to attract new and retain existing passengers.

6. Acknowledgments

Support from the Spanish Ministry of the Economy and Competitiveness (Research Project TRA2015-66235-R) is gratefully acknowledged. The authors would also like to thank FEDER of the European Union for financial support via project “Mejora de la calidad del TP para fomentar la movilidad sostenible: Metro de Sevilla” of the “Programa Operativo FEDER de Andalucía 2007-2013”.

7. References

Amin, M. and Isa, Z., 2008. An examination of the relationship between service quality perception and customer satisfaction: A SEM approach towards Malaysian Islamic banking. Internation Journal of Islamic and Middle Eastern Finance and Management, 1(3), pp. 191-209.

Andreassen, T.W., 1995. (Dis)satisfaction with public services: the case of public transportation. Journal of Services Marketing, 9, pp. 30-41.

Bagozzi, R.P., 1994. Structural equations models in marketing: Basic principles. Principles of Marketing Research. Ed. Cambridge: Blackwell, pp. 317-385.

Beirao, G. and Cabral, J.A.S., 2007. Understanding attitudes towards public transport and private car: A qualitative study. Transport Policy, 14(6), pp. 478-489.

Bollen, K.A., 1989. Structural Equations with Latent Variables.Wiley, New York.

Brons, M., Givoni, M. and Rietveld, P., 2009. Access to railway stations and its potential in increasing rail use. Transportation Research Part A: Policy and Practice, 43(2), pp. 136-149.

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

Celik, E., Bilisik, O.N., Erdogan, M., Gumus, A.T., Baracli, H. (2013). An Integrated Novel Interval Type-2 Fuzzy MCDM Method to Improve Customer Satisfaction in Public Transportation for Istambul. Transportation Research Part E, 58, 28-51.

Chen, C.F., 2008. Investigating Structural Relationship Between Service Quality, Perceived Value, Satisfaction, and Behavioral Intentions for Air Passengers: Evidences from Taiwan. Transportation Research Part A: Policy and Practice, 42(4), pp. 709-717.

Chiou, Y.C., and Chen, Y.H., 2012. Service quality effects on air passenger intentions: A service chain perspective. Transportmetrica, 8(6), pp. 406-426.

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

Cugnata, F., Kenett, R.S. and Salini, S., 2016. Bayesian Networks in Survey Data: robustness and sensitivity issues. Journal of Quality Technology, 48(3), pp. 253-264.

De Oña, J. and 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), pp. 605-622

De Oña, J., De Oña, R., Diez-Mesa, F., Eboli, L. and Mazzulla, G., 2016a. A composite index for evaluating transit Service quality across different user profiles. Journal of Public Transportation, 19(2), pp. 128-153.

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

De Oña, J., De Oña, R., Eboli, L. and Mazzulla, G., 2016b. Transit passengers´ behavioural intentions: The influence of service quality and customer satisfaction. Transportmetrica A: Transport Science, 12(5), 385-412.

De Oña, J., Mujallli, R.O. and Calvo, F.J., 2011. Analysis of traffic accident injury severity on spanish rural highways using bayesian networks. Accident Analysis & Prevention, 43(1), pp. 402-411.

De Oña, R., de Abreu, J., Muñoz-Monge, C. and de Oña, J., 2017. Users´ satisfaction evolution of a metropolitan transit system in a context of economic downturn.International Journal of Sustainable Transportation (Just-Accepted)

De Oña, R., Machado, J.L. and De Oña, J., 2015. Perceived Service Quality, Customer Satisfaction and Behavioral Intentions Structural Equation Model for the Metro of Seville, Spain. Transportation Research Record, 2538, pp. 76-85.

Dell´Olio, L., Ibeas, A. and Cecin, P., 2011. The quality of service desired by public transport users.Transport Policy, 18(1), pp. 217-227.

Dimitrov, D.M., 2006. Comparing groups on latent variables: A structural equation modeling approach.Speaking of Research, 26, pp. 429-436.

Duarte, C. W., Klimentidis, Y. C., Harris, J. J., Cardel, M. and Fernandez, J. R., 2011.A hybrid Bayesian Network/Structural Equation Modelling (BN/SEM) approach for detecting physiological networks for obesity-related genetic variants. Proceedings IEE International Conference Bioinformatics and Biomedicine, pp. 696-702.

Eboli, L. and Mazzulla, G., 2008. A stated preference experiment for measuring service quality in public transport. Transportation Planning and Technology, 31(5), pp. 509-523.

Eboli L. and Mazzulla, G., 2012. Structural equation modelling for analyzing passengers´ perceptions about railway services.Procedia-Social and Behavioural Science, 54, pp. 96-106.

Fillone, A.M., Montalbo, C.M. and Tiglao., N.C., 2005.Assessing urban travel: A structural equations modelling (SEM) approach. Proceedings of the Eastern Asia Society for Transportation Studies, 5, pp. 1050-1064.

Golob, T.F., 2003. Structural equation modeling for travel behavior research. Transportation Research Part B: Methodological, 37(1), pp. 1-25.

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

Hair, J.F., Black, W.C., Babin, B.J. and Anderson, R.E., 2010. Multivariate Data Analysis: A Global Perspective. Prentice Hall, New Jersey.

Hänninen, M., 2008. Analysis of human and organizational factors in marine traffic risk modeling. Literature Review. Series AM. Espoo, Helsinki University of Technology.Faculty of engineering and architecture.Department of applied mechanics.

Hapsari, R., Clemes, M.D. and Dean, D., 2017. The impact of service quality, customer engagement and selected marketing constructs on airline passenger loyalty. International Journal of Quality and Services, 9(1), pp. 21-40

Heckerman, D., 1998. A tutorial on learning with Bayesian networks.Learning in graphical models.Springer Netherlands.

Ho, C. and Lee, Y., 2007. The development of an e-travel service quality scale. Tourism Management, 28(6), pp. 1434-1449.

Hooper, D., Coughlan, J. and Mullen, M.R., 2008. Structural equation modelling: guidelines for determining model fit. Electronic Journal of Business Research Methods, 6(1), pp. 53-60.

Islam, M.R., Hadiuzzaman, M. Banik, R., Hasnat, M.M., Musabbir, S.R., Hossain, S. (2016). Bus Service Quality Prediction and Attribute Ranking: A Neural Network Approach. Public Transport, 8(2), 295-313.

Kamaruddin, R., Osman, I. and Pei, C.A.C., 2012. Public Transport Services in Klang Valley: Customer expectations and its relationship using SEM. Procedia-Social and Behavioral Sciences, 36, pp. 431-438.

Kashani, A.T. and Mohaymany, A.S., 2011. Analysis of the traffic injury severity on two-lane, two-way rural roads based on classification tree models. Safety Science, 49(10), pp. 1314-1320.

Kenett R.S. and Salini, S., 2011. Modern analysis of customer surveys: with applications using R (Vol. 117). John Wiley & Sons.

Kjaerulff, U.B. and Madsen, A.L., 2008. Bayesian networks and influence diagrams: a guide to construction and analysis. Information Science and Statistics, Jordan, M., Kleinberg, J. and Schölkopf, B. (eds). Springer.

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

Linda, S., 2003. Can public transport compete with the private car?. IATSS Research, 27(2), pp. 27-35.

Machado-Leon, J.L., de Oña, R. and de Oña, J., 2016. The role of involvement in regards to public transit riders´ perceptions of the service. Transport Policy, 48, 34-44.

Mulaik, S.A., James, L.R., Van Alstine, J., Bennett, N., Lind, S. and Stilwell, C.D., 1989.Evaluation of goodness-of-fit indices for structural equation models. Psychological Bulletin, 105, pp. 403-445.

Nathanail, E. (2008). Measuring the Quality of Service for Passengers on the Hellenic Railways. Transportation Research Part A, 42(1), 48-66.

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

Pearl, J., 1988. Probabilistic reasoning in intelligent systems. San Francisco. CA: Morgan Kaufman.

Philip, K., Gus, M., Rodney, A., John, A., 2003.Customer repurchase intention: A general structural equation model. European Journal of Marketing. 37(11,12), pp. 1762-1800.

Rahman, F., Das, T., Hadiuzzaman, M.D. and Hossain, S., 2016. Perceived service quality of paratransit in developing countries: A structural equation approach. Transportation Research Part A: Policy and Practice, 93, pp. 23-38.

Redman, L., Friman, M., Gärling, T. and Hartig, T., 2013. Quality attributes of public transport that attract car users: A research review. Transport Policy, 25, pp. 119-127.

Scheines, R., Hoijtink, H. and Boomsma, A., 1999. Bayesian estimation and testing of structural equation models.Psychometrika, 64(1), 37-52.

Shonk, D., and Chelladurai, P., 2008. Service Quality, Satisfaction, and Intent to Return in Event Sport Tourism. Journal of Sport Management, 22, pp. 587-602.

Trentini, F., Malgaroli, M., Camerini, A.L., Di Serio, C. and Schulz, P., 2015. Multivariate determinants of self-management in Health Care: assessing Health Empowerment Model by comparison between structural equation and graphical models approaches. Epidemiology Biostatistics and Public Health, 12 (1), 1-13.

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

Woods, R., Masthoff, J., 2017. A Comparison of Car Driving, Public Transport and Cycling Experiences in Three European Cities. Transportation Research Part A, 103, pp. 211-222.

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

Yilmaz, V. and Ari, E., 2017. The effects of service quality, image, and customer satisfaction on customer complaints and loyalty in high-speed rail service in Turkey: a proposal of the structural equation model.