



# Customer satisfaction in urban rail: a study on transferability of structural equation models

Amirali Soltanpour<sup>1</sup> · Mahmoud Mesbah<sup>2,3</sup>  · Meeghat Habibian<sup>2</sup>

Accepted: 12 December 2019 / Published online: 30 January 2020  
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

## Abstract

Studies on the quality of service thrive to quantify passenger satisfaction of public transport services which is a key factor in how these services are used. Service quality evaluation is performed by means of objective and subjective approaches. Considering that people are attracted to public transport services based on their own perception, subjective approaches are utilized to capture user judgment. However, there are various subjective approaches in the literature to model service quality in which different lists of factors and model structures are identified. This paper aims to test the spatial transferability of these approaches at various levels. All the adopted approaches are based on a Structural Equation Model (SEM) to evaluate the relationship between perceived service quality and customer satisfaction. First a synthesis of the literature on model transferability is presented to compare these approaches in terms of model characteristics (structural equivalence) and parameter values (measurement equivalence). Three classes of model characteristics, including six distinct SEM structures in total, are tested to identify the factors affecting service quality: (1) recommendations of previously published models from the literature, (2) established customer satisfaction theories, and (3) adopting an exploratory approach and developing it based on the previous two approaches and the nature of locally collected data. Tehran Metro Line 3, which is a heavy urban rail mode, is used as a case study. The dataset consists of three hundred validated responses collected through a customer satisfaction survey. The results indicate that model characteristics are not transferable to this case study (i.e., none of the SEM structures from previous studies were directly transferable). The best model is developed when the entire modeling procedure is repeated using a wide range of affecting factors and gradually narrowing them down through model development. The final model comprises four latent variables, namely major services, comfort, security, and minor services, the first two of which have the largest effect on service quality.

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s12469-019-00223-y>) contains supplementary material, which is available to authorized users.

---

Extended author information available on the last page of the article

**Keywords** Public transport · Customer satisfaction · Service quality · Structural equation model · Perception · Tehran Metro

## 1 Introduction

Retaining public transport (PT) passengers or encouraging a mode shift to PT services is a strategy of many large cities that aims to reduce congestion, improve public health and decrease air and noise pollution. As passengers are the direct users of PT services, it is paramount to gain a deep understanding of service quality based on their perceptions (Eboli and Mazzulla 2015). Perception is an abstract and psychological term that plays a major role in governing behavior and making decisions (Habib et al. 2011).

To evaluate service quality, one should understand the factors affecting the formation of perceptions towards PT service quality. The significance of assessing service quality from the users' perspective has been highlighted in many references including TCRP report 165 (2013) in which a service quality framework is defined. Subjective user perception is usually collected by a Customer Satisfaction Survey (CSS). It is prevalent in these surveys to have both questions about the customers' overall satisfaction and detailed questions about specific attributes of the service. Moreover, obtaining the relative importance of these attributes is essential when an investment prioritization is needed (Cao and Cao 2017; Weinstein 2000).

In developing countries, it is more prevalent to use simple aggregate measurement tools, such as *SERVQUAL*, developed by Parasuraman et al. (1985), and there are few studies that have applied sophisticated models which can deal with users' subjective perceptions (Irfan et al. 2012; Machado-León et al. 2017; Rahman et al. 2016). Therefore, the goal of this paper is to apply a methodology that takes the concept of perception and satisfaction into account. Thus, a structural equation model (SEM) is adapted to investigate the causal relationship between customer satisfaction and perceived service quality. The advantage of an SEM is that it permits the definition of latent variables, which can reflect the subjectivity of perceived satisfaction and service quality. It should be noted that using an SEM model with all its characteristics (e.g., latent structures and parameter values) from a previous study in a new case study is questionable and requires a meticulous investigation on how well they can be transferred. However, applying the SEM structure from previous studies can be considered as a way to accelerate model development in a new study when there are solid reasons to argue for it. On the other hand, what justifies the need for a new study is that by comparing the new model results with the previous ones, an opportunity will be provided for PT researchers, planners, and decision-makers to understand how each PT service works similar to/different from the other studies, and how passengers from various cities around the world perceive their PT service quality differently. These comparisons would enable PT planners to learn how to apply and/or modify the previous successful experiments according to their own region and society. Therefore, this study includes several latent structures for service quality

that were considered in prior theories and previous studies (CEN 2002; de Oña et al. 2013; Eboli and Mazzulla 2015; Parasuraman et al. 1985). Furthermore, various SEM structures suggested in previous studies are examined on a locally collected dataset to confirm the model's spatial transferability. In other words, the precession of detected relationships in previous studies is examined with a locally collected and consistent dataset. It is investigated how previously developed model structures of other cities would perform if they were directly applied to the new case study. After exploring various model alternatives, it is proposed that the SEM model should also undergo a rigorous model development exercise by conducting an exploratory factor analysis (EFA) first. A combination of EFA and prior knowledge about the nature of data can lead to an enhanced model.

The scope of this paper covers the analysis among different models and choosing the most proper one(s), but its results, as suggested in the literature, can be considered as a determinant in further studies to: (1) inform operators, transport managers, policymakers, and researchers about the specific impact of each attribute on customer satisfaction; (2) distinguish how resources should be allocated to these attributes to improve the overall PT service quality; and (3) analyze how PT service quality can affect the ridership's spatial and temporal fluctuation (Boisjoly et al. 2018; Cao and Cao 2017; Chowdhury et al. 2018; Ingvardson and Nielsen 2019; Morfoulaki et al. 2015).

The remainder of this paper consists of the following sections: first, a literature review is provided, then the methodology is explained followed by a brief introduction to the case study. Then the data collection procedure and results are presented. The paper ends with conclusions and recommendations for future research.

## 2 Literature review

In order to move towards a customer-oriented service, PT organizations must pay more attention to service quality measurement (Stuart et al. 2000). There are two methods of measuring service quality: (1) the objective method where service quality can be expressed by performance measures (e.g., capacity, speed, reliability, and frequency). These measures can be collected by different tools, including -but not limited to- terminal surveys, automated vehicle location (AVL), and automated passenger counter (APC) (Ma et al. 2014; Mesbah et al. 2015); and (2) the subjective method that only can be measured by means of customer perceptions (Mouwen and Rietveld 2013; Vitale et al. 2016). Customer satisfaction is, in fact, one of the key determinants of personal perceptions towards a service (Chen 2008; Hensher and Prioni 2002). There is an ongoing debate about how the two concepts of service quality and customer satisfaction are related and/or separated from each other. Although, service quality and customer satisfaction are often used interchangeably in PT studies (de Oña et al. 2013; Eboli and Mazzulla 2007, 2015; Irfan et al. 2012; Stuart et al. 2000), their differences have been considered in some parts of the literature (Hadiuzzaman et al. 2017; Lai and Chen 2011; Minser and Webb 2010). For example, service quality judgments are thought to be more detailed and cognitive while customer satisfaction judgments are considered more holistic and affective

(Lai and Chen 2011; Oliver 2015). CSS is a regular tool used to collect passengers' subjective perceptions in diverse ways (e.g., on-board, online, phone, social media) (Abenoza et al. 2017; Eboli and Mazzulla 2007; Gris  and El-Geneidy 2017; Haghighi et al. 2018). In a CSS, passengers express their level of satisfaction with service quality attributes and overall service based on a pre-defined scale of evaluation [usually a Likert scale (Likert 1932)]. In addition, the relative importance of these attributes is needed to identify the impact rate of each attribute on overall satisfaction.

There are two different methods to estimate the relative importance of these attributes: stated importance (asking customers to rate each attribute on an importance scale) and derived importance (deriving a measure of attribute importance by statistically testing the strength of the relationship of individual attributes with overall satisfaction) (de O a et al. 2013; Weinstein 2000). Stated importance is the more intuitive and simpler of the two methods, but it extends the survey length, which can reduce response rate and precision of the survey. Also, stated importance can lead to deficient discrimination among importance ratings, with customers rating almost all of the measures near the top of the scale. For this reason, in recent years there has been a development in derived importance estimation based on CSS (Dell'Olivo et al. 2010; Habib et al. 2011; Imaz et al. 2015; Joewono and Kubota 2007; Weinstein 2000). Several techniques such as bivariate correlations (Figler et al. 2011; Weinstein 2000), regression analysis (Cao and Cao 2017; Dell'Olivo et al. 2010; Hensher et al. 2010; Weinstein 2000), factor analysis (Eboli and Mazzulla 2012; Javid et al. 2016; Lai and Chen 2011), and SEM (de O a et al. 2013; Eboli and Mazzulla 2015; Joewono et al. 2012; Karlaftis et al. 2001; Lai and Chen 2011; Minser and Webb 2010; Stuart et al. 2000) have been used to investigate the relationship between service quality attributes and customer satisfaction.

Because of the complex and abstract concept of service quality (Parasuraman et al. 1985), SEM has become more frequently used in recent years than the other aforementioned techniques. SEM was adopted in several fields of research and generalized by Wiley (1973) and J reskog (1970). It depends on a series of underlying observed and latent variables. The latent variables, also known as factors, can provide a better understanding of how customers perceive service quality by grouping the attributes that behave similarly into a factor representing them (Hadiuzzaman et al. 2017; Wan et al. 2016). The process of considering an aggregation of observed variables into a factor has been carried out by different methods in the literature. Some of them are based on previous research. For instance, Aydin et al. (2015) proposed a nine-latent model based on Nathanail's study (Nathanail 2008), and Shen et al. (2016) considered the *TCRP report 165* (TCRP report 165 2013). Other studies adopted a general satisfaction theory to a passenger satisfaction study, for example, Irfan et al. (2012) and Celik et al. (2014) propounded models based on the *SERVQUAL* theory (Parasuraman et al. 1985). The last group of studies used heuristic methods such as principal component analysis (PCA), EFA, and confirmatory factor analysis (CFA) (de O a et al. 2013, 2015; Eboli and Mazzulla 2012; Lai and Chen 2011; Minser and Webb 2010; Mouwen and Rietveld 2013). For further information on how to apply SEM for analyzing public transport service quality please refer to Eboli et al. (2018).

There are some issues in each of these methods. Machado-León et al. (2017) pointed out the main reason *SERVQUAL* had lost popularity over the last few years. Tyrinopoulos and Antoniou (2008) suggested that “*Factor analysis should not be blindly applied to a dataset with several variables hoping that some underlying patterns would be uncovered; instead, a theoretical motivation should drive factor analysis applications*” (pp. 263). On the other hand, using proposed models of previous studies could be questionable due to the uniqueness of the initial dataset or differences in sample frames, time frames and community characteristics (Golob 2003; Trompet et al. 2013). Transferability tests can be applied to answer this negligence.

Transferability is an issue of applying a structured and/or estimated model of one context to another one across spatial, temporal, or cultural dimensions (Atherton and Ben-Akiva 1976; Koppelman and Wilmot 1982). It has been widely applied in disaggregate travel demand models (Agyemang-Duah and Hall 1997; Atherton and Ben-Akiva 1976; Ben-Akiva and Bolduc 1987). Although many studies consider transferability as the direct application of the exact parameter values to another context, the process of considering transferability has been suggested to have four different levels: (1) basic theory of travel behavior; (2) mathematical modeling; (3) empirical model characteristics; and (4) parameter values (Hansen 1981; Sikder et al. 2013). In SEM, model transferability has been commonly investigated under the concept of multigroup equivalence, which is defined as testing SEM equivalency across a variety of comparative groups (e.g., gender, age, culture, organizations) (Byrne 2010). Cultural diversity was among the most favored multigroup equivalence studies in the past two decades (Byrne 2015). “Culture” elucidates behavioral norms within a society and evidence suggests that customers’ requirements vary across cultures (Smith and Reynolds 2002). In psychological studies, it is believed that since each cultural context can comprise diverse attitudes, norms, and values, the people from different contexts may have different perceptions about the same set of evaluating questions. Therefore, they may be represented by a distinct structure in each context (Hui and Triandis 1985). A study by Malhotra et al. (1994) investigated differences in perceived service quality between developed and developing countries. The authors illustrated the importance of considering various aspects (e.g., environmental, economic, sociocultural) which may cause distinction cross-nationally, and how they might affect the service quality evaluation. In SEM cross-cultural studies, similar to transferability studies mentioned before, there are two levels of testing for equivalence: structural equivalence and measurement equivalence (Byrne 2010; Hui and Triandis 1985). Adapting the four transferability levels of demand modeling into the transferability of customer satisfaction by SEM, it can be noted that in level 1, one should confirm that customer satisfaction can be modelled by a well-grounded theory. At level 2, SEM can be applied to mathematically model this theory. Structural and measurement equivalence tests in SEM are, respectively, similar to level 3 and 4 (Byrne 2010; Van de Vijver and Leung 1997). Studies suggest that the process should be started by the more abstract cross-cultural equivalence (i.e., structural equivalence or level 3) and then, if passed the prior stage, one can go through the later stage (i.e., level 4 or measurement equivalence) (Smith and Reynolds 2002). Measurement equivalence is commonly investigated by a set of three tests namely, “metric equivalence”, “measurement unit equivalence”, and “scalar

equivalence" (Van de Vijver and Leung 1997) while structural equivalence is investigated by a set of fit indices (Byrne and Van de Vijver 2010; Cheung and Rensvold 2002). A detailed discussion about cross-cultural equivalence has been provided by Berry et al. (2002) which is beyond the scope of this paper.

This paper aims to test the different transferability levels of various service quality structures from cities in developed countries to another city in the context of a developing country. All four levels of transferability may be considered in a cross-cultural comparison. However, due to limited access to the original data of previous studies, their models have been redeveloped by locally collected data; thus, investigating level 1, 2, and 3 (i.e., basic theory, mathematical modeling, and model characteristics, respectively) is accomplished but assessing the fourth level of transferability is impractical by a common multigroup equivalence analysis (Byrne 2016). Finally, with the findings obtained from the structure of previous studies, an enhanced model is developed which can illustrate the service quality and its affecting factors with respect to the nature of locally collected data.

### 3 Methodology

#### 3.1 Structural equation modeling

This approach permits modeling of a causal relationship by considering both observed and latent variables (i.e., indicators and factors, respectively). It comprises a series of statistical methods (e.g., factor analysis, path analysis, and regression models) to analyze the data.

SEM consists of two components: a measurement model reflecting the relationship between latent and observed variables, and a structural model assessing the strengths of causal relationships between latent variables. Moreover, latent variables can be defined as endogenous (dependent) or exogenous (independent). The basic equation of the structural model is defined as (Bollen 1989):

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

where  $\eta$  (eta) is a  $m \times 1$  vector of endogenous latent variables,  $B$  (beta) is a  $m \times m$  coefficient matrix of the endogenous latent variables,  $\Gamma$  (gamma) is a  $m \times n$  coefficient matrix of the exogenous latent variables,  $\xi$  (xi) is a  $n \times 1$  vector of exogenous latent variables, and  $\zeta$  (zeta) is a  $m \times 1$  vector of error terms associated with the endogenous latent variables.

The basic equations of the measurement model are:

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

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

where  $x$  is a  $q \times 1$  vector of the independent observed variables,  $\Lambda_x$  (lambda) is a  $q \times n$  matrix of factor loadings for the effects of independent observed variables on exogenous latent variables, and  $\delta$  (delta) is a  $q \times 1$  vector of measurement error.

Similarly,  $y$  is a  $p \times 1$  vector of dependent observed variables,  $A_y$  is a  $p \times m$  matrix of factor loadings for the effects of dependent observed variables on endogenous latent variables, and  $\varepsilon$  (epsilon) is a  $p \times 1$  vector of measurement error.

There are various SEM estimation methods, most of which are based on covariance analysis. Covariance analysis works by finding model parameters in which the difference between the covariance implied by the model and the covariance of the sample is minimized (Golob 2003). The most commonly used SEM estimation method is Maximum Likelihood (ML). In addition, SEM parameters can be estimated by other methods, such as generalized least squares (GLS), asymptotically distribution-free weighted least squares (ADF-WLS) among others (Bollen 1989; Golob 2003). However, choosing the suitable estimation method depends on different assumptions (e.g., probability distribution, the scale of the variables, sample size) (Golob 2003).

### 3.2 Studied structures

This paper uses three classes of structures to achieve a structure which represents the factors affecting service quality appropriately. These structures are as follows.

#### 3.2.1 Previous studies

As discussed in the literature review section, there have been many studies applying SEM with an exploratory and case-specific structure. de Oña et al. (2013) introduced a model with three exogenous latent variables, *Service*, *Comfort*, and *Personnel* and one endogenous latent variable, *Overall service quality*. This structure is referred to by *S1* in this paper. Eboli and Mazzulla (2015) presented a latent structure which consists of seven exogenous latent variables, *Safety*, *Cleanliness*, *Comfort*, *Service*, *Additional Services*, *Information*, and *Personnel*; these latent variables are linked to an endogenous latent variable of *Service Quality*. This structure is referred to by *S2* in this paper.

#### 3.2.2 Established customer satisfaction theories

Apart from the above studies which apply a case-specific latent structure, available general theories about assessing customer satisfaction in which factors affecting service quality are identified. Parasuraman et al. (1988) introduced the *SERVQUAL* instrument, which consists of five dimensions, *Tangibles*, *Reliability*, *Responsiveness*, *Assurance*, and *Empathy*, to assess customer satisfaction. The European Committee for Standardization recommended a Quality Criteria system of eight factors, *Availability*, *Accessibility*, *Information*, *Time*, *Customer Care*, *Comfort*, *Security*, and *Environmental Impact*, to evaluate customers' perceptions (CEN 2002). In this paper, structures *S3* and *S4* are based on these two theories, respectively.

### 3.2.3 Exploratory approach

When using models and theories proposed in previous studies, there is a critical consideration that originates from the nature of data. In fact, it can cause ambiguity due to the uniqueness of the initial dataset or differences in community characteristics (Golob 2003; Trompet et al. 2013). As a result, in this study, another structure (S5) is considered based on an EFA which leads to the formation of five exogenous latent variables namely *Comfort*, *Security*, *Major Services*, *Cleanliness*, and *Personnel* accompanied by one endogenous latent variable named *Customer Satisfaction*.

As discussed in Sect. 2, applying factor analysis to a dataset without taking the context into account may result in relations in which their causality can not be easily justified. Instead, a cause-and-effect theoretical basis should lead the application of factor analysis. In this paper, structure S6 is developed by a constrained EFA with respect to prior knowledge about the nature of data. Four exogenous latent variables are formed in this approach, *Major Services* (i.e., main attributes related to the performance of service), *Comfort* (i.e., attributes associated with feeling comfortable during a trip) (Duarte et al. 2010), *Security* (i.e., attributes associated with feeling safe and secure), and *Minor Services* (i.e., attributes generally related to staff performance). Moreover, in S6, the conceptual differences between service quality and customer satisfaction, also discussed in Sect. 2, is acknowledged in order to investigate the debate about their uniformity/discrepancy. As a result, a second-order latent variable of *Service Quality* is considered which is reflected by the four exogenous latent variables (*Major Services*, *Comfort*, *Security*, and *Minor Services*). This *Service Quality* is the latent variable which has a direct causal relationship with the endogenous latent variable of *Customer Satisfaction*.

### 3.3 Applied data and modeling procedure

Since the data of each previous study was collected independently, there was neither a unified framework of data collection nor a unique list of attributes among them. Furthermore, access to the original data of all previous studies was not possible. Therefore, those models were redeveloped by the locally collected data. The attributes were matched as much as possible and their factor loadings are estimated by means of a CFA, the details of which is given in Table 2. The reliability and validity of every six structures have been checked and compared with each other. Afterwards, an SEM model is developed for each of the six structures in order to make them also comparable by their goodness-of-fit indices. Figure 1 shows a conceptual formation of S1, S5, and S6. For the sake of consistency, the endogenous latent variable is named “customer satisfaction” in all the six models (in order to occupy less space, observed variables are not drawn).



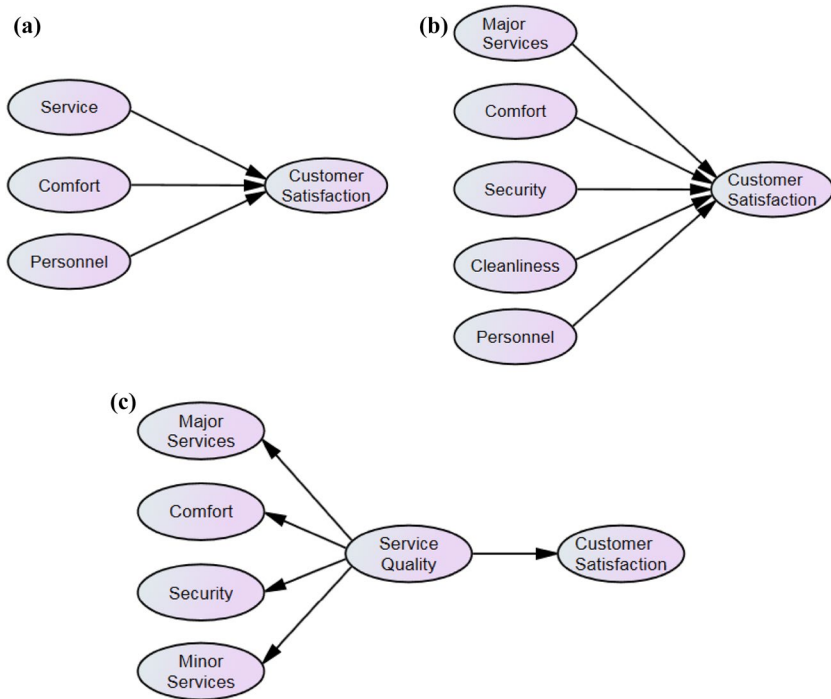


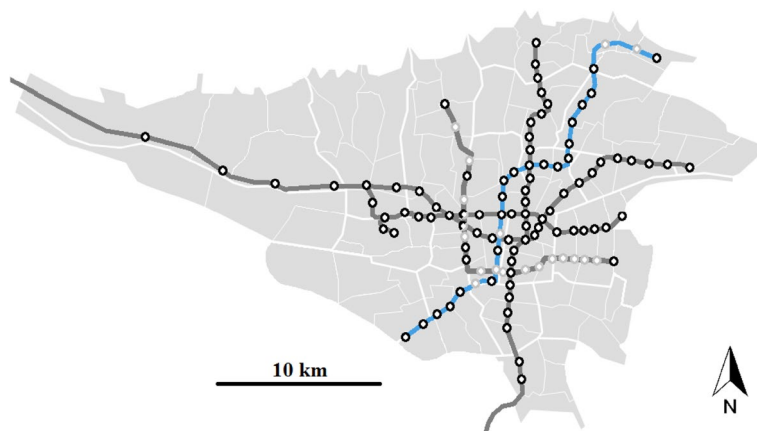
Fig. 1 Conceptual formation of S1 (a), S2 (b), and S6 (c)

## 4 Survey

### 4.1 Case study

Tehran, the capital of Iran, is the most populous city in the country and the second-most populated metropolitan area in the Middle East. It has a population of approximately 8.9 million in the central city and 15.1 million in the metropolitan area. Its area is about 700 km<sup>2</sup> (270.2 mi<sup>2</sup>) (Habibian and Rezaei 2017). There are 18.3 million daily vehicle trips in Tehran showing 23% growth in the past decade. The private vehicle share is about 40% while common public transport modes including minibus, bus, and rail carry 38% of the trips and the rest of the motorized trips are made by a shared taxi mode (also called jitney). The mass rail transport system (also called Metro) consists of five lines with about 135 km (84 mi) of double tracks and more than 100 stations connecting four corners of the city. Another major line is being constructed and two others are planned. The Metro system carries 18% of total motorized trips (Tehran Municipality 2016).

Tehran Metro Line 3 is one of the newest lines opened in 2012. It is the second longest rail transport in Tehran, connecting the northeast to the southwest, with 37 km (23 mi) of double tracks containing 23 stations. It provides about 350,000 daily trips over 18 h a day (Tehran Municipality 2016; Tehran Traffic and



**Fig. 2** Tehran Metro Line 3 (blue line) (color figure online)

Transportation Organization 2013). Figure 2 depicts a schematic map of Line 3 (blue line—starting from upper rightmost).

## 4.2 Data collection

A paper-based questionnaire was designed for this study. The questionnaire included 47 questions divided into three sections: (1) satisfaction and importance level; (2) trip characteristics; and (3) demographic data. In the first section, passengers stated their perceived satisfaction and perceived importance for 22 service quality attributes (see Table 1) based on a 5-point Likert scale. Afterward, their current trip satisfaction and all Metro trips satisfaction were questioned separately. The second section asked questions about the passengers' travel characteristics, such as their access mode, egress mode, the reason for using Metro. The last section, demographic data, included information about the passengers' age, gender, household income, household size, etc. (Soltanpour et al. 2018).

The data were collected through face-to-face interviews conducted on-board the Tehran Metro Line 3. Interviewers were graduate students of the transportation group. Not only their independence from the operator company made passengers more comfortable to express personal explicit judgment, but also their familiarity with transport issues has led them to easily explain any ambiguity in the questionnaire to the respondents with plain and clear examples (e.g., difference between safety and security, definition of attributes' satisfaction and importance). Data collection lasted about 14 weekdays in January and February 2017 (all normal working days). Eventually, after eliminating 37 incomplete questionnaires, there were 300 validated responses for use in the modeling procedure. Table 1 shows the list of collected attributes, the mean and standard deviation of their satisfaction and importance scores stated by the passengers.

**Table 1** Mean and standard deviation of the collected attributes

Service quality attributes	Satisfaction		Importance	
	Mean	Standard deviation	Mean	Standard deviation
Access time to station	3.74	1.11	4.60	0.80
Ticket purchasing or recharging	4.45	0.88	4.50	0.83
Ticket presenting	4.42	0.87	4.49	0.89
Fare amount	4.00	1.14	4.50	0.97
Staff behavior	3.99	1.05	4.54	0.81
Waiting time	2.59	1.32	4.84	0.45
Seat availability on board	2.92	1.44	4.38	0.91
Air conditioning on board	3.82	1.16	4.67	0.67
Travel time	3.83	1.08	4.70	0.62
Transfer time	3.60	1.15	4.50	0.76
Egress time to destination	3.79	1.12	4.71	0.61
Crowdedness on platform	3.30	1.29	4.46	0.82
Crowdedness on board	3.08	1.41	4.67	0.57
Information availability in station	4.00	1.16	4.54	0.83
Information availability on board	3.92	1.07	4.55	0.81
Safety in station	4.10	1.04	4.76	0.55
Safety on board	3.72	1.18	4.74	0.59
Security in station	4.09	1.10	4.84	0.43
Security on board	3.76	1.25	4.87	0.39
Cleanliness in station	4.54	0.74	4.74	0.55
Cleanliness on board	4.40	0.76	4.76	0.52
Ethical and behavioral messages	3.86	1.15	4.25	1.11
Current trip satisfaction	3.91	0.94	–	–
All metro trips satisfaction	3.39	1.00	–	–

### 4.3 Sample characteristics

Following data collection, the first step is to analyze the respondents. Some information about the sample is provided as follows. The sample comprised more males (59%), most were aged below 40 (74.3%), and many had an academic degree (a university graduate) (69.7%) conforming census data in most districts Line 3 passes through which are amongst the most educated districts in Tehran (Atlas of Tehran metropolis 2011). Figure 3 illustrates summarized information about the respondents' demographic and their travel habits.

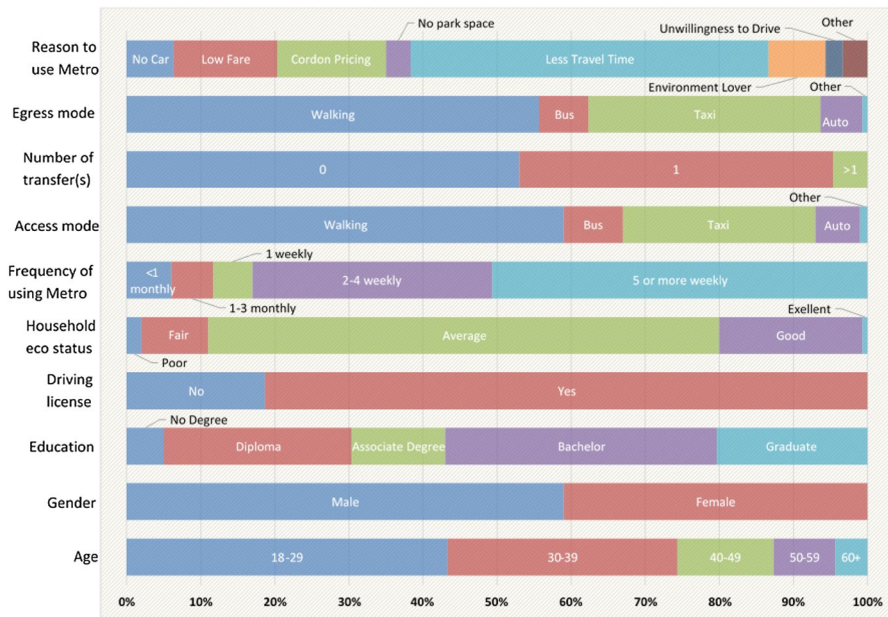


Fig. 3 Sample characteristics

## 5 Results

Three classes of model characteristics, totally include six distinct SEM structures as explained in the methodology (*S1–S6*), are applied in this study. Four levels of the transferability test, which includes structural and measurement equivalence tests, are investigated in this section. The measurement and structural models are evaluated by criteria defined for each. Factor loadings and reliability are checked to evaluate structures' validity and reliability. In the structural model evaluation, a number of goodness-of-fit indices are checked. These models are calibrated using AMOS software (Arbuckle and Wothke 1999).

### 5.1 Transferability at various levels

#### 5.1.1 Basic theory

All the structures have been developed to assess the relationship between PT service quality and customer satisfaction, and to find service quality attributes and their related weights (importance) in Tehran Metro line 3. Therefore, all the structures are theoretically equivalent.

### 5.1.2 Mathematical models

A Structural Equation Model is the common theme for modeling all the structures. Also, it is confirmed that SEM is an applicable methodology in this case study, since *S1* and *S2* have relatively acceptable goodness-of-fit indices in comparison with their original models (i.e., de Oña et al. (2013) and Eboli and Mazzulla (2015) studies) (see Table 4). Thus, all the structures are equivalent in terms of mathematical modeling.

### 5.1.3 Empirical structural characteristics

Confirmatory factor analysis was conducted in order to examine structural characteristics and cross-cultural equivalence of previous structures (*S1*, *S2*, *S3*, and *S4*).

In the first structure (*S1*), three exogenous latent variables were considered, *Service*, *Comfort*, and *Personnel*. There are eight factor loadings out of 22 which are less than 0.40, the cut-off suggested by Stevens (1992) and Hair et al. (2009), showing that 36% (>20%) of observed variables should be dropped from the model, which is not acceptable (Hair et al. 2009) (see Table 2). The Composite Reliability (CR) for each latent variable is greater than 0.6, the cut-off suggested by Bagozzi and Yi (1988); thus *S1* passes the reliability test (see Table 3).

In structure 2 (*S2*), six exogenous latent variables were used, *Safety*, *Cleanliness*, *Comfort*, *Service*, *Information*, and *Personnel*. The latent variable of *Personnel* has a CR less than 0.6 which shows a reliability problem and one factor loading is below the cut-off point of 0.40.

In structure 3 (*S3*), five exogenous latent variables were considered, *Tangibles*, *Reliability*, *Responsiveness*, *Assurance*, and *Empathy*. Latent variables of *Reliability* and *Responsiveness* have a reliability issue with a CR of 0.4 and 0.5, respectively. Also, five factor loading values are below the cut-off of 0.4.

Structure 4 (*S4*) has six exogenous latent variables, *Accessibility*, *Information*, *Time*, *Customer Care*, *Comfort*, and *Security*. Only three latent variables have a CR greater than 0.6 indicating a reliability issue in the structure. Factor loadings of three observed variables are less than 0.40.

Due to the drawbacks of these first four structures, it is indicated that the latent constructs of previous studies and theories are not perfectly transferable to this case study. Thus, an EFA is conducted based on eigenvalues greater than 1, with the maximum likelihood method, and the Promax method of rotation to investigate whether an improvement could be achieved. This constitutes Structure 5 (*S5*). It categorizes 19 service quality attributes (three attributes are dropped) into five factors, *Comfort*, *Security*, *Major Services*, *Cleanliness*, and *Personnel*. Factor loadings and construct reliability are tested by running a CFA which depicts that the latent factor of *Personnel* is not reliable (CR=0.5) and there are two factor loadings less than 0.40. These issues indicate that the structure may still be improved.

Thereafter, a constrained EFA with a fixed number of four factors is conducted for *S6*. It categorizes 21 service quality attributes (only one attribute is dropped)

**Table 2** Factor Loading of the Six Structures

	Service quality attributes						S1		S2		S3		S4		S5		S6	
	LV	FL <sup>a</sup>	LV	FL	LV	FL	LV	FL	LV	FL	LV	FL	LV	FL	LV	FL	LV	FL
<b>1</b>	Access time to station	Com <sup>b</sup>	0.33	<b>0.46</b>	Res	<b>0.43</b>	Acc	0.39	<b>0.43</b>	Maj	<b>0.46</b>	Maj	<b>0.43</b>	Maj	<b>0.46</b>	Maj	<b>0.43</b>	<b>0.43</b>
<b>2</b>	Ticket purchasing or recharging	Per	0.31	<b>0.48</b>	Tan	<b>0.40</b>	Acc	0.19	<b>0.40</b>	Per	<b>0.50</b>	Per	<b>0.50</b>	Per	<b>0.50</b>	Per	<b>0.50</b>	<b>0.50</b>
<b>3</b>	Ticket presenting	Serv	0.32	<b>0.45</b>	Res	<b>0.39</b>	Cus	0.35	<b>0.39</b>	Maj	<b>0.39</b>	Maj	<b>0.39</b>	Maj	<b>0.39</b>	Maj	<b>0.39</b>	0.38
<b>4</b>	Fare amount	Serv	0.37	<b>0.41</b>	Tan	<b>0.40</b>	Cus	0.27	<b>0.40</b>	Maj	<b>0.42</b>	Maj	<b>0.42</b>	Maj	<b>0.42</b>	Maj	<b>0.42</b>	<b>0.40</b>
<b>5</b>	Staff behavior	Per	0.36	<b>0.52</b>	Assu	<b>0.40</b>	Cus	<b>0.40</b>	<b>0.40</b>	Per	<b>0.55</b>	Per	<b>0.55</b>	Per	<b>0.55</b>	Per	<b>0.55</b>	<b>0.42</b>
<b>6</b>	Waiting time	Serv	<b>0.49</b>	<b>0.51</b>	Rel	<b>0.56</b>	T	<b>0.56</b>	<b>0.56</b>	Maj	<b>0.50</b>	Maj	<b>0.50</b>	Maj	<b>0.50</b>	Maj	<b>0.53</b>	<b>0.53</b>
<b>7</b>	Seat availability on board	Com	<b>0.68</b>	<b>0.71</b>	Tan	<b>0.71</b>	Com	<b>0.68</b>	<b>0.68</b>	Com	<b>0.71</b>	Com	<b>0.68</b>	Com	<b>0.71</b>	Com	<b>0.71</b>	<b>0.71</b>
<b>8</b>	Air conditioning on board	Com	<b>0.57</b>	<b>0.52</b>	Emp	<b>0.61</b>	Com	<b>0.61</b>	<b>0.61</b>	Com	<b>0.55</b>	Com	<b>0.55</b>	Com	<b>0.55</b>	Com	<b>0.56</b>	<b>0.56</b>
<b>9</b>	Travel time	Serv	<b>0.49</b>	<b>0.54</b>	Rel	<b>0.52</b>	T	<b>0.52</b>	<b>0.52</b>	Maj	<b>0.53</b>	Maj	<b>0.53</b>	Maj	<b>0.53</b>	Maj	<b>0.54</b>	<b>0.54</b>
<b>10</b>	Transfer time	Serv	0.24	0.33	Rel	0.28	T	0.28	0.28	Maj	0.33	Maj	0.33	Maj	0.33	Maj	0.34	0.34
<b>11</b>	Egress time to destination	Com	0.35	<b>0.43</b>	Res	<b>0.40</b>	Acc	<b>0.40</b>	<b>0.40</b>	Maj	<b>0.43</b>	Maj	<b>0.43</b>	Maj	<b>0.43</b>	Maj	<b>0.40</b>	<b>0.40</b>
<b>12</b>	Crowdedness on platform	Com	<b>0.65</b>	<b>0.73</b>	Tan	<b>0.74</b>	Com	<b>0.74</b>	<b>0.74</b>	Com	<b>0.68</b>	Com	<b>0.68</b>	Com	<b>0.71</b>	Com	<b>0.71</b>	<b>0.71</b>
<b>13</b>	Crowdedness on board	Com	<b>0.67</b>	<b>0.75</b>	Tan	<b>0.76</b>	Com	<b>0.76</b>	<b>0.76</b>	Com	<b>0.70</b>	Com	<b>0.70</b>	Com	<b>0.74</b>	Com	<b>0.74</b>	<b>0.74</b>
<b>14</b>	Information availability in station	Serv	<b>0.40</b>	<b>0.66</b>	Res	<b>0.50</b>	Inf	<b>0.50</b>	<b>0.50</b>	Com	<b>0.69</b>	Com	<b>0.69</b>	Com	<b>0.74</b>	Com	<b>0.74</b>	<b>0.74</b>
<b>15</b>	Information availability on board	Serv	0.37	<b>0.74</b>	Res	<b>0.47</b>	Inf	<b>0.47</b>	<b>0.47</b>	Com	<b>0.69</b>	Com	<b>0.69</b>	Com	<b>0.74</b>	Com	<b>0.74</b>	<b>0.74</b>
<b>16</b>	Safety in station	Per	<b>0.60</b>	<b>0.59</b>	Emp	<b>0.61</b>	Secu	<b>0.61</b>	<b>0.61</b>	Com	<b>0.59</b>	Com	<b>0.59</b>	Com	<b>0.57</b>	Com	<b>0.57</b>	<b>0.57</b>
<b>17</b>	Safety on board	Per	<b>0.52</b>	<b>0.50</b>	Emp	<b>0.58</b>	Secu	<b>0.58</b>	<b>0.58</b>	Com	<b>0.51</b>	Com	<b>0.51</b>	Com	<b>0.54</b>	Com	<b>0.54</b>	<b>0.54</b>
<b>18</b>	Security in station	Per	<b>0.81</b>	<b>0.83</b>	Assu	<b>0.75</b>	Secu	<b>0.75</b>	<b>0.75</b>	Com	<b>0.82</b>	Com	<b>0.82</b>	Com	<b>0.86</b>	Com	<b>0.86</b>	<b>0.86</b>
<b>19</b>	Security on board	Per	<b>0.76</b>	<b>0.77</b>	Assu	<b>0.72</b>	Secu	<b>0.72</b>	<b>0.72</b>	Com	<b>0.77</b>	Com	<b>0.77</b>	Com	<b>0.77</b>	Com	<b>0.77</b>	<b>0.78</b>
<b>20</b>	Cleanliness in station	Com	<b>0.44</b>	<b>0.79</b>	Assu	<b>0.50</b>	Com	<b>0.50</b>	<b>0.50</b>	Com	<b>0.41</b>	Com	<b>0.41</b>	Com	<b>0.75</b>	Com	<b>0.75</b>	<b>0.68</b>
<b>21</b>	Cleanliness on board	Com	<b>0.45</b>	<b>0.73</b>	Assu	<b>0.48</b>	Com	<b>0.48</b>	<b>0.48</b>	Com	<b>0.43</b>	Com	<b>0.43</b>	Com	<b>0.77</b>	Com	<b>0.77</b>	<b>0.69</b>
<b>22</b>	Ethical and behavioral messages	Serv	0.38	<b>0.44</b>	Emp	<b>0.35</b>	Inf	<b>0.35</b>	<b>0.35</b>	Com	<b>0.42</b>	Com	<b>0.42</b>	Com	<b>0.49</b>	Com	<b>0.49</b>	<b>0.49</b>

LV latent variable, FL factor loading

<sup>a</sup>Factor loadings equal or greater than 0.40 are shown in bold

<sup>b</sup>In each structure, the initials (bold font) are shown in the text before Table 1. (e.g., for S1: **Com**fort, **Per**sonnel, **Ser**vice)

**Table 3** Reliability of the six structures

S1	S2		S3		S4		S5		S6		
	Latent variable	CR	Latent variable	CR	Latent variable	CR	Latent variable	CR	Latent variable	CR	
Service	0.6	Safety	0.8	Tangibles	0.7	Accessibility	0.4	Comfort	0.8	Major Services	0.7
Comfort	0.7	Cleanliness	0.7	Reliability	0.4	Information	0.6	Security	0.8	Comfort	0.8
Personnel	0.7	Comfort	0.8	Responsiveness	0.5	Time	0.4	Major services	0.6	Security	0.8
		Service	0.6	Assurance	0.7	Customer care	0.3	Cleanliness	0.7	Minor services	0.6
		Information	0.6	Empathy	0.6	Comfort	0.7	Personnel	0.5		
		Personnel	0.5		0.8	Security					

CR construct reliability

into four factors, *Major Services*, *Comfort*, *Security*, and *Minor Services*. Results indicate that there is no construct reliability issue, but two factor loadings are less than the cut-off point. The results of measurement models of all six structures are presented in Tables 2 and 3.

Six different SEMs based on the proposed structures (*S1–S6*) are developed. A number of goodness-of-fit indices are used for comparing models and then choosing the best one. Chi squared (CMIN), degrees of freedom (DF), goodness-of-fit index (GFI), Tucker–Lewis index (TLI), comparative fit index (CFI), and root mean square error of approximation (RMSEA) are common fit indices used to investigate how well a model fits the sample data. Akaike information criterion (AIC) is a fit index for comparing models; the closer its value to zero, the better the model fits the data (Akaike 1987). Chi squared is exceedingly sensitive to sample size, so the Chi squared to degrees of freedom ratio (CMIN/DF) is used instead. Ratios less than 5 are indicative of an acceptable fit (the lower, the better) (Marsh and Hocevar 1985). The GFI, TLI, and CFI greater than 0.90, and the RMSEA less than 0.08 indicate a close fit of the model to the data (Browne and Cudeck 1993; Hair et al. 2009; Jöreskog and Sörbom 1984).

The values of these indices for the six models are shown in Table 4. Also, GFI, RMSEA, and CFI which are reported in the studies of de Oña et al. (2013) and Eboli and Mazzulla (2015) are listed in the table to facilitate comparing *S1* and *S2*, respectively, with their original models. Among the six structures, three structures (*S2*, *S5*, and *S6*) fit the data in a better way. Between these three, *S2* has a reliability issue mentioned above (i.e., having a latent variable with low CR). Comparing *S5* and *S6*, *S6* passes all of the criteria confirming that the modifications made to *S5* have remarkably enhanced the proposed structure.

#### 5.1.4 Estimated parameters

Testing transferability of parameters (i.e., measurement equivalence) requires access to previous studies' data. For *S3* and *S4*, as they are developed based on a general theory, there is no previous data to make this level of transferability test happen. Also, there are some mode-specific and case-specific attributes in de Oña et al. (2013) and Eboli and Mazzulla (2015) studies which make it impractical to test the transferability of the structures in this level. Generally, as the previously studied structures (*S1–S4*) do not pass the prior level of the transferability test, leaving the test in the previous level is inevitable.

## 5.2 Results of the proposed model (*S6*)

From the analysis and comparisons made in previous sections, it is seen that *S6* is the best model for this study's data. This final model consists of four latent variables defining the second-order latent variable of *Service Quality* namely *Major Services*, *Comfort*, *Security*, and *Minor Services*. *Major Services* and *Comfort* have the



**Table 4** Fit Indices for Proposed Models

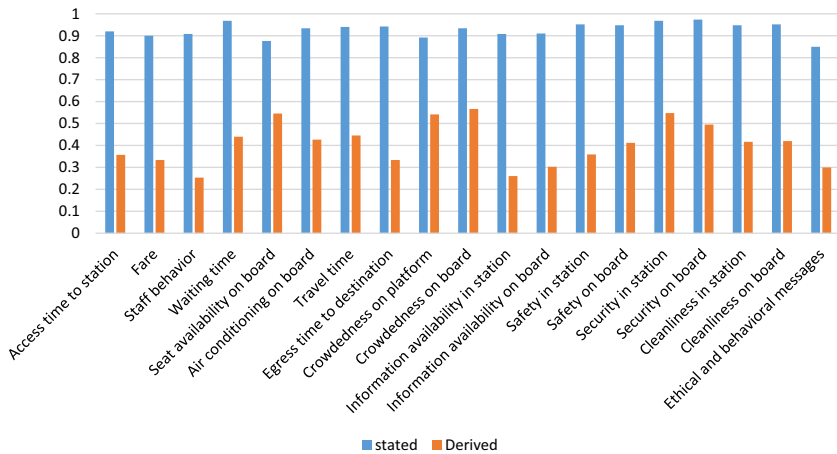
Fit indices	de Oña et al. (2013)	Eboli and Mazzulla (2015)	S1	S2	S3	S4	S5	S6
Chi squared (CMIN)			925.71	760.90	996.09	918.63	590.17	22.91
CMIN/DF			3.70	3.09	4.03	3.73	3.21	2.86
Absolute fit indices								
Goodness of fit index (GFI)	0.86	0.65	0.78	0.80	0.75	0.76	0.82	0.98
Root mean square error of approximation (RMSEA)	0.12	0.10	0.10	0.08	0.10	0.10	0.09	0.08
Incremental fit indices								
Comparative fit index (CFI)	0.71	0.79	0.61	0.71	0.57	0.62	0.74	0.98
Tucker–Lewis index (TLI)			0.57	0.67	0.52	0.57	0.70	0.97
Parsimony fit index								
Akaike information criterion (AIC)			1025.71	868.90	1102.09	1026.63	684.17	48.91

**Table 5** Estimated weights of model (S6)

Variables		Non-st. W	S.E.	C.R.	P	St. W
Independent observed	Exogenous latent					
Egress time to destination	Major services	0.95	0.21	4.62	***	0.40
Travel time		1.20	0.22	5.37	***	0.54
Waiting time		1.47	0.28	5.34	***	0.53
Fare amount		0.96	0.21	4.61	***	0.40
Access time to station		1.00				0.43
Crowdedness on board	Comfort	1.61	0.19	8.72	***	0.74
Crowdedness on platform		1.41	0.17	8.52	***	0.71
Seat availability on board		1.59	0.19	8.55	***	0.71
Safety on board		0.97	0.14	7.15	***	0.54
Air conditioning on board		1.00				0.55
Security in station	Security	1.62	0.17	9.32	***	0.86
Security on board		1.66	0.18	9.21	***	0.78
Safety in station		1.00				0.57
Cleanliness on board	Minor services	1.22	0.20	5.99	***	0.69
Cleanliness in station		1.17	0.19	5.98	***	0.68
Information availability on board		1.23	0.23	5.28	***	0.50
Information availability in station		1.15	0.24	4.88	***	0.43
Ethical and behavioral messages		1.31	0.25	5.25	***	0.49
Staff behavior		1.00				0.42
<i>First-order latent</i>	<i>Second-order latent</i>					
Major Services	Service Quality	1.46	0.10	15.00	***	0.94
Comfort		1.97	0.14	14.19	***	0.88
Security		1.48	0.13	11.86	***	0.72
Minor Services		1.00				0.70
<i>Endogenous latent</i>	<i>Exogenous latent</i>					
Customer Satisfaction	Service Quality	1.22	0.23	5.34	***	0.88
<i>Dependent observed</i>	<i>Endogenous latent</i>					
Current trip satisfaction	Customer satisfaction	2.27	0.40	5.63	***	0.88
All trips satisfaction		1.00				0.37

\*\*\*The weight is significantly different from zero at the 0.001 level (two-tailed)

largest effect on service quality (standardized weights of 0.94 and 0.88, respectively) (Table 5). The latent variable of *Customer Satisfaction* consists of two observed variables, *All trips satisfaction* and *Current trip satisfaction*, and is chiefly reflected by the latter (standardized weight of 0.88) which is an important finding here showing that perceptions about the trip that passengers are doing at the moment better describe their satisfaction level. Table 5 includes non-standardized weights (Non-st. W), standard errors (S.E.), critical ratios (C.R. = Non-st. W/S.E.), significance test of weights (P), and standardized weights (St. W) of the final model.



**Fig. 4** Comparison of stated and derived importance

The total effect of each service quality attribute on customer satisfaction is calculated. These are the values which can be defined as the derived importance of service quality attributes. Comparison between derived importance and perceived importance as stated by passengers in the survey shows significant differences. Among the top 3 important attributes ranked based on these two methods, there is only one shared attribute (*Security in station*). The comparison of attributes based on derived and stated importance is shown in Fig. 4. It provides another evidence that the derived and stated importance values can be very different.

## 6 Conclusions

This study presents experimental results on developing a customer satisfaction model for an urban public transport case study in a developing country. In addition, this study focuses on two topics: (1) examining various levels of model transferability i.e. from previous studies or fundamental theories into locally collected data, and (2) investigating the impact of service quality attributes on overall customer satisfaction. A comprehensive literature review on previous studies indicates that SEM is a strong method to be used for exploring a latent structure that can define service quality. Although, a new structure may be statistically better than borrowing it from the previous studies, it is worthwhile to examine previous structures if there is a solid reason to argue for them. Accordingly, it is important to confirm the transferability of previous models when they are used in any case study apart from their origin. In order to evaluate the customer satisfaction in the case study of Tehran Metro, a number of credible proposed structures and theories have been taken into account. However, findings of this paper demonstrate that previously proposed SEM structures show a weaker fit when applied to the present case study compared to an SEM structure that is developed based on an exploratory analysis. The main idea of the

transferability test in this paper is on structural characteristics, which is suggested to be checked before a transferability test of estimated parameters.

The best structure of the latent variables is found after analyzing six distinct structures in terms of different levels of transferability test. The final model identifies four latent variables to describe service quality: *Major Services* (i.e., main attributes related to the performance of service), *Comfort* (i.e., attributes associated with feeling comfortable during a trip), *Security* (i.e., attributes associated with feeling safe and secure), and *Minor Services* (i.e., attributes generally related to staff performance). Among them, *Major Services* is the most influential latent variable on service quality, while *Minor Services* is the least influential latent variable.

Furthermore, this paper determines that the observed variable of “Current trip satisfaction” is a better indicator of the latent variable of Customer Satisfaction compared to “All trips satisfaction”. Therefore, on-board surveys should be given more attention, in order to obtain accurate perceptions about the trip passengers are doing at that moment, for better evaluation of service quality and customer satisfaction.

With the determination of the importance of service quality attributes, this paper confirms previous findings about the difference between stated and derived importance, and also helps PT operators and managers to justify and plan for PT improvement with a prioritization on how to invest the limited resources. In fact, attributes can be divided into four quadrants and an importance-performance analysis (IPA) can be undertaken to help operators better understand each attribute’s contribution to customer satisfaction. For instance, in the case of Tehran, attributes associated with train frequency (e.g., crowdedness, security, seat availability) require the highest attention. Allocating limited resources more efficiently can improve PT services, which can retain current users while encouraging new ones to use PT.

For future research, segmentation of trips into several sections (e.g., access, waiting, in-vehicle, transfer, and egress) and evaluation of service quality for each section separately could be considered. Also, other PT modes of the case study area could be considered for a more comprehensive evaluation with a larger sample size.

## References

- Abenoza RF, Cats O, Susilo YO (2017) Travel satisfaction with public transport: determinants, user classes, regional disparities and their evolution. *Transp Res Part A Policy Pract* 95:64–84. <https://doi.org/10.1016/j.tra.2016.11.011>
- Agyemang-Duah K, Hall FL (1997) Spatial transferability of an ordered response model of trip generation. *Transp Res Part A Policy Pract* 31:389–402. [https://doi.org/10.1016/S0965-8564\(96\)00035-3](https://doi.org/10.1016/S0965-8564(96)00035-3)
- Akaike H (1987) Factor analysis and AIC. *Psychometrika* 52:317–332. [https://doi.org/10.1007/978-1-4612-1694-0\\_29](https://doi.org/10.1007/978-1-4612-1694-0_29)
- Arbuckle JL, Wothke W (1999) Amos 4.0 user’s guide. SmallWaters Corporation, Chicago
- Atherton TJ, Ben-Akiva M (1976) Transferability and updating of disaggregate travel demand models. In: Paper presented at the Transportation Research Board 55th annual meeting, Washington D.C., January 1976
- Atlas of Tehran metropolis (2011) Municipality of Tehran. <http://www.irancarto.cnrs.fr>. Accessed 05 June 2018

- Aydin N, Celik E, Gumus AT (2015) A hierarchical customer satisfaction framework for evaluating rail transit systems of Istanbul. *Transp Res Part A Policy Pract* 77:61–81. <https://doi.org/10.1016/j.tra.2015.03.029>
- Bagozzi RP, Yi Y (1988) On the evaluation of structural equation models. *J Acad Mark Sci* 16:74–94. <https://doi.org/10.1007/BF02723327>
- Ben-Akiva M, Bolduc D (1987) Approaches to model transferability and updating: the combined transfer estimator. *Transp Res Rec J Transp Res Board* 1139:1–7
- Berry JW, Berry JW, Poortinga YH, Segall MH, Dasen PR (2002) *Cross-cultural psychology: research and applications*. Cambridge University Press, New York
- Boisjoly G, Grisé E, Maguire M, Veillette M-P, Deboosere R, Berrebi E, El-Geneidy A (2018) Invest in the ride: a 14 year longitudinal analysis of the determinants of public transport ridership in 25 North American cities. *Transp Res Part A Policy Pract* 116:434–445. <https://doi.org/10.1016/j.tra.2018.07.005>
- Bollen KA (1989) A new incremental fit index for general structural equation models. *Sociol Methods Res* 17:303–316. <https://doi.org/10.1177/0049124189017003004>
- Browne MW, Cudeck R (1993) Alternative ways of assessing model fit. In: Bollen KA, Long JS (eds) *Testing structural equation models*. Sage focus editions, Sage, pp 136–154
- Byrne BM (2010) *Testing for the factorial equivalence of instrument scores*. Structural equation modeling with AMOS: basic concepts, applications, and programming, 2nd edn. Routledge, London, pp 197–230
- Byrne BM (2015) Testing across nations and cultures: issues and complexities. *Int J Bus Glob* 14:170–186
- Byrne BM (2016) *Structural equation modeling with AMOS: basic concepts, applications, and programming*, 3rd edn. Routledge, New York
- Byrne BM, Van de Vijver FJ (2010) Testing for measurement and structural equivalence in large-scale cross-cultural studies: addressing the issue of nonequivalence. *Int J Test* 10:107–132. <https://doi.org/10.1080/15305051003637306>
- Cao J, Cao X (2017) Comparing importance-performance analysis and three-factor theory in assessing rider satisfaction with transit. *J Transp Land Use* 10:65–82
- Celik E, Aydin N, Gumus AT (2014) A multiattribute customer satisfaction evaluation approach for rail transit network: a real case study for Istanbul, Turkey. *Transp Policy* 36:283–293. <https://doi.org/10.1016/j.tranpol.2014.09.005>
- CEN (2002) EN 13816—transportation—logistics and services—public passenger transport—service quality definition, targeting and measurement, Bratislava
- Chen C-F (2008) Investigating structural relationships between service quality, perceived value, satisfaction, and behavioral intentions for air passengers: evidence from Taiwan. *Transp Res Part A Policy Pract* 42:709–717. <https://doi.org/10.1016/j.tra.2008.01.007>
- Cheung GW, Rensvold RB (2002) Evaluating goodness-of-fit indexes for testing measurement invariance. *Struct Equ Model* 9:233–255. [https://doi.org/10.1207/S15328007SEM0902\\_5](https://doi.org/10.1207/S15328007SEM0902_5)
- Chowdhury S, Hadas Y, Gonzalez VA, Schot B (2018) Public transport users' and policy makers' perceptions of integrated public transport systems. *Transp Policy* 61:75–83. <https://doi.org/10.1016/j.tranpol.2017.10.001>
- de Oña J, de Oña R, Eboli L, Mazzulla G (2013) Perceived service quality in bus transit service: a structural equation approach. *Transp Policy* 29:219–226. <https://doi.org/10.1016/j.tranpol.2013.07.001>
- de Oña R, Machado JL, de Oña J (2015) Perceived service quality, customer satisfaction, and behavioral intentions: structural equation model for the Metro of Seville, Spain. *Transp Res Rec J Transp Res Board* 2538:76–85. <https://doi.org/10.3141/2538-09>
- Dell'Olio L, Ibeas A, Cecín P (2010) Modelling user perception of bus transit quality. *Transp Policy* 17:388–397. <https://doi.org/10.1016/j.tranpol.2010.04.006>
- Duarte A, Garcia C, Giannarakis G, Limão S, Polydoropoulou A, Litinas N (2010) New approaches in transportation planning: happiness and transport economics. *NETNOMICS Econ Res Electron Netw* 11:5–32. <https://doi.org/10.1007/s11066-009-9037-2>
- Eboli L, Mazzulla G (2007) Service quality attributes affecting customer satisfaction for bus transit. *J Public Transp* 10:21–34. <https://doi.org/10.5038/2375-0901.10.3.2>
- Eboli L, Mazzulla G (2012) Structural equation modelling for analysing passengers' perceptions about railway services. *Procedia Soc Behav Sci* 54:96–106. <https://doi.org/10.1016/j.sbspro.2012.09.729>

- Eboli L, Mazzulla G (2015) Relationships between rail passengers' satisfaction and service quality: a framework for identifying key service factors. *Public Transp* 7:185–201. <https://doi.org/10.1007/s12469-014-0096-x>
- Eboli L, Forciniti C, Mazzulla G (2018) Formative and reflective measurement models for analysing transit service quality. *Public Transp* 10:107–127. <https://doi.org/10.1007/s12469-017-0168-9>
- Figler SA, Sriraj P, Welch EW, Yavuz N (2011) Customer loyalty and Chicago, Illinois, transit authority buses: results from 2008 customer satisfaction survey. *Transp Res Rec J Transp Res Board* 2216:148–156. <https://doi.org/10.3141/2216-16>
- Golob TF (2003) Structural equation modeling for travel behavior research. *Transp Res Part B Methodol* 37:1–25. [https://doi.org/10.1016/S0191-2615\(01\)00046-7](https://doi.org/10.1016/S0191-2615(01)00046-7)
- Grisé E, El-Geneidy A (2017) Evaluating the relationship between socially (dis) advantaged neighbourhoods and customer satisfaction of bus service in London, UK. *J Transp Geogr* 58:166–175. <https://doi.org/10.1016/j.jtrangeo.2016.11.016>
- Habib KMN, Kattan L, Islam MT (2011) Model of personal attitudes towards transit service quality. *J Adv Transp* 45:271–285. <https://doi.org/10.1002/atr.106>
- Habibian M, Rezaei A (2017) Accounting for systematic heterogeneity across car commuters in response to multiple TDM policies: case study of Tehran. *Transportation* 44:681–700. <https://doi.org/10.1007/s11116-015-9672-4>
- Hadiuzzaman M, Das T, Hasnat MM, Hossain S, Rafee Musabbir S (2017) Structural equation modeling of user satisfaction of bus transit service quality based on stated preferences and latent variables. *Transp Plan Technol* 40:257–277. <https://doi.org/10.1080/03081060.2017.1283155>
- Haghighi NN, Liu XC, Wei R, Li W, Shao H (2018) Using Twitter data for transit performance assessment: a framework for evaluating transit riders' opinions about quality of service. *Public Transp* 10:363–377. <https://doi.org/10.1007/s12469-018-0184-4>
- Hair JF, Black WC, Babin BJ, Anderson RE (2009) *Multivariate data analysis*, 7th edn. Pearson Prentice Hall, Englewood Cliffs
- Hansen S (1981) Pro cross-cultural transferability. In: *New horizons in behavioural travel research*. Lexington Books, Lexington
- Hensher DA, Prioni P (2002) A service quality index for area-wide contract performance assessment. *J Transp Econ Policy (JTEP)* 36:93–113
- Hensher DA, Mulley C, Yahya N (2010) Passenger experience with quality-enhanced bus service: the tyne and wear 'superoute' services. *Transportation* 37:239–256. <https://doi.org/10.1007/s11116-009-9240-x>
- Hui CH, Triandis HC (1985) Measurement in cross-cultural psychology: a review and comparison of strategies. *J Cross-cult Psychol* 16:131–152. <https://doi.org/10.1177/0022002185016002001>
- Imaz A, Habib KMN, Shalaby A, Idris AO (2015) Investigating the factors affecting transit user loyalty. *Public Transp* 7:39–60. <https://doi.org/10.1007/s12469-014-0088-x>
- Ingvardson JB, Nielsen OA (2019) The relationship between norms, satisfaction and public transport use: a comparison across six European cities using structural equation modelling. *Transp Res Part A Policy Pract* 126:37–57. <https://doi.org/10.1016/j.tra.2019.05.016>
- Irfan SM, Kee DMH, Shahbaz S (2012) Service quality and rail transport in Pakistan: a passenger perspective. *World Appl Sci J* 18:361–369
- Javid MA, Okamura T, Nakamura F, Tanaka S, Wang R (2016) People's behavioral intentions towards public transport in Lahore: role of situational constraints, mobility restrictions and incentives. *KSCE J Civ Eng* 20:401–410. <https://doi.org/10.1007/s12205-015-1123-4>
- Joewono TB, Kubota H (2007) User satisfaction with paratransit in competition with motorization in Indonesia: anticipation of future implications. *Transportation* 34:337–354. <https://doi.org/10.1007/s11116-007-9119-7>
- Joewono TB, San Santoso D, Ningtyas DU (2012) The causal relationship of the service quality of the TransJakarta Busway. *Public Transp* 4:77–100. <https://doi.org/10.1007/s12469-012-0053-5>
- Jöreskog KG (1970) A general method for estimating a linear structural equation system. *ETS Res Rep Ser*. <https://doi.org/10.1002/j.2333-8504.1970.tb00783.x>
- Jöreskog KG, Sörbom D (1984) *LISREL-VI user's guide*, 3rd edn, Mooresville
- Karlaftis MG, Golias J, Papadimitriou E (2001) Transit quality as an integrated traffic management strategy: measuring perceived service. *J Public Transp* 4:27–44. <https://doi.org/10.5038/2375-0901.4.1.2>
- Koppelman FS, Wilmot CG (1982) Transferability analysis of disaggregate choice models. *Transp Res Rec J Transp Res Board* 895:18–24

- Lai W-T, Chen C-F (2011) Behavioral intentions of public transit passengers—the roles of service quality, perceived value, satisfaction and involvement. *Transp Policy* 18:318–325. <https://doi.org/10.1016/j.tranpol.2010.09.003>
- Likert R (1932) A technique for the measurement of attitudes. *Arch Psychol* 22:1–55
- Ma Z, Ferreira L, Mesbah M (2014) Measuring service reliability using automatic vehicle location data. *Math Probl Eng* 2014:1–12. <https://doi.org/10.1155/2014/468563>
- Machado-León JL, de Oña R, Baouni T, de Oña J (2017) Railway transit services in Algiers: priority improvement actions based on users perceptions. *Transp Policy* 53:175–185. <https://doi.org/10.1016/j.tranpol.2016.10.004>
- Malhotra NK, Ulgado FM, Agarwal J, Baalbaki IB (1994) International services marketing: a comparative evaluation of the dimensions of service quality between developed and developing countries. *Int Mark Rev* 11:5–15. <https://doi.org/10.1108/02651339410061937>
- Marsh HW, Hocevar D (1985) Application of confirmatory factor analysis to the study of self-concept: first-and higher order factor models and their invariance across groups. *Psychol Bull* 97:562–582. <https://doi.org/10.1037/0033-2909.97.3.562>
- Mesbah M, Lin J, Currie G (2015) “Weather” transit is reliable? Using AVL data to explore tram performance in Melbourne, Australia. *J Traffic Transp Eng (Engl Ed)* 2:125–135. <https://doi.org/10.1016/j.jtte.2015.03.001>
- Minsler J, Webb V (2010) Quantifying the benefits: application of customer loyalty modeling in public transportation context. *Transp Res Rec J Transp Res Board* 2144:111–120. <https://doi.org/10.3141/2F2144-13>
- Morfoulaki M, Myrovali G, Kotoula K (2015) Increasing the attractiveness of public transport by investing in soft ICT based measures: going from words to actions under an austerity backdrop—Thessaloniki’s case, Greece. *Res Transp Econ* 51:40–48. <https://doi.org/10.1016/j.retrec.2015.07.006>
- Mouwen A, Rietveld P (2013) Does competitive tendering improve customer satisfaction with public transport? A case study for the Netherlands. *Transp Res Part A Policy Pract* 51:29–45. <https://doi.org/10.1016/j.tra.2013.03.002>
- Municipality Tehran (2016) An overview to Tehran Traffic Data. Deputy of Traffic and Transportation, Tehran
- Nathanail E (2008) Measuring the quality of service for passengers on the Hellenic railways. *Transp Res Part A Policy Pract* 42:48–66. <https://doi.org/10.1016/j.tra.2007.06.006>
- Oliver RL (2015) *Satisfaction: a behavioral perspective on the consumer*, 2nd edn. Routledge, New York
- Parasuraman A, Zeithaml VA, Berry LL (1985) A conceptual model of service quality and its implications for future research. *J Mark* 49:41–50. <https://doi.org/10.1177/002224298504900403>
- Parasuraman A, Zeithaml VA, Berry LL (1988) Servqual: a multiple-item scale for measuring consumer perception. *J Retail* 64:12–37
- Rahman F, Das T, Hadiuzzaman M, Hossain S (2016) Perceived service quality of paratransit in developing countries: a structural equation approach. *Transp Res Part A Policy Pract* 93:23–38. <https://doi.org/10.1016/j.tra.2016.08.008>
- Shen W, Xiao W, Wang X (2016) Passenger satisfaction evaluation model for Urban rail transit: a structural equation modeling based on partial least squares. *Transp Policy* 46:20–31. <https://doi.org/10.1016/j.tranpol.2015.10.006>
- Sikder S, Pinjari AR, Srinivasan S, Nowrouzian R (2013) Spatial transferability of travel forecasting models: a review and synthesis. *Int J Adv Eng Sci Appl Math* 5:104–128. <https://doi.org/10.1007/s12572-013-0090-6>
- Smith AM, Reynolds NL (2002) Measuring cross-cultural service quality: a framework for assessment. *Int Mark Rev* 19:450–481. <https://doi.org/10.1108/02651330210445>
- Soltanpour A, Mesbah M, Habibian M (2018) Perceived service quality in urban rail transit: a comparison of structural equation models. Paper presented at the Transportation Research Board 97th annual meeting, Washington D.C., January 2018
- Stevens J (1992) *Applied multivariate statistics for the social sciences*. Lawrence Erlbaum, Hillsdale
- Stuart K, Mednick M, Bockman J (2000) Structural equation model of customer satisfaction for the New York City subway system. *Transp Res Rec J Transp Res Board* 1735:133–137. <https://doi.org/10.3141/1735-16>
- TCRP report 165 (2013) *Transit capacity and quality of service manual (TCQSM)*, 3rd edn. Transportation Research Board, Washington, D.C

- Tehran Traffic and Transportation Organization (2013) An overview of Tehran Transportation Master Plan. Deputy of planning and studies, Tehran
- Trompet M, Parasram R, Anderson R (2013) Benchmarking disaggregate customer satisfaction scores of bus operators in different cities and countries. *Transp Res Rec J Transp Res Board* 2351:14–22. <https://doi.org/10.3141/2351-02>
- Tyrinopoulos Y, Antoniou C (2008) Public transit user satisfaction: variability and policy implications. *Transp Policy* 15:260–272. <https://doi.org/10.1016/j.tranpol.2008.06.002>
- Van de Vijver FJ, Leung K (1997) *Methods and data analysis for cross-cultural research*, vol 1. Sage, Newbury Park
- Vitale A, Guido G, Rogano D (2016) A smartphone based DSS platform for assessing transit service attributes. *Public Transp* 8:315–340. <https://doi.org/10.1007/s12469-016-0133-z>
- Wan D, Kamga C, Hao W, Sugiura A, Beaton EB (2016) Customer satisfaction with bus rapid transit: a study of New York City select bus service applying structural equation modeling. *Public Transp* 8:497–520. <https://doi.org/10.1007/s12469-016-0135-x>
- Weinstein A (2000) Customer satisfaction among transit riders: how customers rank the relative importance of various service attributes. *Transp Res Rec J Transp Res Board* 1735:123–132. <https://doi.org/10.3141/1735-15>
- Wiley DE (1973) The identification problem for structural equation models with unmeasured variables. *Structural equation models in the social sciences*. Seminar Press, New York, pp 69–83

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Affiliations

Amirali Soltanpour<sup>1</sup> · Mahmoud Mesbah<sup>2,3</sup>  · Meeghat Habibian<sup>2</sup>

✉ Mahmoud Mesbah  
mmesbah@aut.ac.ir

Amirali Soltanpour  
soltanpo@msu.edu

Meeghat Habibian  
habibian@aut.ac.ir

<sup>1</sup> Department of Civil and Environmental Engineering, Michigan State University, East Lansing, MI 48824, USA

<sup>2</sup> Department of Civil and Environmental Engineering, Amirkabir University of Technology, Hafez St., Tehran 15875-4413, Iran

<sup>3</sup> School of Civil Engineering, The University of Queensland, Brisbane, QLD 4072, Australia