ARCHIVES OF TRANSPORT
Volume 65, Issue 1, 2023

ISSN (print): *0866-9546*
e-ISSN (online): *2300-8830*
DOI: *10.5604/01.3001.0016.2477*

HOW TRANSIT IMPROVEMENTS ARE PERCEIVED BY PASSENGERS? RESULTS OF A BEFORE-AFTER CUSTOMER SATISFACTION SURVEY

Mina SAHRAEI[1], Mahmoud MESBAH[2 3], Meeghat HABIBIAN[4], Tara SAEIDI[5], Amirali SOLTANPOUR[6 7]

[1, 2, 4, 5, 6] *Department of Civil and Environmental Engineering, Amirkabir University of Technology, Tehran, Iran*

[3] *School of Civil Engineering, The University of Queensland, Brisbane, Australia*

[7] *Department of Civil and Environmental Engineering, Michigan State University, East Lansing, USA*

Abstract:

Although customer satisfaction surveys are widely utilized by transit agencies, there are limited analyses in the literature on the perception of passengers as a result of service improvements. A before-after study can help to evaluate the effect of changes from customer's points of view and thus guarantee a continuous improvement in the service. In this paper, customer satisfaction was directly observed through a Customer Satisfaction Survey (CSS) before and after certain changes. Furthermore, Structural Equation Modeling (SEM) is utilized to evaluate passenger's perception of the service attribute importance. Finally, an Importance-Performance Analysis (IPA) is adapted to analyze the changes in satisfaction and importance from the passenger's perspective on each service attribute. In both before and after cases, a consistent SEM structure is used. The follow-up IPA provides transit agencies with priorities to improve service attributes and helps managers to devote their resources to key attributes that matter to the riders. Metro line 3 in Tehran was selected as the case study which is 33.7 km long with 25 stations. Two surveys were performed one before (with the sample size of 300), and one after (with the sample size of 384) a set of changes the most important of which was a headway reduction. The SEM was developed with five latent variables of main service, comfort, information, protection, and physical appearance. This structure was assessed on both the before and after data collections and showed to be valid. Security at the station and security on board were the most important service attributes in both waves according to their factor loadings, while ethics and behavioral messages had the smallest factor loading and the least importance. Comparing the attributes in both surveys suggested that reducing the headway was effective, although it did not seem to be sufficient for enhancing the overall customer satisfaction and improvements need to be continued.

Keywords: *customer satisfaction, structural equation modelling (SEM), importance-performance analysis (IPA), before-after study*

To cite this article:

Sahraei, M., Mesbah, M., Habibian, M., Saeidi, T., Soltanpour, A., (2023). How transit improvements are perceived by passengers? Results of a before-after customer satisfaction survey. Archives of Transport, 65(1), 53-66. DOI: https://doi.org/10.5604/01.3001.0016.2477

Contact:
1) minasahraei@aut.ac.ir [https://orcid.org/0000-0001-5615-041X]; 2) mmesbah@aut.ac.ir [http://orcid.org/0000-0002-3344-1350] – corresponding author; 3) habibian@aut.ac.ir [https://orcid.org/0000-0002-4130-9042]; 4) tara.saeidi1374@gmail.com [https://orcid.org/0000-0002-5787-6631]; 5) soltanpo@msu.edu [https://orcid.org/0000-0002-6648-8956]

1. Introduction

Increasing travel demand in urban areas due to an increase in people activities can influence the use of public transit and its level of service (Gärling et al., 2002). Public transit can be adopted as an effective solution to tackle urban problems, such as air pollution and traffic congestion and help to achieve sustainable development (Hu et al., 2022). Therefore, public transit managers need to better understand rider needs and expectations. A higher satisfaction of service quality in public transit can encourage current passengers to use public transit continuously, and attract new passengers to the system by improving the public image (Morpace International, 1999) which is considered in many studies (Zehmed and Jawab, 2021, Sukwadi et al., 2021, Harreman-Fernandes et al., 2021, Park et al., 2021). Transit agencies need to identify passenger's preferences to understand how to prioritize their investments and attract more users (Abenoza et al., 2017). Keeping customers satisfied will result in maintaining current travelers, attracting new users, and improving the public image of transit (Morpace International, 1999). Furthermore, understanding users' opinions about service quality can lead to an understanding of areas in need of improvement.

There are several approaches to evaluate the importance of each service attribute from the customer's perspective. Two major methods are 1) asking users to rate the importance of attributes and 2) deriving the importance by statistical models (Eboli and Mazzulla, 2015). In the former method, one passenger might declare all attributes as highly important which does not differentiate attributes; thus, the latter method is preferred (Dell'Olio et al., 2010; Soltanpour et al., 2020). Structural Equation Modeling (SEM) is an advanced technique for derived importance estimation, which is appropriate for describing latent variables of service quality (Eboli and Mazzulla, 2015). Latent variables help to better define the relationships among service attributes and the overall service quality (Eboli and Mazzulla, 2015). A Customer Satisfaction Survey (CSS) is used to collect customer's perspective about service attributes, it can be run regularly (e.g. annually or every three months (Ryus et al., 2003)) in order to monitor service (de Oña et al., 2014).

Assessing regular and periodic data helps monitoring the status quo and improvements in service quality. It ascertains that the operator strategies met the goals or there is a need for further improvements (de Oña et al., 2014). Monitoring service quality in public transit is performed in two ways: (1) using "aggregated indices", or (2) assessing and comparing each service attribute (de Oña et al., 2014). Although an aggregated analysis is simple, evaluating individual attributes of a service can identify the impact of any minor change on service. Comparison between attributes helps to concentrate on improvement priorities. However, implementing both approaches is ideal for a before-after study (de Oña et al., 2014).

Importance-Performance Analysis (IPA) is a technique that suggests management strategies (Martilla and James, 1977). The objective of an IPA is to identify the strengths and weaknesses of service attributes, to diagnose the priorities, and to suggest possible improvements that should be focused (Abalo et al., 2007); therefore, an IPA is an effective tool to find priorities and improvements in a before-after study. The simplicity and effectiveness of IPA have turned it into a widely used method (Wu and Jason, 2017).

This paper inspects the effects of some major changes in a metro service on the customer satisfaction. Two surveys have been implemented: one before the changes (Customer Satisfaction Survey 1, or CSS1), and one after the changes (Customer Satisfaction Survey 2, or CSS2). The changes included opening a new station, reducing the headway (i.e., it has been halved from 18 minutes to 9 minutes). It should be noted that sometime between the two surveys, the fare was increased by 30%, but this increase was very marginal since the precedent fare was very cheap (Statistical Report of Tehran (Municipality, 2017) notes that the cost of metro ticket for a trip was about 0.7% of the average daily income). An IPA is used to demonstrate how changes affect users' points of view.

This paper is structured as follows. In section 2, the literature review is provided. The methodological framework and concepts are explained in section 3. In section 4, the case study and the dataset are presented in greater details. Results of the SEM and the IPA for both surveys are described in section 5. Finally, section 6 summarizes the primary findings.

2. Literature review

Enhancing service quality in public transit is one of the most effective strategies to attract people from private modes, especially in metropolitan areas.

Monitoring continuous improvements of transit service through analyzing periodic data helps planners and authorities to develop appropriate policies (de Oña et al., 2014). According to Friman and Fellesson (2004), on passengers' behavior, transit users typically retain negative experiences more than positives in their minds. Transit authorities are willing to promote a positive image of the services by understanding how to increase rider's satisfaction. If passengers become more satisfied, they will continue using public transit, and will recommend it to other people. There are suitable tools to measure and monitor the service quality of public transit systems, and various methods have been introduced to estimate the relative importance of each service attribute on the satisfaction. Eboli and Mazzulla (2011), among others, measured the relationships between transit service quality and passengers' perception.

Increasing satisfaction leads to a boost in passenger retention. Passenger's satisfaction can be measured by exploiting the data collected in a Customer Satisfaction Survey (CSS) (Abenoza et al., 2017). de Oña and de Oña (2014) classified customer satisfaction analysis in two categories: aggregated models (making an overall service quality index from aggregating individual attributes) and disaggregated models (measuring service attributes individually). IPA is a disaggregated model represented by Martilla and James (1977), in which attributes are divided into four quadrants based on their performance and importance. Shen et al., (2016) applied IPA for determining attributes priorities, in which the indicators of performance were satisfaction rates, and the importance indicators were obtained from SEM. Furthermore, IPA is a widely used technique, specifically, to find priorities for service improvements (Putra et al., 2014; Machado-León et al., 2017; Wu and Jason, 2017; Echaniz et al., 2019; Yuan et al., 2019).

There are two main approaches to evaluate the importance of service attributes. One way is obtaining the stated importance, in which travelers are directly asked to rate the importance of attributes in a CSS. The second method is based on derived importance that is obtained by statistical models and assess the strength of individual attributes on the overall satisfaction. Examples are regression analysis, path analysis, and SEM (Eboli and Mazzulla, 2015). SEM is used as a general statistical modeling technique and is a combination of factor analysis and path analysis or regression, in the process of which latent variables are constructed (Shen et al., 2016). SEM was developed by Wiley (1973) and Jöreskog (1970). It is known as a large sample technique usually more than 200 (Lei and Qiong, 2007) and it is applied to various fields (Madha et al., 2016, Merlin et al.,2021, Halakoo et al., 2022). Eboli and Mazzulla (2007) used SEM to obtain the importance of each service attribute and identified key factors of service for customer satisfaction in transit. They considered latent variables that were measured by 16 service attributes and satisfaction was measured by two indicators of "perception" and "expectation". Also, de Oña et al. (2013) described latent variables in service and its relationship with overall service quality. Soltanpour et al., (2020) and Soltanpour et al., (2018) studied customer satisfaction of Tehran metro line 3 and confirmed that a latent structure for each case study is exclusive and cannot be readily transferred. Also, they found that the indicator 'current trip satisfaction' is better than a general satisfaction over 'all trips'. They used SEM to find derived importance in service. Mesbah et al., (2022) studied the customer satisfaction of Tehran metro in different groups of users and found that despite differences across groups, a convenient and secure transit system are the most important latent factors for all customers.

In this paper, the model obtained from the new dataset of CSS2 can be applied to the CSS1 data since both datasets are from the same system and area. Sample size in SEM studies depends on complexity of model, distributional characteristics of observed variables, and the method used for estimation (Kline, 2005). There are rules of thumb to estimate sample size (Bentler and Chou, 1987, Tanka, 1987), although several published studies did not follow such rules (Hadiuzzman et al., 2017). Table 1 represents some studies with SEM and their sample size.

The importance of understanding the past performance is because of evaluating trends and determining the effect of new policies and other interventions. A before-after study enable transit authorities to monitor effects of changes (Ryus, 2003). In addition to performance measurement, before-after study helps researchers in several fields in transit (Persaud et al., 2001; Patel et al., 2007; Fayish and Gross, 2010; Høye and Laureshyn, 2019; Pu et al., 2020). Dziekan and Kottenhoff (2007) documented

a reduction of 20% in waiting time after the installation of electronic displays in a before-after study on tramline 15 in the Hague, Netherland. Results were attained by a questionnaire given to users. de Oña et al., (2014) suggested service improvements based on indices obtained from an aggregated method and assessed each attribute by analyzing passenger's perception. They used regular and periodic data to analyze performance measures from 2006 to 2012 in metropolitan Granada, Spain. de Oña et al., (2015) measured level of service quality over time in Granada (Spain). As service quality is a "multidimensional construct", changes of user's perception are related to variation of customer satisfaction with the service attributes over time (de Oña et al., 2015). In some cities (e.g. Nashville, US), a CSS is performed monthly to control performance, safety, and cleanliness (Ryus, 2003). However, neither of these studies either did not investigate the cause of change in customer satisfaction or considered the perception of customers as a result of an improvement.
A study on passengers' satisfaction is carried out to reveal the effectiveness of improvements in transit. It should be noted that most of previous studies did not concentrate on user's perception of before and after implementing changes in service attributes. Moreover, there is no study before and after applying changes in a public transit system in a populated city in developing countries. Therefore, the objective of this study is to compare transit performance before and after some changes in the service quality attributes.

3. Methodology

Service quality can be monitored in two ways of a) aggregated and b) disaggregated models (de Oña et al., 2014). In this paper, the importance of each attribute is estimated by SEM and an IPA is then applied to compare the effect of changes on each attribute in the two surveys, and to prioritize the future enhancements. A customer satisfaction study on the 'before' conditions was conducted and called CSS1. Similarly, an 'after' survey is called CSS2. The performance of the metro line is first compared by the direct observations in the surveys using a two-sampled t-test. Afterwards, an Exploratory Factor Analysis (EFA) is applied to develop SEM structure for CSS2. The same structure is then verified by a Confirmatory Factor Analysis (CFA) on CSS1. Since the same structure is deployed in both models, a fair comparison can be made between importances of attributes which is presented in this paper. Finally, the effect of changes in service quality on passenger's satisfaction is determined.

3.1. Structural Equation Modeling (SEM)

An Exploratory Factor Analysis (EFA) is conducted to develop the model structure. Afterwards, factor loadings are attained from a Confirmatory Factor Analysis (CFA). Next, the validity and reliability of the measurement model are examined and for each latent variable, internal consistency is determined by Composite Reliability (CR) (Wang et al., 2016). Then, SEM is used to obtain the relative importance of each attribute.
SEM is a powerful multivariate analysis technique that combines regression, path analysis and factor analysis. SEM can identify and estimate the causality among variables (Sun, 2018). There are two main type of variables in SEM: observed variables (could be directly measured) and latent variables or factors (inferred from observed variables).
Therefore, the basic framework of SEM comprise structural model and measurement model. The relationship between observed variables and latent variables is determined in the measurement model. Latent variables are divide into endogenous and exogenous latent variables. These unobserved variables and their relationships are indicated through a structural model (Bollen, 1989).

Table 1. Literature review results for satisfaction with SEM

Year	Authors	Location	Kind of transportation	Sample size	Type of analysis
2007	Eboli and Mazzulla, (2007)	Cosenza, Italy	Bus	763	SEM
2010	Githui et al., (2010)	Nairobi, Kenya	Bus and matatu (mini-bus)	140	SEM
2016	Shen et al., (2016)	Suzhou, China	Urban rail transit	813	PSI model, SEM, IPA
2017	Das et al. (2017)	Dhaka, Bangladesh	Bus	400	SEM
2017	Hadiuzzman et al. (2017)	Dhaka, Bangladesh	Bus	655	SEM
2020	Soltanpour et al., (2020)	Tehran, Iran	Metro	300	SEM

The Maximum Likelihood method (ML is more common than other techniques to estimate SEM which has been selected in this study. Bollen (1989) and Washington et al., (2010) described these methods and their goodness of fit measures. Also the latent variable of overall service quality has been built through Second-order SEM (for more information about Second-order SEM see (Soltanpour et al., 2018) and (Soltanpour et al., 2020)).

Fit indices such as Chi-square (CMIN) the ratio of chi-square to degrees of freedom (CMIN/DF), goodness of fit (GFI), comparative fit index (CFI), Tucker-Lewis index (TLI) and root mean square error of approximation (RMSEA) estimate the model adequacy, and the cut-off values for CMIN/DF are recommended lower than 5 (Marsh and Hocevar, 1985), CFI and TLI should be more than 0.9 for a good fit, GFI to be more than 0.9 and RMSEA should be below 0.08 (Bollen, 1989; Browne and Cudeck, R, 1993; Hair et al., 1998; Hair, J. F., 2009).

3.2. Importance-Performance Analysis (IPA)

IPA is a technique that makes a four-quadrant plot by incorporating customers' perceived performance and importance (Martilla and James, 1977). Each quadrant in this plot suggests a different strategy for managers to improve customer satisfaction. All service attributes are categorized in IPA by their importance and performance in respectively vertical and horizontal axes which makes a four-quadrant plot (Q1 to Q4) as shown in Figure 1. The four areas are defined as follows:

Q1 (Keep up the good work): This quadrant indicates attributes with high performance and high importance that are performing well and investments on them should be continued.

Q2 (Concentrate here): This quadrant includes attributes with high importance and low performance that is the most crucial area. Attributes situated in Q2 represent the major service weakness and should be improved immediately with prioritized investments on them.

Q3 (Low priority): This part comprises of service attributes with low importance and low performance. Although the performance of service attributes in this area is not sufficient, they are relatively unimportant to the passengers.

Q4 (Possible overkill): Attributes fall into this area have high performance, and low importance meaning that excessive investments have been made in them.

4. Survey

Tehran metro line 3 is the case study of this research. Tehran, the capital of Iran, is the most populated city in this country. It has a population of about 9 million people in the central city, and its area has been extended to over 700 km2 (Habibian and Rezaei, 2017). The mass rail transit in Tehran included five lines when CSS1 was carried out and it had remained constant in CSS2. Tehran metro line 3 is one of the newest lines that connects southwest to northeast. Several service quality attributes have been changed from CSS1 to CSS2.

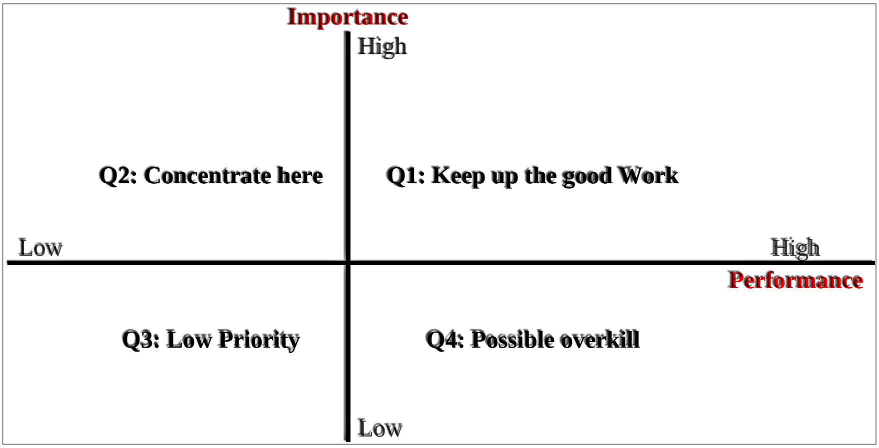

Fig. 1. Importance-Performance Grid

In our considerations, we assume that the vehicle blocks start their trips at the beginning of a minute and end their trips at the end of a minute. This means that if one trip starts at the minute *zm* and the other ends at that minute, then two vehicle blocks are needed to service these two "overlapping" trips.
Given the above assumption, the lower bound (LB) on the number of vehicle blocks can be determined
Based on the above presented formulas, we propose the following variants (depending on the level of detail of the data held) for computing the LB on the number of vehicle blocks (assuming that the LBs α and β refer to the same minute):

4.1. Data

The interviews were conducted based on a paper-based questionnaire proposed to the passengers while on board. A total number of 684 respondents participated in the two surveys: 300 responses were collected in February, and another 384 in December 2017. All interviews were during the weekdays between 8:00 AM and 9:00 PM, and the interviewees were selected randomly. The interviewees were asked to indicate their satisfaction with service quality attributes in a 5-point Likert scale (one denotes very dissatisfied, and five very satisfied). Questionnaires were structured into three main sections:

(1) passengers' satisfaction with individual service attributes (22 service attributes), plus two questions on overall satisfaction:
  - Overall satisfaction with current metro trip
  - Overall satisfaction with all metro trips made so far,

(2) trip characteristics,
(3) sociodemographic characteristics.

4.2. Descriptive statistics

Figure 2 presents the summary statistics of sociodemographic and trip characteristics of the dataset in both surveys. Although these variables are also important in customer satisfaction analysis (Saeidi et al., 2020), this paper only investigates the effect of changes in service attributes since the descriptive statistics indicates that the sociodemographic and trip characteristics in both surveys were similar. Respondents built up a sample of almost equally divided between males and females. The major part of the interviewees were younger than 40 (more than 70%). Most of the sample were university-educated (more than 50%), and there was a small group with no university degree. A considerable part of the respondents were not the head of household (in the Iranian social structure men are obliged to support a household with income). Most passengers traveled daily (about 50%). Most of the respondents were interviewed in off-peak hours.

5. Results and discussion

The changes in service attributes of metro line 3 between the two surveys included a decreased headway (from 18 to 9 minutes), addition of a new station, and an increase in fare. This paper assesses how the combination of these changes have affected the customer satisfaction. The SEM model is developed based on CSS2 and the same model is applied to CSS1 (CFA is implemented to accept or reject the model), then the derived importance of attributes in each survey are compared in the IPA diagram.
The mean and standard deviation of passengers' satisfaction level of each service attribute are presented in Table 2 as well as the t-test results of the differences between CSS1 and CSS2. Table 2 indicates that cleanliness at station, cleanliness on board, ticket purchasing or recharging, and ticket presenting were attributes with the highest satisfaction in both CSSs. Also, waiting time and seat availability on board were the attributes with the lowest satisfaction levels in CSS1, while seat availability on board and crowdedness on board were attributes with the lowest satisfaction scores in CSS2. Furthermore, it is clear that reducing the headway (which is represented by waiting time) boosted the waiting time, and the improvement was statistically significant (at the 95% confidence interval, with a Tcritical value equal to 1.96). However, the satisfaction levels of air conditioning on board, egress time to destination, information availability on board, safety at the station, safety on board, security at the station, security on board, cleanliness at the station and cleanliness on board have been reduced with t-test values more than the critical value (1.96 at 95%) showing the significance of this reduction. The average overall satisfaction of passengers with their current trip in both surveys remained almost the same, while the satisfaction of passengers with 'all trips' increased; yet the increase was not significant. Standard deviations in Table 2 are all below 2.07, which is the critical value demonstrating the homogeneous opinions between users (de Oña, Eboli and Mazzulla, 2014). Therefore, passengers' opinions had homogeneity in

this data. In brief, increasing the frequency has boosted satisfaction with the waiting time, but at the same time increase in demand reduced satisfaction with attributes such as safety, security, and cleanliness. The net of these two led to no change in the overall trip satisfactions (attributes 23 and 24).

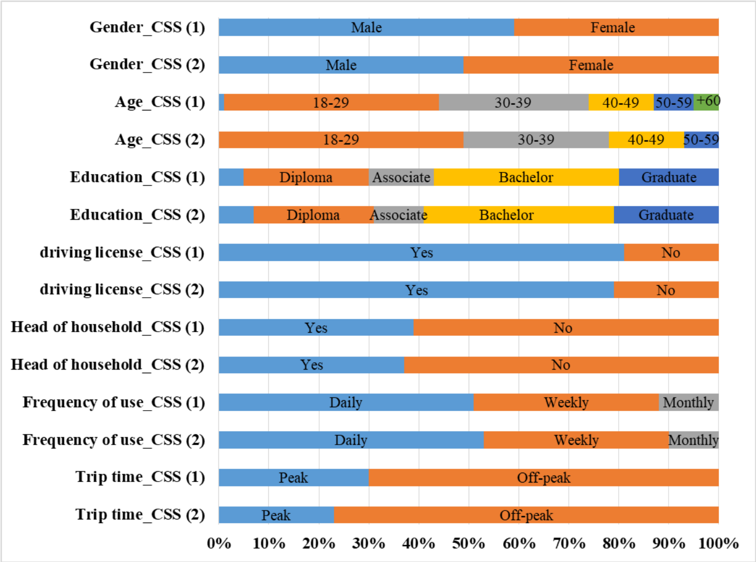

Fig. 2. Descriptive statistic of both surveys

Table 2. Mean, Standard deviation and T-TEST for performance of service attribute in both surveys

Service Quality Attributes	**Mean**		**Standard deviation**		**t-test**
	CSS (1)	**CSS (2)**	**CSS (1)**	**CSS (2)**	
Access time to station	3.74	3.73	1.11	1.07	0.12
Ticket purchasing or recharging	**4.45**	**4.39**	0.88	0.77	0.93
Ticket presenting	**4.42**	**4.32**	0.87	0.81	1.54
Fare	4.00	3.91	1.14	1.04	1.06
Staff behavior	3.99	3.90	1.05	1.03	1.12
Waiting time	2.59	3.08	1.32	1.24	**4.95**
Seat availability on board	2.92	2.81	1.44	1.40	1.00
Air conditioning on board	3.82	3.47	1.16	1.11	**3.99**
Travel time	3.83	3.90	1.08	0.92	0.90
Transfer time	3.60	3.50	0.91	1.03	1.35
Egress time to destination	3.79	3.52	1.12	1.12	**3.13**
Crowdedness on platform	3.30	3.19	1.29	1.21	1.14
Crowdedness on board	3.08	2.88	1.41	1.30	1.90
Information availability at the station	4.00	3.88	1.16	1.03	1.41
Information availability on board	3.92	3.69	1.07	1.07	**2.79**
Safety at the station	4.10	3.90	1.04	1.03	**2.51**
Safety on board	3.72	3.52	1.18	1.13	**2.24**
Security at the station	4.09	3.79	1.10	1.05	**3.61**
Security on board	3.76	3.49	1.25	1.18	**2.87**
Cleanliness at the station	**4.54**	**4.34**	0.74	0.77	**3.44**
Cleanliness on board	**4.40**	**4.19**	0.76	0.89	**3.32**
Ethical and behavioral messages	3.86	3.72	1.15	1.16	1.57
Current trip	3.91	3.90	0.94	0.83	0.14
All trips	3.39	3.52	1.00	0.92	1.75

5.1. Structural Equation Modeling (SEM)

Application of SEM to CSS1 and CSS2 data lead to models S(1) and S(2). An EFA determined the structure of these two models. Therefore, a similar structure was utilized to assess the influences of service changes on customer satisfaction. The applied structure resulted in five latent variables including "Main service", "Comfort", "Information", "Protection", and "Physical appearance" and is shown in Figure 3. Each attribute is explained by service attributes shown in Table 3.

1) Main service comprises of ticket purchasing or recharging, ticket presenting, fare value, staff behavior, waiting time, and travel time which are the main operational attributes in the service.
2) Comfort includes physical comfort attributes for travelers such as seat availability on board, air conditioning on board, crowdedness on platform, and crowdedness on board.
3) Information is related to the attributes that transmit some information to travelers from station and on board. Information availability at station, and information availability on board are assigned to this latent variable.
4) Protection includes safety at the station, security at station, and security on board. When passengers feel protected from physical injury they feel safe, and when they are kept from emotional harm it makes them feel secure.
5) Physical Appearance consists of cleanliness of the station, cleanliness on board, and behavioral messages.

The results of CFA that was applied to both S(1) and S(2) are presented in Table 3. Factor loadings for all attributes in S(1) and S(2) were above the cut off criteria of 0.4 (Stevens, 2002; Hair, J. F., 2009), thus none of the attributes were eliminated from the structure. Also, the Composite Reliability (CR) for all latent variables were more than the recommended threshold of 0.6 (Bagozzi and Yi, 1998). Thus, according to the factor loadings and CR, the structure was significant for both surveys.

With the above factor analysis, SEM is implemented. The derived importance obtained from SEM in S(1) and S(2) are presented in Table 4. Also, Figure 3 exhibits the conceptual framework of the second-order SEM in S(1) and S(2). Since observed variables occupied a lot of space, they are omitted in Figure 3 but shown in Table 4. According to this table, security at the station, security on board, crowdedness on the platform, crowdedness on board, seat availability on board, and travel time were the most important attributes in both surveys, and the ethical and behavioral messages were the least important ones.

Table 3. Factor loading and CR for both S(1) and S(2)

Service Quality Attributes	Latent variables	Factor loading		CR	
		S(1)	S(2)	S(1)	S(2)
Ticket purchasing or recharging	Main service	0.40	0.44	0.6	0.7
Ticket presenting		0.41	0.54		
Fare		0.42	0.40		
Staff behavior		0.43	0.53		
Waiting time		0.40	0.60		
Travel time		0.46	0.66		
Seat availability on board	Comfort	0.71	0.83	0.8	0.8
Air conditioning on board		0.51	0.60		
Crowdedness on platform		0.73	0.79		
Crowdedness on board		0.77	0.83		
Information availability at the station	Information	0.64	0.77	0.7	0.7
Information availability on board		0.82	0.78		
Safety at the station	Protection	0.56	0.53	0.8	0.8
Security at the station		0.87	0.85		
Security on board		0.78	0.86		
Cleanliness at the station	Physical Appearance	0.74	0.79	0.7	0.7
Cleanliness on board		0.76	0.81		
Ethical and behavioral messages		0.44	0.40		

Latent exogenous variables with the highest positive effect on service quality in S(1) were the main service (0.86), comfort (0.77), protection (0.68), physical appearance (0.59) and information (0.45) and in S(2) were the main service (0.87), comfort (0.71), protection (0.7), information (0.62) and physical appearance (0.6). It's interesting that the value of latent variables coefficient in both surveys were similar. All coefficients were significant with a p-value of less than 0.05.

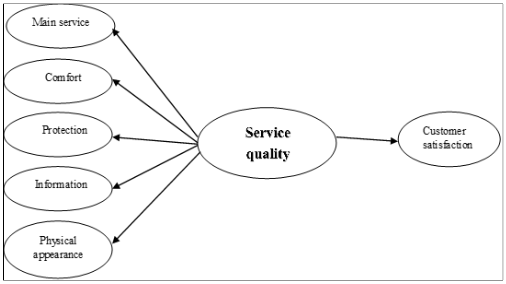

Fig. 3. The conceptual framework of SEM

Analyzing the goodness-of-fit of SEM models plays an important role in the model validity (Shen et al., 2016). Chi-square, RMSEA and GFI are absolute fit indices that demonstrate how well a "baseline model" fits the sample data (Wiley, 1973). Due to the restrictions of Chi-square, which is sensitive to sample size, researchers use CMIN/DF (Jöreskog, 1970) with a suggested ratio lower than 5.0 (Marsh and Hocevar, 1985). The RMSEA for a good fit is below 0.08, and a value of 0.08 to 0.1 is a mediocre fit (Browne and Cudeck, 1993, Hair, 2009). The Goodness-of-Fit Index (GFI) ranges between 0 to 1 and a value of 0.9 or greater demonstrates a well-fitting model (Jöreskog, 1970). Incremental fit indices compare "the chi-square value to a baseline model" (Jöreskog, 1970) consisting of TLI and CFI that range from 0 to 1; again, a value greater than 0.9 demonstrates a good fit. Fitness indices of the model are presented in Table 5. CMIN/DF is between 2 to 5. GFI, CFI and TLI demonstrate a good fit, and RMSEA is below 0.08. All fit indices confirm that both models fit well to the samples.

Table 4. Derived importance from SEM

Service Quality Attributes	Attribute (IPA)	Latent variables	Derived Importance	
			S(1)	S(2)
Ticket purchasing or recharging	Ticket_purch	Main service	0.32	0.29
Ticket presenting	Ticket_pre		0.39	0.36
Fare	Fare		0.29	0.26
Staff behavior	Staff		0.39	0.36
Waiting time	Waiting_t		0.44	0.42
Travel time	Travel_t		0.48	0.46
Seat availability on board	Seat	Comfort	0.49	0.48
Air conditioning on board	vent		0.36	0.36
Crowdedness on platform	Crowd_st		0.47	0.46
Crowdedness on board	Crowd_tr		0.49	0.47
Information availability at the station	Info_ st	Information	0.43	0.37
Information availability on board	Info_tr		0.43	0.40
Safety at the station	Safe_ st	Protection	0.32	0.30
Security at the station	Secur_st		0.51	0.48
Security on board	Secur_tr		0.52	0.49
Cleanliness at the station	Clean_st	Physical Appearance	0.44	0.38
Cleanliness on board	Clean_tr		0.45	0.41
Ethical and behavioral messages	Ethic		0.25	0.21

Table 5. Fit indices for S(1) and S(2)

	CMIN	CMIN/DF	GFI	RMSEA	CFI	TLI
S(1)	361.95	2.21	0.891	0.064	0.871	0.851
S(2)	404.67	2.47	0.904	0.062	0.904	0.890

5.2. Importance – Performance Analysis (IPA)

The improvement in service attributes can be prioritized based on the derived importance calculated in Table 4, and the performance presented in Table 2. Four quadrants of the IPA are formed and presented in Figure 4. The blue attributes and dividing line are related to CSS (1), and the red attributes with their red dividing line are related to CSS (2). Those attributes with a change in their quadrants during two the surveys are demonstrated with red arrows.

Q1: security at station, and cleanliness on board remained in Q1 before and after changes. So the current related procedures need to be continued.

Q2: waiting time, seat, crowdedness on the platform, crowdedness on board, and security on board were located in Q2 before and after the changes. It shows that reducing waiting time improved the performance, however it was insufficient to considerably satisfy the passengers. Moreover, according to Al-Amin et al., (2021) seat availability is an attribute that can affect customer satisfaction. As a result, more improvements in the Q2 attributes need to be made.

Q3: attributes located in this quadrant should be improved, but they have a lower priority than Q2. Air conditioning was in Q3 in both surveys.

Q4: staff behavior, ticket presenting, ticket purchasing or recharging, safety at station, ethical ads, and fare were in Q4. The quadrant of these attributes did not change between the two surveys, which shows an unnecessary expenditure on them.

According to Figure 4, travel time in CSS (1) was in Q2, whereas in CSS (2) it was improved and moved to Q1. Also, information availability on board was located in Q1, which is moved to Q2 after the changes, and became a concern. Furthermore, information availability at station, and cleanliness at station were moved from Q1 to Q4. Therefore, excessive resources were spent on them during the study period.

Overall, improvement priorities were identified through an IPA. The poorly-performed attributes with a high importance should be the top priorities since they have the most negative effects. In general, the results of the IPA showed that improvement priorities did not make a big difference regarding many attributes before and after applying service changes. Users were very satisfied with cleanliness (at the station and on board), and security at the station in both surveys. Passengers remained dissatisfied with waiting time and crowdedness. Metro is clean but not comfortable (Since the comfort attributes are located in Q2 with low performance ratings, the service is not comfortable enough in passengers' point of view) and should be more secure on board. Travel time needed to be improved in CSS (1), which has been achieved in CSS (2). Information on board was an attribute that moved from Q1 to Q2 and it should be improved more.

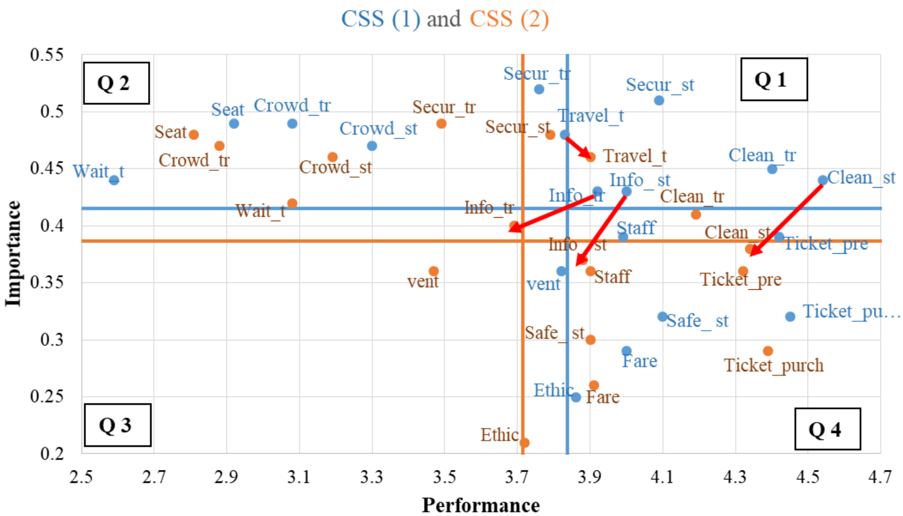

Fig. 4. Importance – performance plot

6. Conclusions

Considering service quality in public transit is an important issue for transit managers. Understanding how changes in service quality can influence passenger satisfaction helps transit managers to plan a better investment and identify key attributes.

While there are no before-after studies in the metro system of developing countries, this paper proposed how customer satisfaction is affected by different attributes of service quality before and after implementing a set of changes in Tehran metro line 3. SEM is an appropriate methodology for establishing the relationships between customer satisfaction and service quality which was used in this paper. Based on the performance and importance of attributes obtained from SEM, an IPA specified the attributes that needed to be improved, and those attributes that unnecessary resources were spent on them. The IPA helped classifying attributes into different categories to find the priorities of improvements for investments in service. The data was collected in two separate surveys; one with 300 participants in the before study, and one with 384 participants in the after study. The changes consisted of opening a new station, increasing the fare, and reducing the headway. Based on the data collected in the surveys, a SEM was developed with five latent variables obtained from EFA, namely main service, comfort, information, protection, and physical appearance. This structure was assessed on both the before and after surveys and showed to be valid. A CFA was applied to confirm the structure. Security at the station and security on board were the most important service attributes in both surveys according to their factor loadings, while ethics and behavioral messages had the smallest factor loading and the least importance. An IPA was used to devise a better investment plan to improve the service quality. The top priority attributes were seat availability on board, crowdedness at the station, crowdedness on board, security on board, and waiting time in both before and after studies. Reducing headway influenced the satisfaction of waiting time, but it remained in quadrant 2 (the priority area) for improvement. Travel time was improved and moved from the quadrant 2 (the 'concentrate here' area) to quadrant 1 (the area of good work). Information on board was completely vice versa as it was moved from quadrant 1 (the good work area) to quadrant 2 (the priority area). For the second priority, air conditioning was in the same area before and after the changes, and it was not improved. 'Cleanliness on board' and 'security at station' were considerable for passengers before and after the changes. According to the results, Tehran metro line 3 was clean, safe with an appropriate fare, whereas security, crowdedness, and waiting time needed to be improved.

For future research, it is suggested to periodically measure the public transit performance so that the effect of seasonal and annual changes on passengers' satisfaction could be revealed as well. In addition to that, this study could be conducted on other metro lines or other public transit modes to compare and contrast service quality in the whole system, and to more precisely analyze the effect of service changes on the passengers' satisfaction.

References

[1] Abalo J., Varela J., Manzano V., (2007). Importance values for importanceperformance analysis: a formula for spreading out values derived from preference rankings. *Journal of Business Research*, *60(2)*,115–121.

[2] Abenoza, R. F., Cats, O. Susilo, Y. O., (2017). Travel satisfaction with public transport: Determinants, user classes, regional disparities and their evolution. *Transportation Research Part A: Policy and Practice 95*, 64–84. DOI: 10.1016/j.tra.2016.11.011.

[3] Al-Amin, M., Islam, M. S., Ahammed, M. S., (2021). Service Quality Analysis of Intercity Bus Service of Pabna-Dhaka Highway. *American Journal of Traffic and Transportation Engineering*, *6*(4), 116-127.

[4] Bagozzi, R. P., Yi, Y., (1998). On the evaluation of structural equation models. *Journal of the academy of marketing science*, *16*(1), 74-94.

[5] Bentler, P. M., Chou, C. P., (1987). Practical issues in structural modeling. *Sociological methods & research*, *16*(1), 78–117.

[6] Bollen, K. A., (1989). *Structural equations with latent variables* (Vol. 210). John Wiley & Sons.

[7] Browne, M. and Cudeck, R (1993). Alternative ways of assessing model fit. Sage focus editions.

[8] Das, T., Apu, N., Hoque, M. S., Hadiuzzaman, M., Xu, W., (2017). Parameters Affecting the Overall Performance of Bus Network System at

Different Operating Conditions: A Structural Equation Approach. *Transportation Research Procedia. Elsevier B.V., 25*, 5059–5071. DOI: 10.1016/j.trpro.2017.05.206.

[9] Dell'Olio, L., Ibeas, A., Cecin, P., (2010). Modelling user perception of bus transit quality. *Transport Policy*, 388–397.

[10] de Oña, J. de Oña, R., Eboli, L., Eboli, L. Mazzulla, G., (2013). Perceived service quality in bus transit service: A structural equation approach. *Transport Policy*, 219–226. DOI: 10.1016/j.tranpol.2013.07.001.

[11] de Oña, J. de Oña, R., Eboli, L., Mazzulla, G., (2016). Index numbers for monitoring transit service quality. *Transportation Research Part A: Policy and Practice, 84*, 18-30. DOI: 10.1016/j.tra.2015.05.018.

[12] de Oña, J., de Oña, R., (2014). Quality of Service in Public Transport Based on Customer Satisfaction Surveys : A review and assessment of methodological approaches quality of service in public transport based on customer satisfaction surveys : A review and assessment of methodological'. *Transportation Science*. , *49*(3), 605-622. DOI: 10.1287/trsc.2014. 0544.

[13] de Oña, R., Eboli, L., Mazzulla, G., (2014). Monitoring changes in transit service quality over time, *Procedia - Social and Behavioral Sciences*, 974–983. DOI: 10.1016/j.sbspro. 2014.01.132.

[14] Dziekan, K., Kottenhoff, K., (2007). Dynamic at-stop real-time information displays for public transport: effects on customers. *Transportation Research Part A: Policy and Practice*. 489–501. DOI: 10.1016/j.tra.2006. 11.006.

[15] Eboli, L., Mazzulla, G., (2007). Service quality attributes affecting customer satisfaction for bus transit. *Journal of Public Transportation*, 21–34.

[16] Eboli, L., Mazzulla, G., (2011). A methodology for evaluating transit service quality based on subjective and objective measures from the passenger's point of view. *Transport Policy*, 172–181.

[17] Eboli, L., Mazzulla, G., (2015). Relationships between rail passengers' satisfaction and service quality: a framework for identifying key service factors. *Public Transport, 7(2), 185-201*. DOI: 10.1007/s12469-014-0096-x.

[18] Echaniz, E. Ho, C.Q., Rodriguez, A., dell'Olio, L., (2019). Comparing best-worst and ordered logit approaches for user satisfaction in transit services 130. *Transportation Research Part A: Policy and Practice, 130*, 752–769.

[19] Fayish, A. C., Gross, F., (2010). Safety Effectiveness of Leading Pedestrian Intervals Evaluated by a Before – After Study with Comparison Groups'. *Transportation research record 2198.1*,. 15–22. DOI: 10.3141/2198-03.

[20] Friman, M., Fellesson, M., (2004). Service supply and customer satisfaction in public transportation: the quality paradox. *Journal of Public transportation, 12(4), 4*. 57–69.

[21] Gärling, T., Eek, D., Loukopoulos, P. Fujii, S., Johansson-Stenman, O., Kitamura, R., Pendyala, R., Vilhelmson, B., (2002). A conceptual analysis of the impact of travel demand management on private car use. *Transport Policy*, 59–70.

[22] Githui, J., Okamura, T., Nakamura, F., (2010). The structure of users' satisfaction on urban public transport service in developing country: the case of Nairobi, Proceedings of the Eastern Asia Society for Transportation Studies Vol. 7 (The 8th International Conference of Eastern Asia Society for Transportation Studies, 2009). *Eastern Asia Society for Transportation Studie*.

[23] Habibian, M., Rezaei, A., (2017). Accounting for systematic heterogeneity across car commuters in response to multiple TDM policies: case study of Tehran. *Transportation*. 44 (4). 681–700.

[24] Hadiuzzman, M., Das, T., Hasnat, M.M., Hossain, S., Rafee Musabbir, S., (2017). Structural equation modeling of user satisfaction of bus transit service quality based on stated preferences and latent variables, *Transportation Planning and Technology*, *40*(3), 257–277. DOI: 10.1080/03081060. 2017.1283155.

[25] Hair, J. F., B. W. C., Babin, B. J., Anderson, R. E., (2009). Multivariate data analysis. Upper Saddle River, NJ [etc.]. *Pearson Prentice Hall, New York, NY: Macmillan, 24*, 899.

[26] Hair, J. F., Anderson, R.E., Tatham, R.L., et al., (1998). Multivariate data analysis, Prentice hall Upper Saddle River, NJ.

[27] Halakoo, M., Mesbah, M., Habibian, M.,

Mohamed, M., (2022). Modelling quality of service in a fixed route shared taxi (Jitney). *Case Studies on Transport Policy*, *10*(4), 1988-2000.

[28] Harreman-Fernandes, M., Diab, E., Cui, B., DeWeese, J., Crumley, M., El-Geneidy, A., (2021). Transit customer satisfaction research: is the customer always right?. *Handbook of Public Transport Research,* 57-69.

[29] Høye, A., Laureshyn, A., (2019). SeeMe at the crosswalk: Before-after study of a pedestrian crosswalk warning system, *Transportation Research Part F: Traffic Psychology and Behaviour. Elsevier Ltd, 60,* 723–733. DOI: 10.1016/j.trf.2018.11.003.

[30] Hu, X., Chen, X., Zhao, J., Yu, K., Long, B., Dai, G., (2022). Comprehensive service quality evaluation of public transit based on extension cloud model. *Archives of Transport, 61*(1), 103-115.

[31] Jöreskog, K. G., (1970). A general method for estimating a linear structural equation system, *ETS Research Bulletin Series, 1970*(2), i-41

[32] Kline, R. B., (2005). Principles and Practice of Structural Equation Modeling. 2nd ed. *New York: Guilford, 3.*

[33] Lei, P., Qiong, W., (2007). Introduction to structural equation modeling: Issues and practical considerations, *Educational Measurement: issues and practice 26,* 33–43.

[34] Machado-León, J. L. de Oña, R., Baouni, T., de Oña, J., (2017). Railway transit services in Algiers: priority improvement actions based on users perceptions, *Transport Policy*, 175–185. DOI: 10.1016/j.tranpol.2016.10.004.

[35] Madha, A., Salman, A. D., Hussain, H. D., Borhan, N., Riza Atiq, O. K. R., (2016). Analysis of travel behaviour in Petaling Jaya, Malaysia: An application of the theory of planned behaviour. *Archives of Transport, 38.*

[36] Marsh, H. W., 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, *Psychological bulletin*, *97*(3), 562.

[37] Martilla, J. A., James, J. C., (1977). Importance–performance analysis, *Journal of marketing*, *41*(1), 77–79.

[38] Merlin, L. A., Singer, M., Levine, J., (2021). Influences on transit ridership and transit accessibility in US urban areas. Transportation Research Part A: *Policy and Practice, 150,* 63-73.

[39] Mesbah, M., Sahraei, M., Soltanpour, A., Habibian, M., (2022). Perceived service quality based on passenger and trip characteristics: A structural equation modeling approach. *Journal of Rail Transport Planning & Management, 23,* 100340.

[40] Morpace International, Transit Cooperative Research Program and Cambridge Systematics, (1999), United States, Federal Transit Administration, Transit Development Corporation, *A Handbook for Measuring Customer Satisfaction and Service Quality (Vol. 47). Transportation Research Board.*

[41] Municipality, I. and communication of Tehran., (2017). Statistical Report of Tehran.

[42] Park, K., Farb, A., Chen, S., (2021). First-/last-mile experience matters: The influence of the built environment on satisfaction and loyalty among public transit riders. *Transport policy, 112,* 32-42.

[43] Patel, R. B., Council, F. M., Griffith, M. S., (2007). Estimating Safety Benefits of Shoulder Rumble Strips on Two-Lane Rural Highways in Minnesota Empirical Bayes Observational Before-and-After Study, *Transportation Research Record 2019.1,* 205–211. DOI: 10.3141/2019-24.

[44] Persaud, B. N. Retting, R.A., Garder, P.E., Lord, D., (2001). Safety Effect of Roundabout Conversions in the United States Empirical Bayes Observational Before-After Study, *Transportation Research Record 1751*, 1–8.

[45] Pu, Z., Li, Z., Jiang, Y., Wang, Y., (2020). Full Bayesian Before-After Analysis of Safety Effects of Variable Speed Limit System, *IEEE Transactions on Intelligent Transportation Systems*, *28*(2), 1–13. DOI: 10.1109/tits.2019.2961699.

[46] Putra, A. A. Jinca, M.Y., Riyanto, B., Mulyono, A. T., (2014). The Satisfaction Analysis for the Performance of Public Transport Urban Areas, *International Refereed Journal of Engineering and Science*, 38–44.

[47] Ryus, P., (2003). A guidebook for developing a transit performance-measurement system (TCRP Report 88)., *Transportation Research*

Board: Washington, DC, USA.

[48] Saeidi, T., Mesbah, M., Habibian, M., (2020). Sequenced Ordered Logit Model Considering Latent Variables for Determining Trip Satisfaction of Metro Passengers, *Transportation Research Record: Journal of the Transportation Research Board, 7 September*, 755–766. DOI: 10.1177/03611981 20931846.

[49] Shen, W., Xiao, W., Wang, X., (2016). Passenger satisfaction evaluation model for Urban rail transit: A structural equation modeling based on partial least squares, *Transport Policy*, 20–31.

[50] Soltanpour, A., Mesbah, M., Habibian, M., (2018). Percieved 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*, 1–18.

[51] Soltanpour, A., Mesbah, M., Habibian, M., (2020). Customer satisfaction in urban rail : a study on transferability of structural equation models, *Public Transport 12.1*, 123–146. DOI: 10.1007/s12469-019-00223-y.

[52] Stevens, J. P., (2002). *Applied multivariate statistics for the social sciences (4th ed.)*. Hillsdale, NS: Erlbaum.

[53] Sukwadi, R., Susanto, A., & Liang, Y. C., (2021). Airline service quality evaluation for Indonesian low-cost carriers based on Extenics innovation theory. *Archives of Transport, 58*.

[54] Sun, S. C., (2018). Public Transit Loyalty Modeling Considering the Effect of Passengers' Emotional Value: A Case Study in Xiamen, China. *Journal of Advanced Transportation*.

[55] Tanka, J. S., (1987). How big is big enough? Sample size and goodness of fit in structural equation models with latent variables, *Child Development*, 134–146.

[56] Wang, S. Fan, J., Zhao, D., Yang, S., Fu, Y., (2016). Predicting consumers' intention to adopt hybrid electric vehicles: using an extended version of the theory of planned behavior model, *Transportation*, 123–143.

[57] Washington, S. P., Karlaftis, M. G., Mannering, F., (2010). *Statistical and Econometric Methods for Transportation Data Analysis, Second Edition.* Chapman and Hall/CRC. DOI: 10.1201/9781420082869.

[58] Wiley, D. E., (1973). Identification Problem For Structural Equation Models With Unmeasured Variables, *structural equation models in the social science*, Seminar Press, Inc United States of America, 69–83.

[59] Wu, X., Jason, C., (2017). *Using the Three-Factor Method to Identify Improvement Priorities for Express and Local Bus Services in the Twin Cities, Minnesota*, No. 17-02464.

[60] Yuan, Y., Yang, M., Wu, J., Rasouli, S., Lei, D., (2019). Assessing bus transit service from the perspective of elderly passengers in Harbin China, *International Journal of Sustainable Transportation 13*(10). 761–776. DOI: 10.1080/15568318.2018.1512691.

[61] Zehmed, K., Jawab, F., (2021). The performance of tramway service from the users' viewpoint: A comparative analysis between two Moroccan cities. *Archives of Transport,* 60, 7-21. DOI: 10.5604/01.3001.0015.5223.