International Journal of Transportation Science and Technology 11 (2022) 780–789

Contents lists available at ScienceDirect

International Journal of Transportation Science and Technology

journal homepage: www.elsevier.com/locate/ijtst

Applying an integrated approach to metro station satisfaction evaluation: A case study in Shanghai, China

Linbo Li *, Tianshuang Gao, Ling Yu, Yahua Zhang

[a] *Key Laboratory of Road and Traffic Engineering of Ministry of Education, Tongji University, Shanghai 201804, China*
[b] *School of Commerce, University of Southern Queensland, Toowoomba, Queensland, Australia*

ARTICLE INFO

Article history:
Received 1 June 2021
Received in revised form 5 October 2021
Accepted 17 October 2021
Available online 26 October 2021

Keywords:
Metro station
Passenger satisfaction
Evaluation index system
Structural equation mixture model
Cloud model

ABSTRACT

Improving the service level of metro station is the key to attracting passengers to achieve sustainable development. The satisfaction evaluation approach from the passenger's point of view can intuitively and meticulously reflect the passenger's experience when taking the metro and provide a basis for improving the service. Therefore, this paper proposes an integrated approach for a metro station satisfaction evaluation based on passenger perception. In this approach, the evaluation index system is first dynamically screened and established. Then, the structural equation mixture model (SEMM) is adopted to determine the index weight based on the customer satisfaction index model and the benchmark. The result clouds are generated by the basic algorithm of the cloud model. Finally, the targeted improvement suggestions are proposed based on the evaluation results. Taking the Shanghai Metro People's Square Station as a case, the evaluation approach was verified, which reflected the effectiveness and rationality of the approach. As the results show, this approach can not only obtain the desired level of passenger satisfaction, but also reflect the differences and consensus of passengers' perception, making it practically useful and valuable.

1. Introduction

Metro plays a vital role in the urban transportation system and generates great social and economic benefits. Improving the quality of service to meet passengers' demand and retain customers is key to the sustainable development of metro. At present, the evaluation system for metro stations is largely planning-oriented and does not reflect the actual operational status. An effective evaluation system should be user-oriented, that is, able to judge the quality of operations and determine the direction of improvement through the users' experience. Hence, identifying the key factors affecting the quality of metro service via passenger satisfaction evaluation is an important means for service improvement.

Satisfaction is a kind of emotional response, which is the attitude of passengers towards travel perception and expectation. Passenger satisfaction depends on passengers' perception of multiple attributes (Tyrinopoulos and Antoniou, 2008; Eboli and Mazzulla, 2009) of travel service quality, and is related to passengers' subjective feelings.

Peer review under responsibility of Tongji University and Tongji University Press.
* Corresponding author.
E-mail address: llinbo@tongji.edu.cn (L. Li).

https://doi.org/10.1016/j.ijtst.2021.10.004
2046-0430/

The concept of customer satisfaction was first formally proposed by Cardozo in 1965 and applied to marketing studies (Cardozo, 1965), and then some scholars began to introduce econometric models into the analysis of customer satisfaction. Later, Fornell et al. (1996) summarized the Customer Satisfaction Index (CSI) using previous studies and put forward the American customer satisfaction index (ACSI) model. The ACSI model measured the relationships running from the antecedents of overall customer satisfaction (customer expectations, perceived quality and perceived value) to its consequences (customer complaints and customer loyalty). In addition, many countries have conducted researches in the area of customer satisfaction based on Fornell's econometric models.

At present, there are no systematic and mature approaches for passenger satisfaction evaluation of metro stations. Many previous studies focused on the influencing factors of passenger satisfaction (Hensher et al., 2003; Habib et al., 2011; Del Castillo and Benitez, 2012; De Oña et al., 2013; Mouwen, 2015), including accessibility, information, time, customer service, comfort, safety, convenience, reliability, economy, capacity, etc. However, in the specific evaluation process, researchers often choose indicators based on literature or experience (Xiao and Li, 2011; Fan et al., 2013; Zhai et al., 2018). Because of the subjective nature, it is difficult to apply these indicators to the changing dynamics of metro. In addition, for different regions, objectives, and periods, there are different understandings and concerns about the indicators. The evaluation indicators need to be further screened and modified according to the characteristics of the research objectives.

The methods for determining the attribute weights of indices can be roughly divided into two categories: one is the subjective weighting method such as Delphi, Analytic Hierarchy Process (AHP), and expert grading method (Zhai et al., 2018); another is objective weighting method such as standard deviation method (SDM), entropy weight method (EWM), and the criteria importance through intercriteria correlation (CRITIC) method (Li et al., 2018). The subjective weighting methods have strong subjective randomness, and the evaluation results can be easily biased by the influence of the evaluator. In contrast, the objective methods mainly determine the weights based on the variability of the data within the indicator and the conflict between the indicators, but the degree of dispersion of the data may not accurately reflect the evaluator's judgment on the importance of the indicator. In addition, the value of the most important indicator does not necessarily have the greatest difference, meaning that the impact of the evaluation index on the overall satisfaction may not be adequately captured. In some cases, the determined weight could be contrary to the actual degree of importance.

In terms of evaluation methods, many studies adopted the customer satisfaction index model based on business service backgrounds, and solved it by using a structural equation model (SEM) (Shen et al., 2016; Zhang et al., 2016; Zhang et al., 2019). Additionally, evaluation systems are often integrated by various methods. Celik et al. (2014) proposed an evaluation framework which integrates statistical analysis, a two-type fuzzy logic model, and the VIKOR algorithm. Aydin et al. (2015) introduced an approach combining statistical analysis, fuzzy analytic hierarchy process, trapezoidal fuzzy sets and Choquet integral, to evaluate customer satisfaction of urban rail transit (URT). Machado-Leóna et al. (2017) used a tool that combines an Importance-Performance Analysis and a decision tree model. Concepción Garrido et al. (2014) proposed the use of Artificial Neural Networks (ANN) to analyse the service quality perceived by passengers. Wang et al. (2018) presented an intuitionistic fuzzy group decision model in evaluating the customer satisfaction level of URT lines in China. In short, there seems to be a consensus to adopt an integrated approach to evaluate customer satisfaction.

China has issued a national standard on urban public transit passenger satisfaction evaluation in which Part 3 details the indicators, models, data collection and processing, and report writing for URT service evaluation. Descriptive statistics are used for the satisfaction evaluation of each index, and SEM is used to analyse the weight of each index. This paper further develops a dynamic passenger satisfaction evaluation framework specifically for metro stations, which emphasizes the following aspects. First, the evaluation index set used is screened according to the passenger's cognition, that is, the passengers' actual feeling of each service index. So it can be dynamically adjusted and changed with the change of passenger's cognition. Second, the evaluation results not only quantitatively present the scored of satisfaction of passengers to metro station, but also visually show the randomness, fuzziness, and difference of the evaluation results of passenger groups. The use of cloud model can easily reveal the degree of dispersion and distribution range of the evaluation data, allowing researchers to put forward more targeted suggestions and priorities of measures for improvement in services.

The paper is structured in four sections. After the introduction and literature review in Section 1, Section 2 introduces the methodology of structural equation model, cloud model and the framework of passenger satisfaction evaluation. Section 3 takes the People's Square Station in Shanghai as a case study, introduces the questionnaire design and passenger survey, and then demonstrates the feasibility of proposed evaluation approach. Section 4 presents some conclusions about the study.

2. Methods

2.1. Structural equation model

The SEM is a multivariate statistical method based on a covariance matrix of variables to analyse the relationship between variables (Jöreskog, 1993). The core advantage of SEM is that it can reflect the relationship between observable manifest variables and latent variables that cannot be observed directly (Kline, 1998). A complete SEM includes both a measurement and structural model. The measurement model mainly reflects the relationship between manifest and latent variables, while the structural model captures the relationship between latent variables. Establishing an SEM generally includes five steps: model specification, model identification, model estimation, testing of model fit, and model modification (Hau et al., 2004). It is widely used to analyse travel behaviour and evaluate public transport service quality (Shen et al., 2016; Golob, 2003).

A passenger satisfaction model can be constructed on the basis of the theory of customer satisfaction index. At present, the American Customer Satisfaction Index (ACSI) is the most widely used customer satisfaction index measurement model, which is based on the Sweden Customer Satisfaction Barometer (SCSB) model. Considering that the characteristics of metro services do not need to take into account the brand image and involve the possible complaints of customers, this paper uses ACSI model to establish a passenger satisfaction evaluation framework. The concept model of passenger satisfaction at metro station is shown in Fig. 1.

Passenger expectation is an exogenous latent variable and perceived quality, perceived value, passenger satisfaction, passenger complaint and passenger loyalty are endogenous latent variables. The positive and negative signs indicate a positive and negative correlation, respectively. The meaning of each latent variable is as follows:

- Passenger expectation refers to the expectation that passengers place on the service quality of metro station before they travel;
- Perceived quality refers to the actual feeling and perception of quality of the metro station service when passengers receive the service;
- Perceived value refers to the subjective feelings on actual benefits of passengers after experiencing the metro service with reference to the fare;
- Passenger satisfaction refers to an evaluation of the overall satisfaction of the passengers with the metro station service they receive. It is also the result of the passenger's satisfaction with the service of the metro station and the perceived value of the passenger;
- Passengers complaint means that when the actual feeling of the passengers on the metro station service cannot meet the passenger's expectation, passengers will complain;
- Passenger loyalty refers to the long-term choice of passengers to travel by metro after experiencing satisfactory service quality.

In addition to constructing a satisfaction evaluation conceptual model, SEM is also used to obtain the weight of each index for passenger satisfaction evaluation. As mentioned above, the SEM can reflect the relationship between variables, and the path coefficient represents the contribution of independent variables to the dependent variable, that is, the degree of influence of each indicator on overall satisfaction, representing the weight. Compared with the traditional objective weighting method, SEM can better reflect the correlation between each index and overall satisfaction, and it is more reasonable in determining the weight. Therefore, the weights of the indices are obtained by normalizing the path coefficients.

2.2. Cloud model

Cloud model is an uncertainty conversion model between qualitative concepts and quantitative expression in order to analyse the fuzziness and randomness of linguistic terms and the relevance of these two features (Li and Du, 2017), which has been successfully applied to the fields of natural language processing, data mining, decision analysis, intelligent control, image processing among others.

A cloud can be expressed by the following three digital characteristics: expectation E_x, entropy E_n and hyper entropy H_e. For convenience, we can describe a cloud by using A = G (E_x, E_n, H_e). Among them, E_x represents the basic certainty measure of the qualitative concept, that is, the most representative value or sample of a qualitative concept in the universe; E_n is the uncertainty measure of the qualitative concept, which is measured by both fuzziness and randomness of the qualitative concept; H_e represents the uncertainty measure of the entropy value, which can reflect the degree of general acceptance of the qualitative concept. Cloud drops can be generated by these digital characteristics and aggregated into a cloud after a certain amount of accumulation. There are two kinds of cloud generation algorithms expressed as follows, called forward cloud generator and backward cloud generator respectively (Li and Du, 2017). Forward cloud generator is a mapping from qualitative

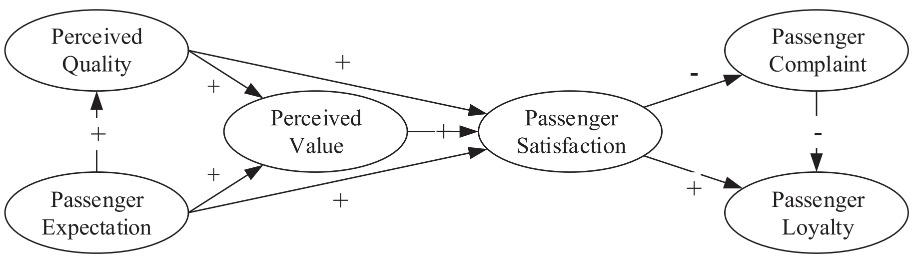

Fig. 1. Conceptual model of passenger satisfaction at metro station.

concept to its quantitative representation. While backward cloud generator is a transformation model from quantitative value to qualitative concept.

This paper regards the whole evaluation index system as a universe of discourse and regards each passenger as a cloud droplet. The overall characteristics of the cloud formed by the comprehensive evaluation results of all passengers reflect the passenger satisfaction of metro station. The forward cloud generator is used to describe the features of manifest variables based on the data, and the features of latent variables are obtained by the backward cloud generator.

2.3. Passenger satisfaction evaluation framework

The passenger satisfaction evaluation framework is divided into the following five parts:

Step 1: Establish an evaluation indicator set of perceived quality in different dimensions. This process includes several sub-steps. Firstly, the initial index set is established through literature review. Secondly, the questionnaire is designed according to the index description and through in-depth interviews. Thirdly, collect questionnaire data, and apply the critical ratio method, reliability analysis method and exploratory factor analysis to screen and reduce dimensions of indicators. Finally, a multi-dimensional evaluation index system is established with the indicator set U.

Step 2: Design the questionnaire and collect data. The design of this questionnaire is based on the evaluation index system of metro station established above. It strictly follows the principles of conciseness, adaptability, purpose, and pertinence of the questionnaire design. Use the satisfaction evaluation questionnaire designed to conduct SP (Stated Preference) survey and collect passenger satisfaction evaluation data.

Step 3: Determine the weight set. With the satisfaction survey data, we can use SEM to fit, modify and evaluate the established model to find the optimal satisfaction evaluation model. By normalizing the path coefficients of the optimal model, the weight of each first-level indicator $\{\mathrm{w}_1, \mathrm{w}_2, \ldots \mathrm{w}_i\}$, the weight of each second-indicator under the corresponding first-level indicator $\{\mathrm{w}_{11}, \mathrm{w}_{12}, \ldots \mathrm{w}_{ij}\}$ are determined respectively. Then the weight w'_{ij} of each indicator is got by $w_i * w_{ij}$. Finally, the weight set W of the whole system is obtained as $\left\{w'_{11}, w'_{12}, \ldots w'_{jk}\right\}$.

Step 4: Generate the cloud model of evaluation results. We can calculate the three actual parameters of each index based on the satisfaction survey data, and use the backward cloud generator to generate the corresponding parameter matrix Z of the index set (Equation (1)). Then we can obtain the cloud model C of the evaluation result by fuzzy operations between W and Z as shown in Eq. (2).

$$Z = \begin{Bmatrix} c_1 \\ c_2 \\ \ldots \\ c_n \end{Bmatrix} = \begin{Bmatrix} E_{x_1} & E_{n_1} & H_{e_1} \\ E_{x2} & E_{n_2} & H_{e_2} \\ & \ldots & \\ E_{x_n} & E_{n_n} & H_{e_n} \end{Bmatrix} \tag{1}$$

$$C = W \times Z \tag{2}$$

Since the operation process involves a mixture of cloud parameters and conventional parameters, the weighted synthesis method of cloud model is shown as Eq. (3):

$$E_x = \sum E_{x_i} w_i, E_n = \sqrt{\sum (E_{n_i}^2 w_i)}, H_e = \sum H_{e_i} w_i \tag{3}$$

Step 5: Compare and determine the evaluation results. By using the forward cloud generator, the comprehensive evaluation cloud model C can be converted into a normal cloud, and compared with the benchmark cloud.

For the benchmark cloud, the evaluation set is a set formed by the attributes of each qualitative comment in the actual survey. If there are g qualitative comments with different levels of satisfaction in the actual survey, the expression of the evaluation set V is as $\boldsymbol{V} = \{V_1, V_2, V_3, \ldots, V_g\}$, where V_g corresponds to g qualitative comments on different levels of satisfaction.

If the lower and upper bounds of each qualitative comment are and respectively, then we can calculate the digital characteristics of each qualitative comment with the formulas under bilateral constraints, and use the forward cloud generator to obtain the benchmark cloud respectively. In this way, the fuzzy comment set is transformed into an actual satisfaction evaluation scale. These three parameters are calculated as Eq. (4):

$$E_x = (t_{\min} + t_{\max})/2, E_n = (t_{\min} + t_{\max})/6, H_e = k \tag{4}$$

where k is a constant adjusted according to the fuzzy threshold of the variable itself. Only when the value of k is greater than 0 and less than $E_n/3$, can it be proved that the qualitative concept has a certain consensus. Therefore, k often takes the value of 0.1 in practical evaluations.

According to the range and shape of the normal cloud in all benchmark clouds, the corresponding evaluation results can be summarized. It can also be specific to the comparison of each indicator, so as to judge the differences of passengers in satisfaction evaluation, evaluation fuzziness, and consensus. According to the results, the data can be retrospectively analysed to find the reasons for the differences and improvement measures.

3. Case study

In this paper, People's Square Station of Shanghai metro is taken as the object of case study. The People's Square Station is located in the People's Square of Huangpu District, Shanghai. Metro Line 1, Line 2 and Line 8 pass through this station. The station has 18 entrances and 18 exits, and it is the only transfer station between Line 1 and Line 2, with a large number of passengers each day.

3.1. Building an evaluation index system

The service quality of a metro station includes a series of activities such as ticket, guide, driving, inquiry, and emergency services, which have an impact on passenger perception in many aspects. Therefore, the evaluation index system of the perceived quality contains indicators of multiple dimensions. Following the principles of practicality and scientificity of the evaluation index system, 38 basic indicators are summarized in the process of literature analysis, and then a representative, reliable and easily recognizable set of initial evaluation indicators is screened out by sociological in-depth interviews. The main purpose of the questionnaire design is to collect passengers' judgment on the impact of initial indicators on satisfaction evaluation. Therefore, the 5-point Likert scale, with "1" denoting "strongly unimportant" and "5" denoting "strongly important", is used in the questionnaire. The questionnaire consists of 45 questions corresponding to 38 initial evaluation indicators and basic passenger information items.

A total of 222 questionnaires were collected, of which 206 were valid. The validity rate of the questionnaires was 92.8%, which met the statistical requirements. The critical ratio method, reliability test, and factor analysis were adopted to analyse the survey data from the aspects of identification, overall reliability, representativeness, and practical significance. The 38 initial evaluation indicators are reduced to 22, and divided into five dimensions: comfort, efficiency, convenience, timeliness, and characteristic to reflect the real perception of passengers. For perceived value, passenger expectation, passenger satisfaction, passenger complaint, and passenger loyalty, there is only one dimension. The corresponding evaluation index system of passenger satisfaction of metro station is shown in Table 1.

3.2. Survey design and data collection

The questionnaire is divided into three parts: the satisfaction survey, personal background information, and travel information. The first part is mainly to evaluate the manifest variables in the index system. For perceived quality, the question is designed to show satisfaction of passengers with various indicators. The Likert 5 scale is used, in which "1" denotes "strongly disagree" and "5" denotes "strongly agree", to measure the respondents' recognition of a particular statement. For other latent variables, the degree of passenger recognition was investigated. The second part investigates the personal

Table 1
Index system of passenger satisfaction evaluation for metro station.

Latent variable		Manifest variable	Symbol
Perceived quality	Comfort	Indoor temperature	Q1
		Ventilation and air quality	Q2
		Cleanliness in station	Q3
		Crowdedness in station	Q4
		Information-oriented signs in station	Q5
		Number, position, and width of stairs	Q6
		Number and speed of security inspection facilities	Q7
		Location and length of transfer channel in station	Q8
		Transfer information-oriented signs outside station	Q9
	Efficiency	Number and location of ticket vending machines and recharging machines	Q10
		Number, location, and form of gates at inbound and outbound station	Q11
		Number, location of gates for large baggage at inbound and outbound station	Q12
		Number and location of manual service counters	Q13
		Number and location of entrances and exits	Q14
	Convenience	Diversity of payment forms	Q15
		Number and location of convenience facilities such as ATM and vending machines	Q16
		Frequency, accuracy, and timeliness of relevant information prompted by station broadcasting	Q17
	Timeliness	Waiting time in station	Q18
		Number of idle parking spaces around the station	Q19
		Availability of shared bicycles around station	Q20
	Characteristic	Lighting color and brightness in the station	Q21
		Decoration design of roof, floor, and wall in station	Q22
Perceived value		Performance of the metro service quality relative to its fares	PV
Passenger expectation		Overall expectation of the metro service	PE
Passenger satisfaction		Overall satisfaction of the metro service	PS
Passenger complaint		Frequency of complaints about the metro service	PC
Passenger loyalty		Be willing to take the subway at this station in the future	PL

characteristics of passengers, including gender, age, income, education and so on. The third part focuses on the frequency and time of passengers taking the metro. The questionnaire survey was carried out through a combination of online and on-site surveys from March 12 to March 14, 2019. In the online survey, a screening part was added to confirm that the respondents had taken the subway at the People's Square Station within the month. The field survey used a random sampling method to select the passengers in the station.

In this survey, a combined total of 260 questionnaires were collected on the People's Square Station. 11 invalid questionnaires with incomplete or random answers were excluded. Finally, 239 questionnaires were validated, which met the requirements of 95% confidence level and 6% relative error. A minimum sample size of 200 is required to reduce bias to an acceptable level for any type of SEM estimation (Kline and Santor, 1999). Effective questionnaires accounted for 91.9% of the total sample, meeting the requirements of effective recovery rate. As the descriptive statistics of the basic characteristics of the respondents shown (Table 2), it can be seen that the distribution of the sample is reasonable. It is worth noting that because people's Square station is an administrative and business center, there are relatively more young people.

In order to test the reliability of the questionnaire, the Cronbach α coefficient method is used. The results show that the α consistency coefficients of all second-level indicators are greater than 0.7, and the overall α consistency coefficients are greater than 0.9, which indicates that the overall reliability of the questionnaire is high and has good consistency and stability. For the validity test, the value of Kaiser-Meyer-Olkin (KMO) is 0.932, and the significance is $p < 0.01$, which is very suitable for factor analysis. The results show that the commonality of each index is much greater than 0.4, which shows that the impact of each index problem on passenger satisfaction is significant. Thus the questionnaire is valid and reliable.

3.3. Evaluation indicators weight determination

Based on the questionnaire data, the SEM conceptual model was fitted and modified, and the optimal model was obtained (Fig. 2). Perceived expectation is deleted because it does not reach a statistically significant level. Perceived quality corresponds to five second-level indicators, and along with perceived value, contributes positively to passenger satisfaction, while passenger complaints contribute negatively to passenger satisfaction. Fig. 2 shows the standard path coefficients between latent variables and between latent variables and manifest variables.

NC, CFI, GFI and RMSEA are used to characterize the fitting effect of the model. NC is short for normed chi-square. As the correction and supplement of chi square value, NC value is an ideal value between 1 and 3 in strict cases. According to statistics, the comparative fit index (CFI) is generally between 0–1, and greater than 0.8 indicates that the model fitting effect is good. The goodness-of-fit index (GFI), greater than 0.9 is ideal, and greater than 0.8 is acceptable. RMSEA is the Root Mean Square Error of Approximation. The smaller the value, the better the fitting effect. Less than 0.1 indicates that the fitting result is acceptable, and less than 0.05 means very good fitting. In the optimal model, the NC is 2.38, the CFI is $0.859 > 0.8$, the GFI is $0.823 > 0.8$, and the RMSEA is $0.076 < 0.08$, which shows that the fitting result of the model is good (Hair et al., 2006). Furthermore, the path coefficients are normalized. The weights of passenger satisfaction with each first-level, second-level and third-level indicator are calculated respectively (Table 3).

3.4. Evaluation result

According to the cloud model theory, in order to ensure the error is less than 0.01, the number of cloud droplets $M = 2000$ is selected to generate a normal cloud. Based on the questionnaire design, after determining the evaluation set V = {strongly dissatisfied (1 point), slightly dissatisfied (2 points), ordinary (3 points), slightly satisfied (4 points), strongly satisfied (5

Table 2
Background characteristics of the sample.

Characteristics	Category value	%	Characteristics	Category value	%
Gender	Male	49.37	Monthly income (¥)	<3000	26.78
	Female	50.63		3000–6000	22.59
Age	<18	1.26		6001–10,000	28.87
	18–25	40.59		10,001–15,000	12.97
	26–30	25.52		15,001–30,000	5.02
	31–40	15.48		>30,000	3.77
	41–50	9.62	Occupation	Student	21.34
	51–60	4.60		Civil servant	5.44
	>60	2.93		Teacher or doctor	17.57
Frequency of Taking the Metro	>5 times a week	31.38		Employee	26.36
	3–5 times a week	36.40		Worker	10.46
	1–2 5 times a week	15.90		Self-employed	7.53
	1–3 times a month	8.37		Others	11.30
	1–11 times a year	7.95	Education	Primary or junior high school	5.44
Private car ownership	Yes	63.60		High school	14.23
	No	36.40		College and University	65.69
				Post graduate +	14.64

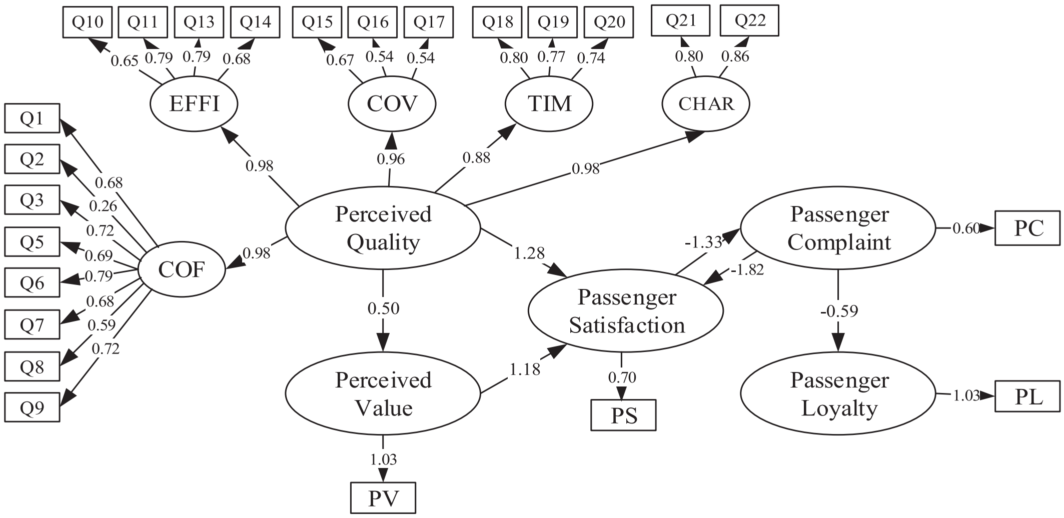

Fig. 2. Optimal SEM with standard path coefficients.

Table 3
Weight set of evaluation index system.

Objective	First-level indicators	Weights	Second-level indicators	Weights within first-level indicators	Third-level indicators	Weight within second-level indicators
Passenger satisfaction	Perceivedquality	0.384	Comfort	0.205	Q1	0.133
					Q2	0.051
					Q3	0.140
					Q5	0.134
					Q6	0.154
					Q7	0.132
					Q8	0.115
					Q9	0.140
			Efficiency	0.205	Q10	0.222
					Q11	0.272
					Q13	0.272
					Q14	0.234
			Convenience	0.202	Q15	0.381
					Q16	0.311
					Q17	0.308
			Timeliness	0.184	Q18	0.347
					Q19	0.334
					Q20	0.319
			Characteristic	0.204	Q21	0.484
					Q22	0.516
	Perceivedvalue	0.242			PV	1.000
	Passenger complaint	−0.374			PC	1.000

points)}, the cloud parameters of each benchmark evaluation were further derived: strongly dissatisfied (0.5, 0.167, 0.10), slightly dissatisfied (1.5, 0.167, 0.10), ordinary (2.5, 0.167, 0.10), slightly satisfied (3.5, 0.167, 0.10), strongly satisfied (4.5, 0.167, 0.10).

Using the backward cloud generator method, the corresponding cloud parameter matrix Z of the indicator set is generated. At the same time, combined with the obtained weight set W, the cloud parameter of the passenger satisfaction comprehensive evaluation result of People's Square station is calculated as C (3.640, 1.070, 0.145). Finally, the cloud parameters of the evaluation results of each third-level indicator and each second-level in perceived quality are shown in Table 4, and that of first-level indicators are shown in Table 5. In order to evaluate the quality of the results more intuitively, the forward cloud generator method is used to generate a normal cloud from the result cloud C and place it in the benchmark clouds (Fig. 3.). The cloud of evaluation results corresponding to each second-level indicator of perceived quality is shown in Fig. 4.

Table 4
Cloud parameters of evaluation results of perceived quality.

Third-level Indicators	E_x	E_n	H_e	Second-level Indicators	E_x	E_n	H_e
Q1	3.703	1.088	0.227	Comfort	3.766	1.142	0.172
Q2	3.042	1.149	0.270				
Q3	3.711	1.064	0.199				
Q5	3.833	1.006	0.368				
Q6	4.017	1.217	0.078				
Q7	3.607	1.315	0.035				
Q8	3.753	1.049	0.265				
Q9	3.967	1.231	0.028				
Q10	3.695	1.093	0.166	Efficiency	3.893	0.903	0.250
Q11	3.983	1.174	0.290				
Q13	3.958	1.266	0.205				
Q14	3.900	1.021	0.491				
Q15	4.038	1.080	0.325	Convenience	3.761	1.077	0.207
Q16	3.732	0.998	0.221				
Q17	3.448	1.152	0.046				
Q18	3.770	1.324	0.170	Timeliness	3.770	1.245	0.153
Q19	3.623	1.188	0.127				
Q20	3.925	1.220	0.163				
Q21	3.996	1.151	0.288	Characteristic	3.970	1.142	0.227
Q22	3.946	1.133	0.169				

Table 5
Cloud Parameters of Evaluation Results of First-level Indicators.

First-level Indicators	E_x	E_n	H_e	Objective	E_x	E_n	H_e
Perceived Quality	3.833	1.099	0.203	Passenger Satisfaction	3.640	1.070	0.145
Perceived Value	3.439	1.118	0.115				
Passenger Complaint	3.569	1.009	0.102				

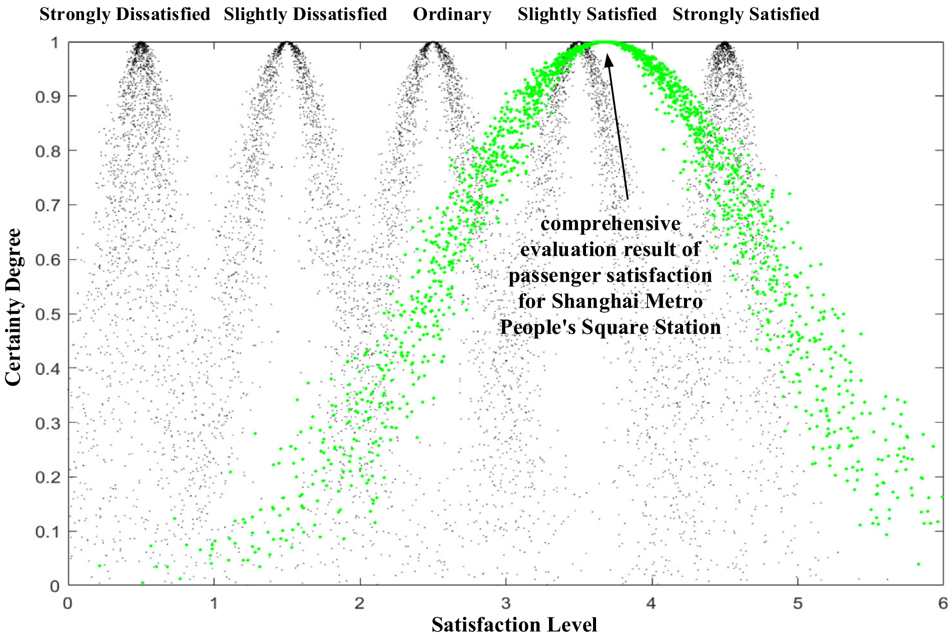

Fig. 3. Normal cloud of comprehensive evaluation result for Shanghai Metro People's Square Station.

3.5. Discussion

As shown in Fig. 3, the normal cloud of evaluation result for People's Square Station is between the benchmark cloud of "slightly satisfied" and "strongly satisfied", but closer to "slightly satisfied", which indicates that the passenger satisfaction of

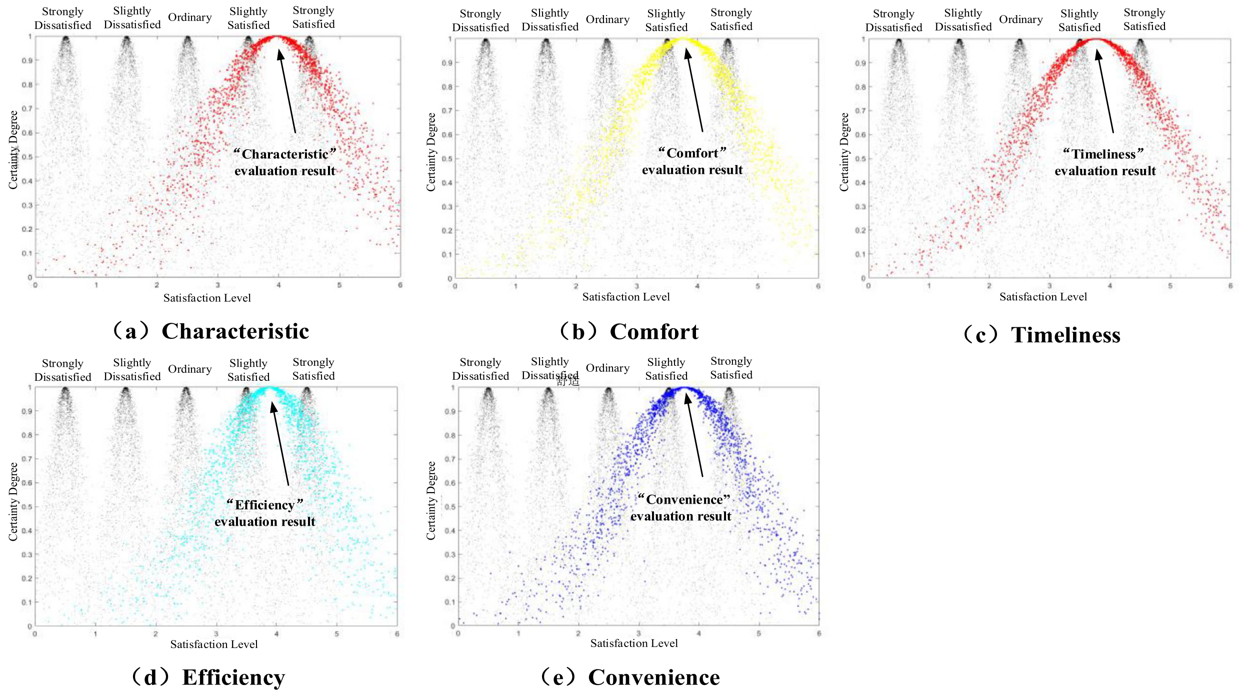

Fig. 4. Normal cloud of evaluation results of different second-level indicators in perception quality of Shanghai Metro People's Square Station.

the station is generally acceptable. However, the span of the evaluation result cloud is much larger than that of the benchmark cloud within the figure, which implies that there is a big cognitive gap of passengers' satisfaction to the service of the station. In addition, the dispersion and thickness of the evaluation result cloud are much larger than that of the benchmark cloud, indicating that different passenger groups have not reached a consensus on the evaluation results. Generally speaking, the evaluation results show that the service level of Shanghai Metro People's Square Station meets the basic expectation needs of passengers, but there is still much room for improvement.

In order to propose measures to improve the passenger satisfaction of the station, the perceived quality is further analysed. Combining with the analysis of Table 3 and Fig. 4, it can be found that the values of *Ex* of the second-level indicators do not differ much, and they are all between "slightly satisfied" and "strongly satisfied".

For the second-level indicators with strong consensus, that is, the dispersion and thickness of cloud are relatively small, the indicators can be directly improved. For example, for the "comfort" dimension, the results in Table 4 show that the evaluation of passenger satisfaction is closer to "slightly satisfied", and the cognitive gap of passengers' satisfaction is small. This indicates that most passengers agree that the People's Square station is satisfactory in terms of "comfort" service, so the satisfaction of this dimension can be improved through direct improvement of indicators. Among them, the evaluation result of the ventilation and air quality (Q2) is low and the value of *Ex* is only 3.042. Combined with the basic situation of People's Square station, its line has a long operating life and many operating stations, and the performance of ventilation and temperature control equipment at People's Square station may be aging. Therefore, it is recommended that the operator's primary task is to perform timely maintenance, repair and replacement of the ventilation equipment, in order to improve this indicator with low evaluation result and high recognition as much as possible, effectively improving the level of passenger satisfaction of the metro station.

For the weak consensus second-level indicators, further analysis of the differences is needed to make targeted improvements. For example, for the "efficiency" indicator, the characteristics of respondents with low score (<12) and high score (>16) are analysed respectively. More than 80% of the unsatisfied people are younger than 30 years old, and more than 70% of them travelled by subway during peak hours, while the satisfied population for these two groups corresponds to 60% and 50%, respectively. So we can consider increasing the number of gates at inbound and outbound station during peak hours to improve efficiency. Similarly, we can make the same analysis for other second-level indicators and give corresponding suggestions for improvement.

4. Conclusions

Metro, as one of the most significant public transport types, plays a crucial role in promoting urban development and residents travel. The passenger's requirements for metro are also constantly increasing with the improvement of living standards. Therefore, achieving a high passenger satisfaction level is vital for regulatory agencies, and it is indispensable to provide methodological guidance in all aspects of passenger satisfaction evaluation so as to create a theoretical basis for evaluating progress. From the perspective of passenger perception, this paper proposes a dynamic evaluation method for metro station. The initial set of indicators is firstly screened through surveys, consisting of 22 indicators. Then using the SEM, based on the conceptual model, the index weights are determined based on the relationship between variables. Next, we develop a passenger satisfaction evaluation model based on the cloud model. Finally, the passenger satisfaction evaluation of Shanghai Metro People's Square Station is taken as an example for application. The results show that the method proposed in this paper can effectively measure the level of passenger satisfaction at metro stations. It can not only obtain the overall satisfaction level of passengers, but also the degree of consistency of evaluation of passengers. It is possible to improve the indicators by directly or further distinguishing passenger characteristics to improve the service level at the metro station.

The correlation analysis of passenger groups and evaluation results is still not the focus of this research. An in-depth analysis is needed in future research to distinguish stratification evaluation of passengers with different characteristics. By quantifying the relationship between different passenger groups and satisfaction, and accurately pinpointing the group characteristics with lower satisfaction evaluation, we can achieve more refined management and optimization, and better improve the service quality of metro station.

Conflict of interest

All the authors have no conflict of interest with the funding entity and any organization mentioned in this article in the past three years that may have influenced the conduct of this research and the findings.

Acknowledgments

This work was funded the National Social Science Foundation of China (No.20BGL291).

References

Aydin, N., Celik, E., Gumus, A.T., 2015. A hierarchical customer satisfaction framework for evaluating rail transit systems of Istanbul. Transp. Res. Part A: Policy Pract. 77, 61–81.
Cardozo, R.N., 1965. An experimental study of customer effort, expectation, and satisfaction. J. Mark. Res. 2 (3), 244–249.
Celik, E., Aydin, N., Gumus, A.T., 2014. A multi attribute customer satisfaction evaluation approach for rail transit network: a real case study for Istanbul, turkey. Transp. Policy 36 (36), 283–293.
Del Castillo, J.M., Benitez, F.G., 2012. A methodology for modeling and identifying users satisfaction issues in public transport systems based on users surveys. Proced. Soc. Behav. Sci. 54, 1104–1114.
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 (3), 219–226.
Eboli, L., Mazzulla, G., 2009. A new customer satisfaction index for evaluating transit service quality. J. Public Transp. 12 (3), 21–37.
Fan, Q., Pu, Q., Yin, C., 2013. Comprehensive evaluation of passenger service quality for urban rail transit based on passengers'perception. Urban Mass Transit 16 (11), 49.
Fornell, C., Johnson, M.D., Anderson, E.W., Cha, J., Bryant, B.E., 1996. The American customer satisfaction index: nature, purpose, and findings. J. Market. 60 (4), 7–18.
Garrido, C., de Oña, R., de Oña, J., 2014. Neural networks for analyzing service quality in public transportation. Expert Syst. Appl. 41 (15), 6830–6838.
Golob, T.F., 2003. Structural equation modeling for travel behavior research. Transp. Res. Part B: Method. 37 (1), 1–25.
Habib, K.M., Kattan, N.L., Islam, M.T., 2011. Model of personal attitudes towards transit service quality. J. Adv. Transp. 45, 271–285.
Hair, J.F., Anderson, R.E., Tatham, R.L., Black, W.C., 2006. Pearson Prentice Hall. New Jersey. humans: critique and reformulation. J. Abnorm. Psychol. 87, 49–74.
Hau, K.T., Wen, Z., Cheng, Z., 2004. Structural equation model and its applications. Educational Science Press, Beijing.
Hensher, D.A., Stopher, P., Bullock, P., 2003. Service quality––developing a service quality index in the provision of commercial bus contracts. Transp. Res. Part A: Policy Pract. 37 (6), 499–517.
Jöreskog, K.G., 1993. Testing Structural Equation Models in Testing Structural Equation Models, KA Bollen & J. Scott Long, éds. Contemporary Sociology, 23 (1), 66-67.
Kline, R.B., 1998. Software review: Software programs for structural equation modeling: Amos, EQS, and LISREL. J. Psychoeduc. Assess. 16 (4), 343–364.
Kline, R.B., Santor, D.A., 1999. Principles and practice of structural equation modelling. Canadian Psychol 40 (4), 381–396.
Li, D., Du, Y., 2017. Artificial intelligence with uncertainty. CRC Press.
Li, L., Fu, J., Guo, X., Zhang, Y., 2018. An Integrated Approach Based on CRITIC Method and Cloud Model for Evaluating Passenger Satisfaction of Urban Rail Transit. the Transportation Research Board, Washington, D.C..
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. Transp. Policy 53, 175–185.
Mouwen, A., 2015. Drivers of customer satisfaction with public transport services. Transp. Res. Part A: Policy Pract. 78, 1–20.
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.
Tyrinopoulos, Y., Antoniou, C., 2008. Public transit user satisfaction: Variability and policy implications. Transp. Policy 15 (4), 260–272.
Wang, Y., Zhang, Z., Sun, H., 2018. Assessing customer satisfaction of urban rail transit network in Tianjin based on intuitionistic fuzzy group decision model. Discr. Dyn. Nat. Soc. 2018, 1–11.
Xiao, J., Li, F., 2011. Assessment on passengers' satisfaction about urban rail transit based on PLS. Urban MassTransit. 14 (7), 56.
Zhai, X., Weng, J., Wang, J., Mao, L., 2018. Evaluation method of bus service quality based on passenger satisfaction survey. Traffic Eng. 18 (02), 52–58.
Zhang, C., Juan, Z., Lu, W., Xiao, G., 2016. Do the organizational forms affect passenger satisfaction? Evidence from Chinese public transport service. Transp. Res. Part A Policy Pract. 94, 129–148.
Zhang, C., Liu, Y., Lu, W., Xiao, G., 2019. Evaluating passenger satisfaction index based on PLS-SEM model: evidence from Chinese public transport service. Transp. Res. Part A: Policy Pract. 120, 149–164.