IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received December 28, 2019, accepted January 8, 2020, date of publication January 13, 2020, date of current version January 21, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2965990

Improving Service Quality of Metro Systems–A Case Study in the Beijing Metro

XINYUE XU[1], YI LU[2], YINGXIANG WANG[1], JIANMIN LI[1], AND HANYU ZHANG[3]

[1]State Key Laboratory of Railway Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China
[2]Subway Operation Technology Centre, Mass Transit Railway Operation Corporation Ltd., Beijing 102208, China
[3]Georgia Institute of Technology, Atlanta, GA 30332, USA

Corresponding author: Xinyue Xu (xxy@bjtu.edu.cn)

This work was supported in part by the State Key Laboratory of Rail Traffic Control and Safety, China, under Grant 2019JBM077.

ABSTRACT In this study, we propose a method that combines Bayesian network, structural equation modeling, and importance-performance analyses to evaluate and improve the service quality of crowded metros from the point of service components. First, service components that influence the service satisfaction of crowded metros are investigated, and 7 service components and 27 factors are collected from the perspective of metro daily operation. Correspondingly, 8,011 sample questioners are collected using both online (i.e. microblog and WeChat) and face-to-face survey methods. The usage of these service components is more practical for metro managers to improve service quality than the factor analysis methods used in previous studies. Second, a hybrid method combining Bayesian network and structural equation modeling is proposed to establish the relationship between service quality and these service components and factors. This hybrid method can learn the uncertain relationship directly from the data, explain the psychological element, and pinpoint the crucial service component. It is concluded that security check and waiting for boarding have the greatest effect on the overall satisfaction. With the help of importance-performance analyses, some practical suggestions (e.g., theme stations, women-only carriages, and one-day ticket) are proposed to optimize bottleneck service components and thus improve the service quality of daily metro operation. Further, heterogeneity of service quality satisfaction for various groups of respondents is discussed, which contributes to effective strategies for satisfaction improvement.

INDEX TERMS Service quality, improvement suggestions, Bayesian networks, structural equation modeling, heterogeneity.

I. INTRODUCTION

Overcrowding in metros is emerging as a major impediment to the improvement of service quality in many countries, including China. In 2018, the average daily passenger volume of the Beijing metro reached 10.54 million, and the train loading factor of 10 lines during peak hours was over 100% according to the data from Mass Transit Railway Operation Corporation LTD. As a consequence, problems such as poor sanitation and insufficient equipment capacity occur, and these problems affect passenger travel experience and reduce their satisfaction with the Beijing metro. In response, transit agencies have taken many measures in the Beijing metro. On one hand, passenger flow control measure has been introduced during peak hours and has become increasingly common in China and other countries to ensure passenger safety [1]–[3]. While this action alleviates congestion and improves the safety from the perspective of managers, it also has the potential to extend the waiting time of some passengers and has little effect on overall passenger satisfaction. On the other hand, new lines have also been built to ease congestion and improve the satisfaction with the whole metro system. However, it takes a long time to build new lines, and the additional capacity seldom catches up with increasing demand [3], [4]. Therefore, identifying the bottleneck of service satisfaction is crucial to improve the service satisfaction of crowded metros, especially in China.

Extensive research has been performed to explore the factors that affect the quality of public transport service using methods such as factor analysis, principal component analysis, cluster analysis, discrete choice model or the multiple regression method. For example, Friman *et al.* [5]

The associate editor coordinating the review of this manuscript and approving it for publication was Young Jin Chun.

leveraged confirmatory factor analysis to examine the psychometric properties of the satisfaction with travel scale. Le-Klähn *et al.* [6] proposed principal component analysis with the varimax orthogonal rotation method to delineate the underlying dimensions that were associated with the satisfaction with public transport in Munich. De Oña *et al.* [7] developed cluster analysis to classify four groups of passengers, and developed a decision tree methodology to identify the most important service quality attributes influencing the overall evaluations by passengers. The above methods have focused on service quality evaluation from an objective perspective; however, service quality is a composite concept in metros, and the satisfaction with metro service is actually psychological [8], which makes it difficult to evaluate the service quality of metros.

An alternative psychological method to evaluate the service quality of metros is structural equation model (SEM). SEM has, nevertheless, been one of the more widely used techniques, and its application in this field has grown over recent years [9], as SEM is able to address large numbers of variables easily, both endogenous and exogenous [10]. For example, Eboli and Mazzulla [11] proposed a structural equation model to explore the impact of the relationship between global customer satisfaction and service quality attributes in public transport. De Oña *et al.* [13] proposed an SEM approach to reveal the unobserved latent aspects describing the service and the relationships between these aspects with the Overall Service Quality. Yilmaz and Ari [14] proposed SEM for delineating the factors affecting the loyalty of passengers traveling by high-speed rail between Eskisehir and Ankara. However, SEM is a confirmatory rather than an exploratory technique because the researcher needs to construct the model by defining unidirectional effects between variables [10]. Therefore, professional knowledge and skills are necessary for model-developing, which usually causes users to miss certain important relationships in the explanation of service quality [9].

Another method is the Bayesian network (BN), which is both mathematically rigorous and an intuitively understandable data analytical tool [15]. BN can determine which predictive variables are important based on the effect of the predictive variables on the target variables. Roos *et al.* [16] proposed an approach based on a dynamic BN to forecast the short-term passenger flows of the urban rail network of Paris, and this approach can provide real-time predictions even in the case of incomplete data. Li *et al.* [17] leveraged a BN to model departure time choice of metro passengers in the Beijing metro, which proved that the proposed model has higher prediction accuracy than typical discrete choice models. Zhu *et al.* [18] proposed a new modeling approach that utilizes a hybrid BN for travel decision inference, and this approach can provide better accuracy than decision tree models and nested logic models. Moreover, BN, as a graphic model, leverage the connection between nodes to indicate whether there is interaction between different nodes, which can well be combined with the SEM. As a result, a new approach combining BN and SEM has been proposed in the field of service evaluation in metros. This approach has already been applied in the health sector with outstanding results (e.g., [19], [20]). The approach combining BN and SEM has been used to evaluate the service quality of a metro [9]. Thus, this new approach is adopted in this study.

Further, a great deal of research on metro service evaluation starts from the following dimensions: accuracy, safety, cleanliness, comfort, reliability [9], [21], [22]. Díez-Mesa *et al.* [9] proposed this method to evaluate the service quality of the LRT Service of Seville (Spain) and considered the following dimensions: Accessibility, Availability of the Service, Security, etc. The BRITISH STANDARD [23] divided the quality criteria into three levels, the first level includes eight dimensions such as availability, accessibility, information, etc. However, these dimensions are based mainly on passenger perceptions and are not friendly to the decision-making of daily operators. Thus, this paper analyzes service quality from the perspective of the metro operation process such as service components (e.g., access (egress), security check, ticket purchases or recharge, card swiping, waiting for boarding, in-vehicle experience, and transfer) [24]. Moreover, each component is further subdivided according to the staff, equipment, and environment, so that the survey and analysis results can be more credible and practical, and better suggestions can be made for the managers.

The primary goal of this paper is to evaluate service quality, identify weak service components, and put forward reasonable suggestions to metro managers. This paper analyzes service quality from the perspective of passengers but also fuses these passenger perceptions into the daily process of metro operation (i.e., access (egress), security check, ticket purchases or recharge, card swiping, waiting for boarding, in-vehicle experience, and transfer). As a result, a service quality evaluation has been established, and this evaluation is the first effort to develop a process-based factor analysis for service quality in Chinese metros, as well as analyze passenger characteristics, especially age and travel frequency. Further, a hybrid method combining BN and SEM is used to determine the relationships between service components and overall service quality satisfaction, as well as the importance factor in each service component. With the help of importance-performance analyses (IPAs), some feasible suggestions are put forward to improve the service quality satisfaction with metros. Finally, heterogeneity of service quality satisfaction for various groups of respondents is discussed, which contributes to effective strategies for service quality satisfaction improvement.

The remainder of the paper is structured as follows: The basic theory of service quality analysis is detailed in Section II. The survey design and data collection are stated in Section III. The results and main findings are discussed in Section IV. Section V summarizes the significant findings and suggests possibilities for metro service evaluation and improvement.

II. THEORETICAL BACKGROUND

This section describes the basic theory of BN and SEM that is used in the evaluation method and explains the main steps of the analysis process.

A. BAYESIAN NETWORK

BN, also called the belief network, is an extension of the Bayesian method. BN is one of the most effective theoretical models in the field of uncertain knowledge representation and inference. After being proposed by Pearl [24], BN has become a research hotspot and is often used as the framework of reasoning under uncertainty that is widely used to express uncertain relationships in recent years.

Let $U = \{x_1, \ldots, x_n\}, x \geq 1$ be the set of parameters. The BN can be defined as $BN = (X, E, P)$, where X, E, P represent nodes, directed line segments, and probability, respectively.

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

where $pa(x_i)$ is the set of antecedent variables of x_i, $i = 1, 2, 3, \ldots n$.

For any BN, its joint probability distribution can be represented as follows:

$$P(U) = \prod\nolimits_{x_i \in U} p(x_i|pa(x_i)) \tag{2}$$

BN can be regarded as a graphic model, where nodes represent different factors. An arrow, representing conditional dependency, points from a parent node to a child node, which means the child node factor depends directly on the parent node factor. Each node represents a random variable.

There are two main approaches to learn the structures of BN: the automatic learning and the manual learning approach. The automatic learning approach aims to achieve the goal of structural learning through some algorithms, while the manual learning approach depends on human experience where high skill and creativity are needed. Thus, the automatic learning approach is adopted in this paper.

Moreover, the automatic learning approach consists of three types of algorithms: constraint-based learning algorithms, scored-based learning algorithms, and hybrid algorithms [12]. Different algorithms might result in different structures, and the algorithm that is selected dominates the robust character of BN. Inspired by Cugnata *et al.* [15], all these three approaches are adopt to learn the BN structure and choose a robust network structure.

B. STRUCTURAL EQUATION MODEL

SEM is a confirmatory rather than an exploratory technique, because the researcher constructs the model by defining unidirectional effects between variables [10]. SEM can establish the relationship between multiple causes and multiple potential variables in the research.

The basic SEM equation is defined in Bollen [25]:

$$\eta = \beta\eta + \Gamma\xi + \zeta \tag{3}$$

where η is an $m \times 1$ vector of endogenous latent variables, ξ is an $n \times 1$ vector of exogenous latent variables, β is an $m \times m$ matrix of path coefficients associated with η, Γ is an $m \times n$ matrix of path coefficients associated with η and ξ, and ζ is an $m \times 1$ residual vector of the equation.

SEM is composed mainly of a measurement model and a structural model. The measurement model is used to represent the relationship between observed variables and latent variables. The basic equations of the measurement model can be expressed as:

$$x = \Lambda_x\xi + \delta \tag{4}$$

$$y = \Lambda_y\eta + \varepsilon \tag{5}$$

where x is a $q \times 1$ vector of exogenous manifest variables, Λ_x is a $q \times n$ factor loading matrix for the effects of the exogenous manifest variables on exogenous latent variables, δ is a $q \times 1$ vector of measuring error, y is a $p \times 1$ vector of endogenous manifest variables, Λ_y is a $p \times m$ factor loading matrix for the effects of endogenous manifest variables on endogenous latent variables, and ε is a $p \times 1$ vector of measuring error [21].

There are many different methods to estimate the parameters of the SEM, such as Maximum Likelihood (ML), generalized least squares estimation (GLS), and weighted least square. The sample size, the complexity of SEM, and the probability distribution should be considered when selecting the estimation method [10].

C. ANALYSIS PROCESS

The proposed combination analysis method of metro service quality is composed of three stages: 1) Questionnaire design and data collection; 2) Exploring the relationship between the dimensions of service evaluation and overall satisfaction with a hybrid method combining BN and SEM; 3) Evaluation of the service quality of metros and obtaining practical suggestions with the help of IPA.

Further, the second stage is composed of two steps: BN structure learning and validation by SEM. The first step aims to extract the relationships among the service components and overall satisfaction from the data. Then, the second step is used to verify the BN structure and estimate the parameters of the final network structure. More details are given as follows.

Step 1 Learning BN Structure:

Step 1.1 Select a robust network structure

Constraint-based learning algorithms, scored-based learning algorithms, and hybrid algorithms are employed to learn the structure of BN. The occurrence frequency of arcs that exist in the BN that is obtained are recorded. The arcs with high occurrence frequency are considered the important arcs. Specifically, we keep the arcs whose occurrence frequency is more than d(i.e., a threshold) retained in the network. The algorithm that generates the BN that contains the most important arcs is the most effective algorithm. Note that if there are networks with the same number of the important

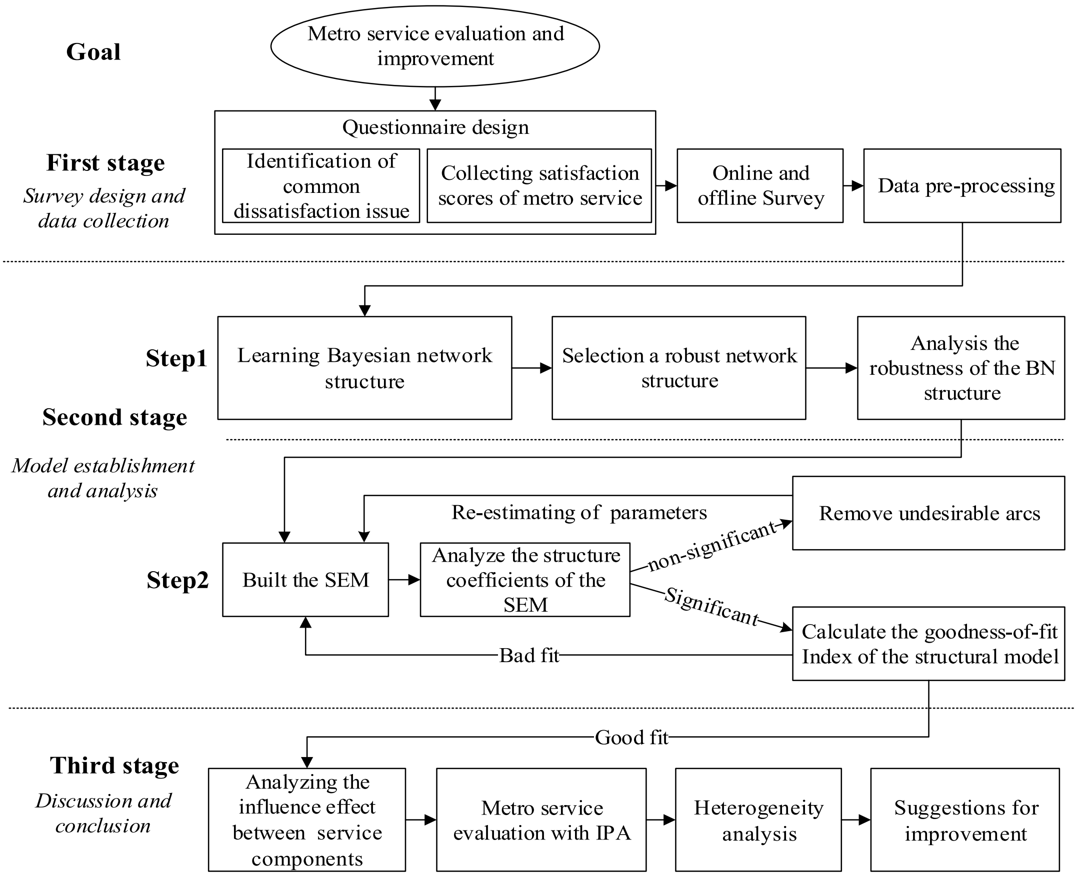


FIGURE 1. Technology roadmap.

arcs, the one with the lowest misclassification rate has been selected.

Step 1.2 Analyze the robustness of the BN structure

When the BN structure is determined, a bootstrap process is used to analyze the robustness of the chosen network. We generate 1000 random subsets with 1000 observations and learn their corresponding BN structure using the selected algorithm in step 1.1. Occurrence Frequency of each arc in these BN structures is recorded, which indicates the robustness of the dimension relationships extracted from the BN. Thus, the arcs whose occurrence frequency is significantly less than 0.5 need to be further validated next.

Step 2 Validation and Evaluation by Using SEM:

Step 2.1 Build the SEM and choose the right method (i.e., ML, GLS) to estimate parameters of the SEM;

Step 2.2 Analyze the structure coefficients of the SEM.

All relationships in the SEM should be significant at a 0.05 level of confidence. That is, the arcs whose $p \geq 0.005$ should be removed, and the parameters should be re-estimated until all relationships are significant.

Step 2.3 Calculate the goodness-of-fit parameters of the structural model:

The following indices are applied to determine the model fit according to Bollen [26]:

a. Absolute fit indices such as the goodness-of-fit index (GFI), adjusted goodness-of-fit index (AGFI), Root Mean Square Error of Approximation (RMSEA), and standardized root mean square residual (SRMSR), which indicate how well the a priori model fits the sample data [28] and demonstrate which model has the superior fit.

b. Incremental fit indices, also known as comparative [29] or relative indices [28], such as the normed fit index (NFI) and comparative fit index (CFI).

c. Parsimony fit indices, such as the parsimony goodness-of-fit index (PGFI), and the parsimonious normed fit index (PNFI).

The acceptable threshold level for each one is suggested in [27]. If some indices show bad fit, some adjustment should be made to the structure of the model. If all indices show good fit, the effects between different service components and overall-satisfaction are estimated.

The roadmap for this method is shown in Fig. 1.

III. DATA COLLECTION

A. QUESTIONNAIRE DESIGN

Based on the service process of metros [2], [3], a questionnaire is designed to investigate the influence of each service component on passenger satisfaction. The questionnaire consists of three parts: identification questions, passenger satisfaction information, and individual information.

The first part aims to exclude the following non-intended respondents: 1) people who are under the age of 18; and

(a) (b) (c)

FIGURE 2. (a) (b) Online service conducted by the official microblog and WeChat of Beijing metro. (c) Face-to-face survey inside of the metros.

2) people whose travel frequency in the Beijing metro is less than once a week in the recent three months. Note that the first part is conducted by the online survey to prevent the phenomenon of malicious filing.

The second part investigates the passengers' overall satisfaction and the passengers' satisfaction for specific service components such as security check and ticketing service. In the overall satisfaction section, there are four 11-point-scale questions where 0 indicates serious dissatisfaction and 10 indicates great satisfaction with the metro service [30]. The first three questions record the individual satisfaction from three different perspectives such as staff, equipment and facilities, and environmental sanitation. And the fourth question investigates passengers' overall satisfaction with their travel experience. Further, there are 105 binary choice questions investigating passengers' satisfaction for seven specific service components, including access (egress), security check, ticket purchases or recharge, card swiping, waiting for boarding, in-vehicle experience, and transfer experience, according to the passenger travel at metros under passenger flow control [3]. For example, for access service component, from the perspective of staff, there are three binary choice questions: whether the staff tone is blunt and impatient, whether staff members are negatively inactive during daily peak hours or not, whether the staff take timely measures to deal with emergencies or not shown in Table 1. Details of survey of other service components are shown in the appendix.

The third part is designed to collect socioeconomic and trip characteristics, for example, gender, education level, occupation, monthly income, and travel frequency. These socioeconomic and trip characteristics can assist metro managers to evaluate the group effects of passengers regarding metro service, and can contribute to taking targeted measures for different groups of passengers. Age, travel frequency (no. of trips/week), and other important characteristics are obtained in the first part.

The questionnaire is designed with in collaboration with managers from the Beijing Subway Operation Technology Centre and the Mass Transit Railway Operation Corporation LTD. A pre-survey has been conducted to validate the effectiveness of the designed questionnaire, and the final questionnaire has also been approved by its experienced managers.

B. DATA COLLECTION

The survey was conducted in the Beijing metro from October 2018 to December 2018 using online and offline methods, simultaneously, as shown in Fig. 2. Specifically, the online survey was conducted by the official microblog and WeChat of the Beijing metro [31], and the offline survey was conducted at stations that belong to the operation lines of Beijing Mass Transit Railway Operation Corporation LTD, including Line 1, Line 2, Line 5, Line 6, Line 7, Line 8, Line 9, Line 10, Line 13, Line 15, Line Fangshan, Line Batong, Line Changping, Line Yizhuang, Line Airport express, and S1 shown in Fig. 3. It was conducted in more than 40% of the metro stations, including all transfer stations and stations with high ridership. Among them, the transfer stations accounted for 43.8% of all the stations surveyed, while the stations of high ridership accounted for 71.4%. Twenty investigators were recruited to conduct the offline survey at stations or trains in the Beijing metro. Because of the time cost for answering the offline survey, the passengers at the metro system are followed by an investigator throughout their journey until the survey is completed. Further, some reward was offered to responders to collect more questionnaires both in the offline and the online survey (i.e., 2¥ for online and 5¥ for offline, respectively). In the end, a total of 8,011 valid questionnaires were collected, which makes this study one of the largest statistical analysis on service quality up till

TABLE 1. Details of survey at the access service component.

A1 Staff	whether the staff tone is blunt and impatient.	□
	whether staff members are negatively inactive during daily peak hours or not.	□
	whether the staff take timely measures to deal with emergencies or not.	□
A2 Equipment and facilities	The access queue is too long or not.	□
	Lack of awnings or awnings in access and egress of the station or not.	□
	Elevators (escalators, escalators) often break down.	□
	Unreasonable guidance marks for entering and leaving stations.	□
	Unreasonable instructions for station surrounding information.	□
	Barrier-free equipment configuration is unreasonable.	□
A3 Environmental sanitation	The sanitary condition of the ground at the entrance/passage is poor.	□
	The sanitary condition of elevator and stair in and out station is poor.	□
On a scale of 0-10, how satisfied are you with the access service component?		
Evaluation	0 indicates serious dissatisfaction and 10 indicates great satisfaction with the metro service.	
Access	0 1 2 3 4 5 6 7 8 9 10 □ □ □ □ □ □ □ □ □ □ □	

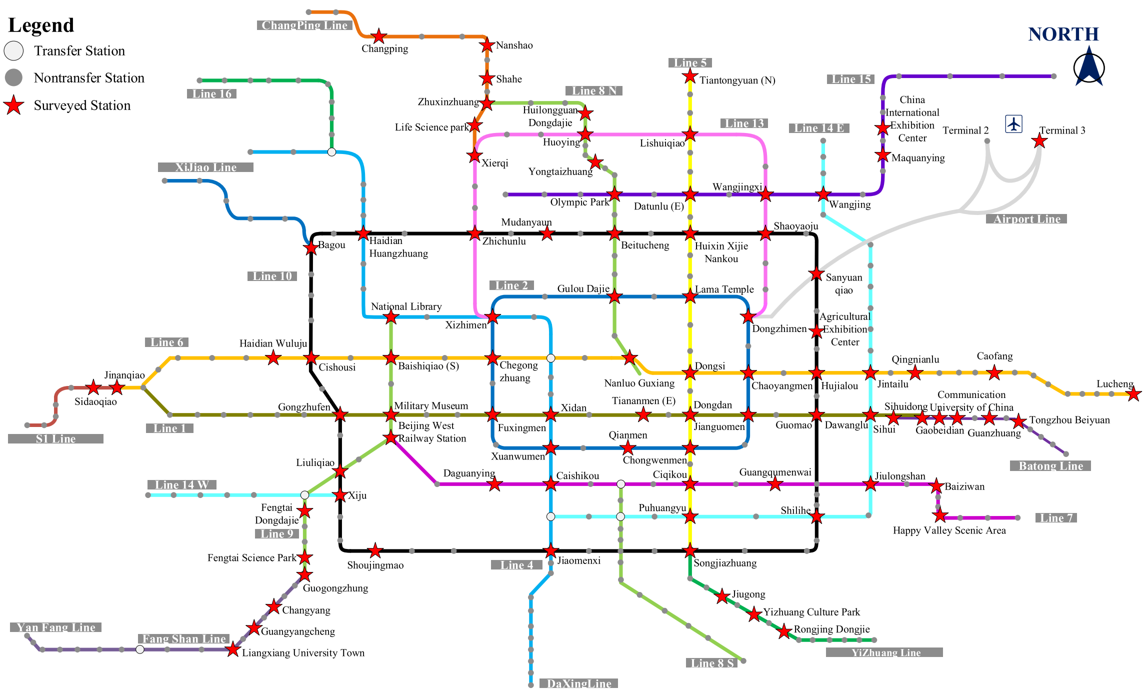


FIGURE 3. The distribution of surveyed stations in the Beijing metro system.

now comparing with other similar studies [8], [9], [21]. The socioeconomic and trip characteristic observations are shown in Table 2.

The Cronbach's alpha coefficient test for the reliability of the survey data is done and the significant value of 0.934 proves the reliability of the survey data [32]. Among all 8,011 respondents, 4,328 are male and 3,683 are female, where the gender ratio of the sample is very close to the gender ratio of the 2015 census [33], as shown in Fig. 4. Thus, we consider that the survey sample can effectively represent the whole passenger group in Beijing metro.

IV. RESULTS AND DISCUSSION

A. PRELIMINARY ANALYSIS

The relationship between socioeconomic status and trip characteristics and the overall satisfaction is discussed based on the questionnaire data. First, the relationship between age and overall satisfaction is identified by cross contingency analysis as shown in Fig. 5. The significance of the chi-squared test was found to be less than 0.005, which means age has a significant effect on overall satisfaction. Specifically, approximately 82.6% of respondents aged 18-30 years old are satisfied with the metro service (i.e., the satisfaction score

TABLE 2. Socioeconomic and trip characteristics of observations.

Socioeconomic and trip characteristics		No. of observations	Percentage
Gender	Male	4328	54.026%
	Female	3683	45.974%
Age group	18-30 years	4083	50.967%
	31-44 years	2770	34.577%
	45-60 years	953	11.896%
	>60 years	205	2.559%
Travel frequency (no. of trips weekly)	1 trip/week	1131	14.1%
	2-3 trips/week	2252	28.1%
	4-5 trips/week	1307	16.3%
	>5 trips/week	3321	41.5%
The most frequent departure time	Weekday peak hours (7:00-9:00, 17:00-19:00)	4977	62.1%
	Off-peak weekday	1642	20.5%
	Weekends/holidays	1392	17.4%
Monthly income	<2000 ¥	565	7.1%
	2001-5000 ¥	1261	15.7%
	5001-8000 ¥	2331	29.1%
	8001-15001 ¥	1826	22.8%
	>15000 ¥	674	8.4%
	Refuse to answer	1354	16.9%

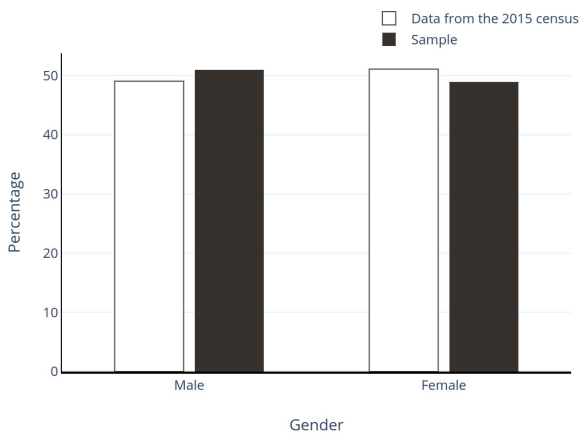


FIGURE 4. Comparison of gender rate.

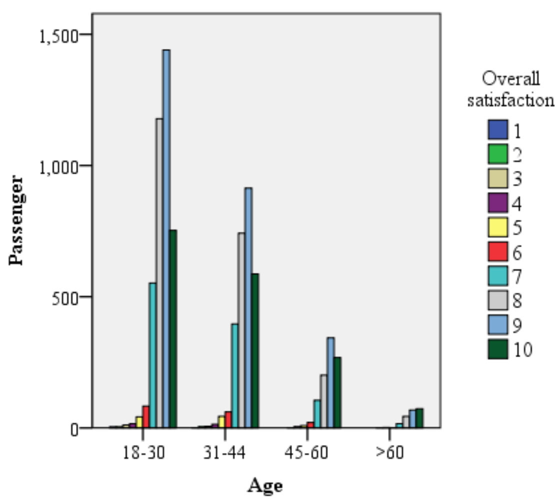


FIGURE 5. Relationship between the age and overall service quality satisfaction.

is greater than or equal to 8). Approximately 81% of the respondents aged 31-44 years are satisfied with the metro service. Approximately 85.2% of respondents aged 45-60 are satisfied with the metro service, and approximately 90.3% of the respondents aged over 60 are satisfied with the metro service. We can conclude that respondents aged between 18 and 44 have the lowest satisfaction with metro service while these respondents account for more than 80% of the whole group of respondents.

Further, the relationship between the travel frequency and overall satisfaction was also investigated by cross contingency analysis as shown in Fig. 6. The result of the chi-square test shows that travel frequency has a significant impact on overall satisfaction. Moreover, approximately 84.7% of the respondents who take the metro once a week are satisfied with

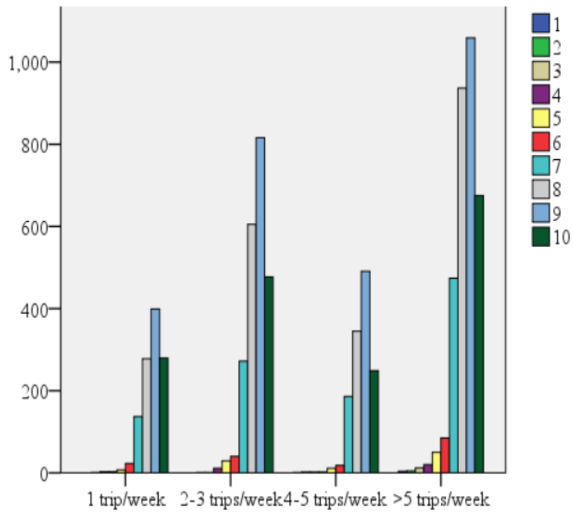


FIGURE 6. Relationship between the travel frequency and overall service quality satisfaction.

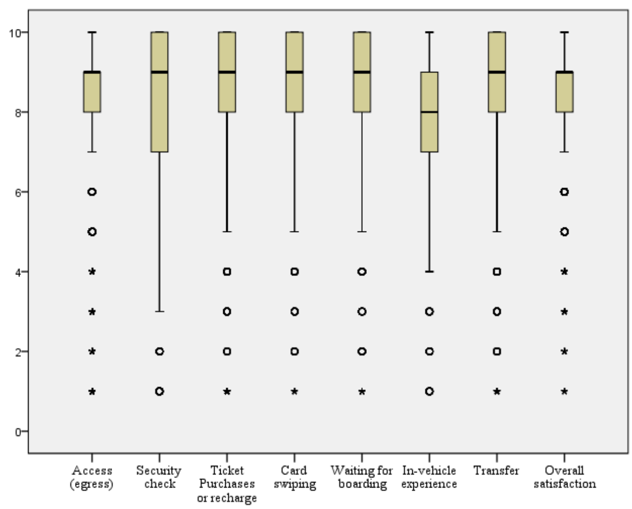


FIGURE 7. Boxplot of seven components and overall satisfaction.

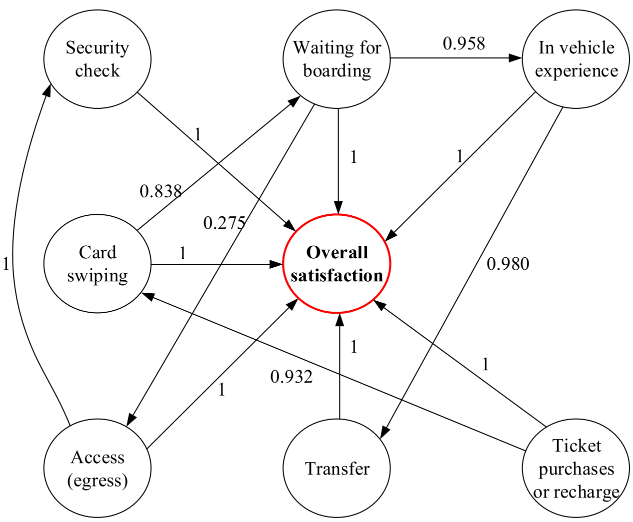


FIGURE 8. Bayesian network structure diagram and the frequency of occurrence for arc.

the metro service. Approximately 84.3% of respondents who take the metro 2-3 times a week are satisfied with the metro service. Approximately 83.1% of respondents who take the metro 4-5 times a week are satisfied with the metro service, and approximately 80.4% of respondents who take the metro above 5 times a week are satisfied with the metro service. Therefore, it is necessary to enhance the travel experience of passengers with higher travel frequency. For example, fare adjustment strategy is a potential solution [34], which can not only help passengers with higher travel frequency to save travel costs but also effectively alleviates congestion during rush hours.

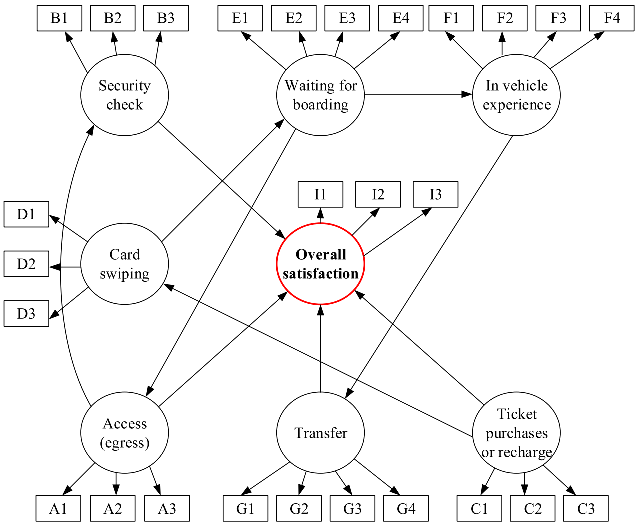
Next, a boxplot of the scores of service components and the overall satisfaction was made to provide a preliminary understanding of the Beijing metro service satisfaction as shown in Fig. 7. Most of the samples have overall satisfaction scores between 7 and 10, with a median of 9 and an average of 8.51. The result shows that the passenger satisfaction is at a good level in the Beijing metro. Moreover, the median satisfaction of the in-vehicle experience service component is low, while those of other service components are 9, suggesting that the in-vehicle experience is poor and needs to be focused on by managers. In addition, more than a quarter of the passengers scored the security check within the range of 3 and 7, which means that a significant amount of passengers are dissatisfied with the security check. Thus, security check is also a key component that needs to be improved.

B. BAYESIAN NETWORK RESULTS

In this section, 10 algorithms (including the mountain climbing algorithm, tabu search algorithm, etc.) are selected to learn the structure of the Bayesian network. The *BNLearn* package in R is used to learn the Bayesian network structure directly from the data in this case. We assume all seven service components directly impact the overall satisfaction. Therefore, the whitelist method is employed via *BNlearn* package to manually add the edges direct from seven service components to the overall satisfaction. Arcs with more than 7 occurrences, which corresponds to a threshold $d = 6.7$, in these algorithms are retained in the model according to [9], and the robust network structure is selected as the final network with the maximum number of important arcs.

After the algorithm is determined, 1000 data sets each of which consists of 1000 randomly selected data points is used to repeat the learning process for the BN structure using the corresponding algorithm (i.e., hc1) that has the most robust network structure. During the process, the frequency of occurrence of each arc in the BN structure is recorded, and the arcs with occurrence frequency below 0.5 are further verified by the structural equation to determine whether the arcs are retained. The final results are shown in Fig. 8.

We can see that all arcs have the frequency of occurrence over 0.5 except the arc directed from waiting for boarding to access (egress). Thus, this dependency relationship should be further examined by SEM. The arrows pointing to overall



Note: the box represents the element of each service component, namely, the observed variable. An ellipse represents a service component, namely, latent variable. Circles represent residuals.

FIGURE 9. Structural equation model.

satisfaction from seven service components have the frequency of occurrence of 1 due to the applied whitelist method in the BN learning. Moreover, the following conclusions can be drawn from Fig. 8.

1) Waiting for boarding affects the satisfaction of in-vehicle experience and access. The waiting time has a slightly important degree of impact on the access (egress) (with a degree of importance of 0.275), where the reason may be that congestion often appeared in these two components during peak hours and congestion will spread. As a result, their satisfaction has an interaction based on their congestion. Even though the degree of importance of the arc is low, it still should be focused on and whether to remove the arc still needs further validation.

2) Access (egress) affects the degree of satisfaction of the security check (with a frequency of occurrence 1). Because the passenger flow control measurement happens in access and security check points, where large passenger flow are accumulated during peak hours ([3], [17], [34]). Passengers can get through the station and security check quickly during nonpeak hours without passenger flow control, and then their satisfaction is independent.

C. STRUCTURAL EQUATION MODEL RESULTS

According to the BN structure obtained above, a further structural equation test was conducted using SPSS Amos. In the process of SEM, binary choice questions for each service component of the questionnaire are selected as observation variables to construct the structure. As Amos does not support the construction of SEM for binary variables, the observed variables should be re-scored. First, let each factor (e.g., A1 staff in Table 1) have 10 points and then set equally the score of the corresponding items (e.g., 3.3 score for each item as shown in Table 1). Second, if passengers are not satisfied with an item, the score of the item is 0. Finally, each item's score becomes the input of the SEM analysis.

Some structural adjustment was made to the rudimentary BN structure according to various indicators of SEM. Both arcs from card swiping, waiting for boarding and in-vehicle experience to overall satisfaction are not significant at the 0.05 confidence level. Therefore, they are removed. The final SEM structure is shown as Fig. 9.

The box in Fig.9 is shown in the following Table:

TABLE 3. The element of each service component.

Service component	The element	No. of observations
Access (egress)	A1	Staff
	A2	Equipment and facilities
	A3	Environmental sanitation
Security check	B1	Staff
	B2	Equipment and facilities
	B3	environmental sanitation
Ticket purchases or recharge	C1	Staff
	C2	Equipment and facilities
	C3	Environmental sanitation
Card swiping	D1	Staff
	D2	Equipment and facilities
	D3	Environmental sanitation
Waiting for boarding	E1	Staff
	E2	Guiding sign
	E3	Equipment and facilities
	E4	Environmental sanitation
In vehicle experience	F1	Train operation
	F2	Staff
	F3	Equipment and facilities
	F4	Environmental sanitation
Transfer	G1	Staff
	G2	Equipment and facilities
	G3	Guiding sign
	G4	Environmental sanitation
Overall satisfaction	I1	Staff
	I2	Equipment and facilities
	I3	environmental sanitation

The results of the SEM are shown in Table 4, Table 5, and Table 6. All the parameters are significant at the level of 0.005, as shown in Table 4, Table 5. Thus, the SEM is suitable for these data.

TABLE 4. The regression weight coefficient of the model was measured.

			Estimate	S.E.	C.R.	P	St
A3	<---	access(egress)	1.000				
A2	<---	access(egress)	0.566	0.011	50.984	***	0.721
A1	<---	access(egress)	0.999	0.018	55.313	***	0.691
B1	<---	security check	1.000				0.749
B2	<---	security check	0.940	0.019	48.611	***	0.744
B3	<---	security check	1.136	0.020	56.419	***	0.652
D3	<---	card swiping	1.000				0.750
D2	<---	card swiping	0.985	0.017	56.988	***	0.768
D1	<---	card swiping	1.020	0.016	63.203	***	0.715
E1	<---	waiting for boarding	1.000				0.781
E2	<---	waiting for boarding	0.771	0.013	60.270	***	0.819
E3	<---	waiting for boarding	0.731	0.012	62.225	***	0.701
E4	<---	waiting for boarding	1.064	0.016	65.902	***	0.718
F1	<---	in-vehicle experience	1.000				0.749
F2	<---	in-vehicle experience	1.158	0.019	60.809	***	0.716
F3	<---	in-vehicle experience	0.711	0.012	60.352	***	0.807
F4	<---	in-vehicle experience	0.789	0.015	51.620	***	0.801
G4	<---	transfer	1.000				0.687
G3	<---	transfer	1.225	0.022	55.143	***	0.767
G2	<---	transfer	1.158	0.027	42.506	***	0.696
G1	<---	transfer	1.310	0.021	61.453	***	0.551
C3	<---	ticket Purchases or recharge	1.000				0.764
C2	<---	ticket Purchases or recharge	0.998	0.016	60.924	***	0.789
C1	<---	ticket Purchases or recharge	0.728	0.012	62.167	***	0.746
I1	<---	overall satisfaction	1.000				**0.758**
I2	<---	overall satisfaction	1.070	0.015	69.809	***	**0.792**
I3	<---	overall satisfaction	1.025	0.015	67.818	***	**0.878**

Note: S.E represents covariance standard error; C.R represents the critical ratio; *** means P<0.001; St represents the normalized estimator.

Further, the absolute fitness index of the structural equation model is as follows: AGFI = 0.824 and GFI = 0.853 are near the recommended value 0.9. And RMSEA = 0.073 is within the range of 0.060 and 0.080, demonstrating a good fit; the comparative fit indices, CFI = 0.905 and NFI = 0.903, are close to 1, indicating good fit; Parsimony fit indices (PGFI = 0.709, PNFI = 0.807) and residual error adaptation index SRMR = $0.0410 < 0.05$ show that the model has a good degree of fit and is acceptable. Therefore, further analysis can be carried out.

The regression weight of I1 (personal overall satisfaction), I2 (equipment overall satisfaction), and I3 (environment

TABLE 5. Regression weight coefficient of structural model.

			Estimate	*S.E.*	*C.R.*	*P*	*St*
card swiping	<---	**ticket Purchases or recharge**	.936	.016	59.485	***	.979
waiting for boarding	<---	**card swiping**	.911	.014	63.300	***	.946
access(egress)	<---	**waiting for boarding**	1.102	.019	57.896	***	.930
in-vehicle experience	<---	**waiting for boarding**	.951	.016	60.433	***	.977
transfer	<---	**in-vehicle experience**	1.006	.018	56.407	***	.978
security check	<---	**access(egress)**	.937	.018	53.054	***	.968
overall satisfaction	<---	**ticket Purchases or recharge**	-.354	.035	-10.163	***	-.464
overall satisfaction	<---	**security check**	.690	.144	4.800	***	.954
overall satisfaction	<---	**transfer**	.580	.047	12.352	***	.699
overall satisfaction	<---	**access(egress)**	-.474	.152	-3.111	0.002	-.677

Note: S.E represents covariance standard error; C.R represents the critical ratio; *** means P<0.001; St represents the normalized estimator.

TABLE 6. Standardized total effects.

	ticket purchases or recharge	*card swiping*	*waiting for boarding*	*in-vehicle experience*	*access(egress)*	*transfer*	*security check*	*overall satisfaction*
card swiping	.979	.000	.000	.000	.000	.000	.000	.000
waiting for boarding	.926	.946	.000	.000	.000	.000	.000	.000
in-vehicle experience	.905	.924	.977	.000	.000	.000	.000	.000
access(egress)	.862	.880	.930	.000	.000	.000	.000	.000
transfer	.885	.904	.955	.978	.000	.000	.000	.000
security check	.834	.851	.900	.000	.968	.000	.000	.000
overall satisfaction	.367	**.848**	**.897**	**.684**	.245	.699	.954	.000

overall satisfaction) on overall satisfaction was 0.758, 0.792, and 0.878, respectively, as shown in Table 4.

The standardized total effects among service components and overall satisfaction are shown in Table 6. Note that the arrows of these three service components (i.e., card swiping, in-vehicle experience, and waiting for boarding) pointing to overall satisfaction have been removed in the SEM compared with the BN structure (see Fig.8 and Fig.9). However, these three components still affect overall satisfaction through other components (0.848, 0.897 and 0.684) illustrated by Table 6. In short, all the seven service components have positive effect on the overall satisfaction, which means the improvement of any individual component will contribute to the improvement of overall satisfaction. Further, key effects are discussed in detail, as follows.

1) The security check has the biggest impact on the overall satisfaction (total effect: 0.954) among all service components. Further observation shows that the staff's attitude and availability of staff has the biggest impact on the satisfaction with the security check service component. Recently the main dissatisfaction part is the lack of security-check equipment, and approximately 34.2% respondents considered that the security check takes too much time. There are similar conclusions with the preliminary analysis, as well as the result of

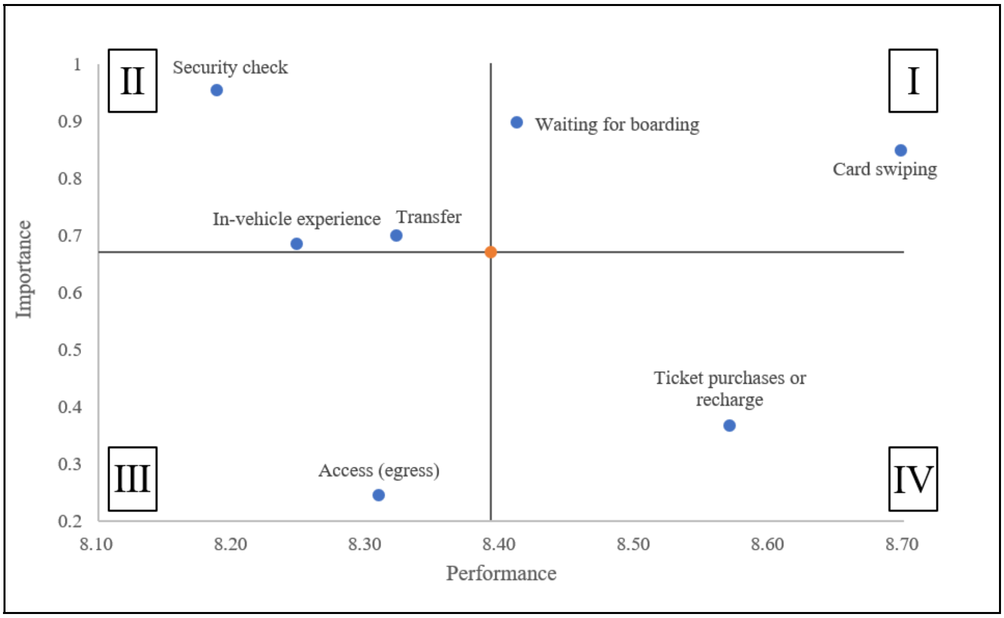


FIGURE 10. IPA matrix of satisfaction.

Zheng et al., (2018) that 46% of the passengers at the Beijing south station needed to wait for more than 4 people to security check during weekday peak hours.

2) Waiting for boarding has the second largest impact on the overall satisfaction (total effect: 0.897). The most influential part of waiting for boarding is the service equipment and facilities: approximately 25.07% of the passengers complained about lack of seats in waiting area, 17.26% considered the temperature in the station is not comfortable, and 16.46% thought the noise was too loud in the station. Moreover, about 30.42% passengers complained that the waiting time is too long for boarding due to the overcrowding. One reason is that passengers have to be left behind because of insufficient train capacity and thus wait for a few more trains during peak hours in China, which has been got considerable attention [36]–[38]. Another reason is that previous studies also prove that passengers are more sensitive to waiting time than in-vehicle time [3], [39].

3) The card swiping service component has the third biggest influence on the overall satisfaction (total effect: 0.848). Specifically, the reliability of card readers has the largest impact on the service satisfaction, and the existing problems include: sensitivity of near-field communication and failure to fix card readers in time. The reliability of card readers directly affects the efficient access of passengers. There were 18.18% of the passengers (the biggest in this component) complain about the bad sensitivity of near-field communication with mobile phones. Metro systems are currently using various new technologies, for example, entering station by QR code and near-field communication [40], [41]. Improving the reliability of these new technologies will result in an improvement of the card swiping service component.

D. DISCUSSION

To make some improvement suggestions, IPA is applied to explore the survey data as shown in Fig. 10. Note that herein the performance is the average satisfaction score of seven service components and overall satisfaction, and the importance is Standardized Total Effects between 7 service components and overall satisfaction. Targeted improvement measures can be proposed as follow.

1) Waiting for boarding and card swiping perform well and have great importance, because these service components are in the first quadrant [21]. The finding in Section 4.3 shows an anxiety of waiting for boarding. To deal with this problem, the decoration can be increased in the platform of the Beijing metro where each station layout has a certain cultural theme, increasing the ornamental value of the platform, which has been done well in several new lines in the Beijing metro (e.g., line 8).

2) Security check, in-vehicle experience, and transfer have great importance while showing worse performance, because these service components are in the second quadrant. As a result, improving these services components will result in a significant improvement of overall satisfaction. Because there is no more space can be used to locate more security or transfer facilities due to the small size of the metro stations in Beijing, the improvement of security check experience and ease the anxiety of passengers while waiting is a practical way to improve the worse performance. The first suggestion is to reduce the height of security equipment so that passengers with large pieces of luggage can pass through the security machines more easily, which is consistent with the results obtained by Eboli and Mazzulla [22]. In addition, the signage of the transfer channel and transfer streamline should be optimized as 25.79% of passenger complained

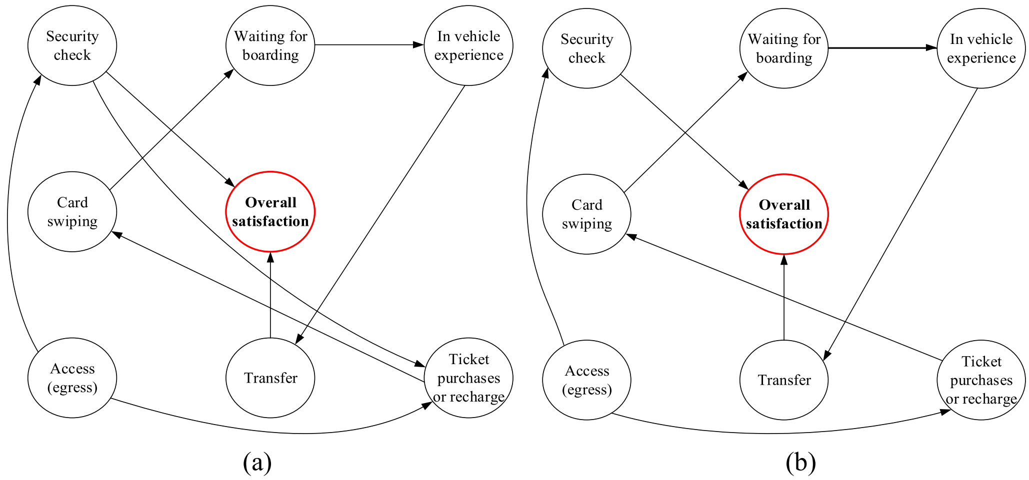


FIGURE 11. (a) The result of respondents with low frequency (b) The result of respondents with high frequency.

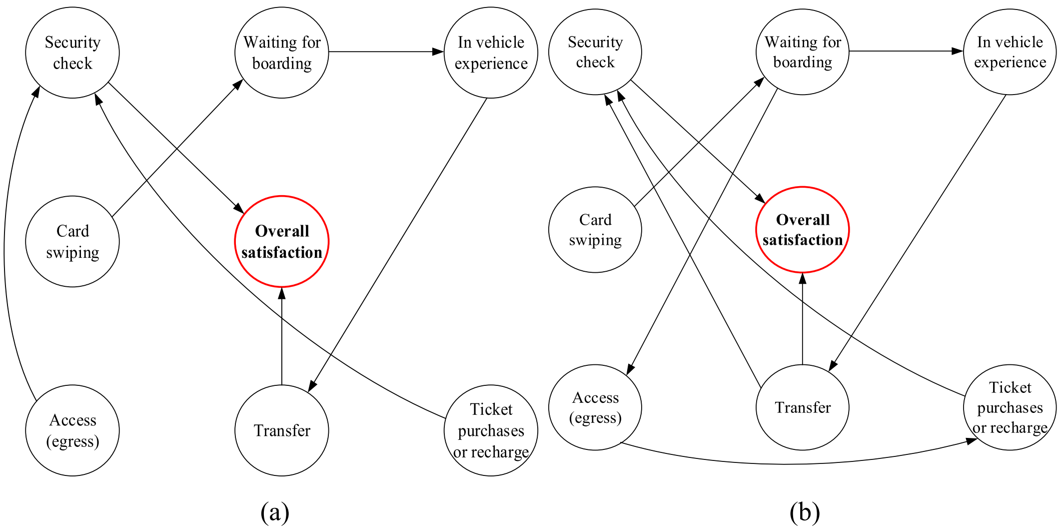


FIGURE 12. (a) The result of respondents who usually take metros during weekday peak hours (b) The result of respondents who usually take metros during off-peak weekday.

the transfer channel is too crowded and 17.91% of passengers complained that the signage was confusing, which has been also investigated by some scholars in recent years [42]. Moreover, according to the questionnaire results, 30.6% of female passengers thought the carriage is too crowded, and 18.16% of female passengers thought the environment in the carriage is not comfortable. Both of these proportions are higher than that of male passengers, therefore women-only carriages is put forward to improve the satisfaction of the in-vehicle experience component, which has been implemented in China [43] and other countries [44].

3) Ticket purchases or recharge performs well while it is not so important, because the service component is in the fourth quadrant. The service needs only to be further maintained. In further investigation, we find that some passengers have a demand for one-day ticket, which is also a good way to sell tickets in a tourist city. This suggestion has been accepted and put into trial operation in Beijing [45].

E. HETEROGENEITY OF RESULTS ACCORDING TO RESPONDENTS' TRIP CHARACTERISTICS

In this section, the effects of respondents' trip characteristics are investigated to capture potential heterogeneity in the model estimates. Because the cross-contingency analysis showed that travel frequency has a significant impact on overall satisfaction, the study also compared the results of the proposed hybrid method according to travel frequency and the most frequent departure time. The structures and model fit indices are shown in Fig.11, Fig.12, Table 7, and Table 8.

As we can see from Table 7, the index of four SEMs show that all the models have a good degree of fit and is acceptable (i.e., all indexes such as AGFI, RMSEA, CFI, NFI, PGFI, PNFI, and SRMR are satisfied with their thresholds as shown in Section IV.C). Therefore, further analysis can be carried out.

According to the result shown in Table 8, top three service which effect the overall satisfaction most is chosen to analyze. Security check (total effect: 0.400), access (egress)

TABLE 7. Goodness of fit measures of SEM.

	Low frequency (1-5 times per week)	High frequency (>5 times per week)	Weekday peak hours (7:00-9:00, 17:00-19:00)	Off-peak weekday
GFI	0.899	0.894	0.904	0.865
AGFI	0.877	0.870	0.881	0.837
RMSEA	0.060	0.063	0.061	0.069
SRMR	0.0345	0.0435	0.0419	0.0437
NFI	0.937	0.913	0.934	0.901
CFI	0.942	0.921	0.938	0.915
PGFI	0.740	0.726	0.727	0.718
PNFI	0.831	0.799	0.809	0.806

TABLE 8. Standardized total effects between seven service components and overall satisfaction of different respondents.

Respondents with different trip characteristics	*Ticket purchases or recharge*	*Card swiping*	*Waiting for boarding*	*In-vehicle experience*	*Access (egress)*	*Transfer*	*Security check*
Low frequency (1-5 times per week)	0.195	0.198	0.208	0.209	**0.383**	**0.217**	**0.400**
High frequency (>5 times per week)	**0.464**	**0.481**	0.375	0.332	**0.537**	0.392	0.184
Weekday peak hours (7:00-9:00, 17:00-19:00)	0.057	0.256	**0.272**	**0.277**	0.165	**0.274**	0.227
Off-peak weekday	0.336	**0.464**	**0.468**	0.157	0.335	0.161	**0.516**
All respondents	0.367	**0.848**	**0.897**	0.684	0.245	0.699	**0.954**

Note that in this study, respondents who usually travel on weekend were not considered, as the result of BN is not good enough due to small sample size.

(0.383), and transfer (0.217) affects passengers with low travel frequency the most. Access (egress) (0.537), card swiping (0.481), and ticket purchases or recharge (0.464) affects passengers with high travel frequency the most. The importance of security check has obvious difference between two group of passengers, one possible reason is that passengers with high frequency are accustomed to security check. Another finding is that transfer, and waiting for boarding is more important than in-vehicle experience for high frequency passengers, the reason is that high frequency passengers consider walking time and waiting time (i.e., the forth, sixth, and seven column) is more important than others such as in vehicle time (i.e., the fifth column) in the passengers' travel decision, which is in line with the conclusion by Raveau et al., (2014).

Passengers who usually take metro during weekday peak hours pay less attention to ticket purchases or recharge (0.057), this is because they seldom need to buy single journey ticket. They pay more attention to in-vehicle experience (0.277), transfer (0.274), and waiting for boarding (0.272). While passenger who usually take metro during off-peak weekday are different, and they pay more attention to security check (0.516) and waiting for boarding (0.486). There is a big difference in the importance of security check in different group of passengers, where the reason is that the first group passengers need to queue up to pass security check during rush hours, and they pay more attention to saving time instead of service quality [17].

Existing research mainly focus on the satisfaction of all passengers to enhance the whole level of service, while few studies consider passenger heterogeneity. In response to this problem, some key targeted suggestions can be put forward as follows:

1) For passengers with high travel frequency, face-scanning can take the place of card swiping service with the help of app and big data analysis tools, which also contribute to build the intelligent metro station. In addition, monthly ticket and promoting online payment with the use of app are good solutions for frequent recharge.

2) During peak hours, more attention should be paid to transfer and waiting for boarding service components, while more attention should be paid to waiting for boarding and security check service components during off-peak hours. Some guiding measures like giving a fare discount for travel 20 minutes before or after the rush hour, should be taken to relieve congestion, which has been studied by Subway [31]. Moreover, foldable seats can be added to the platform to meet the non-peaking hour needs of waiting passengers to response the complaint in Section IV.C (i.e., approximately 25.07% of the passengers complained about lack of seats in waiting area). Further, to improve the service quality of transfer during weekday peak hour, transfer streamlining optimization and improved guidance signage are necessary, which is in line with Zhang *et al.* [40].

V. CONCLUSIONS

This study evaluated the service quality of the crowded Beijing metro using a hybrid method combining BN, SEM, and IPA. BN was used as an exploratory tool to extract the relationships between different service components and overall satisfaction directly from data. SEM was used as a confirmatory tool due to its superiority. The final model was formed with 8 latent variables (i.e., access (egress), security check, ticket purchases or recharge, card swiping, waiting for boarding, in-vehicle experience, transfer, and overall satisfaction) and various observed variables. By this hybrid method, the total effect (importance) of each service component on overall satisfaction can be obtained according to the total of 8,011 samples in the Beijing metro, as well as the total effects between service components. With the help of IPA, some feasible suggestions to improve service level are provided for Beijing metro operators.

1) Passengers pay more attention to security check (total effect: 0.954) and are more sensitive to passing time (e.g., approximately 34.2% respondents considered that the security check takes too much time) in the Beijing metro. That is, passengers are sensitive to saving travel time in a crowded metro, which is line with the previous study [34].

2) Waiting for boarding has the second largest impact on the overall satisfaction (total effect: 0.897). About 30.42% passengers complained that the waiting time was too long for boarding due to the overcrowding. The dissatisfaction in waiting for boarding has been got considerable attention ([36]–[38], [47]), and design rich themes at station platforms is put forward to solve the problem, which has been done well in new lines in the Beijing metro (e.g., line 8).

3) Security check, in-vehicle experience, and transfer have great importance while showing worse performance, according to IPA. One suggestion is to reduce the height of security equipment so that passengers with large pieces of luggage can pass through the security machines more easily, which is consistent with the results obtained by Shen *et al.* [21]. Optimization transfer streamline and transfer signage is another choice, which has been investigated by previous studies [42]. Another suggestion is women-only carriages to improve the women's satisfaction of the in-vehicle experience component, which has been implemented in China [43] and other countries [44].

4) Key service components to improve the overall service quality are identified for different groups of passengers. For example, card swiping is more important for passengers with high frequency (total effect: 0.481, ranked second) than others (total effect: 0.198, ranked sixth). Waiting for boarding is important for both group of passengers, but their interesting points are different: the first group complained about overcrowding, and the second group complained seats are scarce in platforms. According to these results, targeted suggestions are given, such as introducing the face-scanning technology to take the place of card swiping for passenger with high frequency, offering a fare discount to guide passengers through adjusting their original departure time during weekday peak hours [34], and adding extra foldable seats at platforms during non-peak hours.

The findings in this research will contribute to analyzing the mechanism of service quality and providing theoretical basis for improvement strategies in metros. The following aspects will be further studied in the future:

1) The potential improvement effect of the proposed suggestions needs to be further verified.

2) Fusion more multiple data mining passengers' satisfaction is another possible area of future research.

3) It is necessary to fully consider the satisfaction of passengers of all age groups for the metro to improve the service quality of metro.

4) Exploring the heterogeneity of service quality satisfaction for various groups of respondents should be supported by different SEM models, such as SEM-MIMIC, SEM Multi-group, etc. [48], [49].

APPENDIX
DETAILS OF SURVEY OF EACH SERVICE COMPONENT

B. Security check		
B1 Staff	Whether the appearance of security personnel is not standard.	□
	Whether the staff work in a stiff tone and unfriendly manner.	□
	Whether the security inspectors are not serious and have missed the inspection and played with mobile phones.	□
B2 Equipment and facilities	The security machine often breaks down.	□
	There are few security devices and the speed of security check is slow.	□
B3 Environmental sanitation	The conveyor belt of the security inspection machine is in poor sanitary condition, with dust, stains, etc.	□
	The health status of security inspection area is poor.	□
C. Ticket purchases or recharge		
C1 Staff	Whether the appearance of security personnel is not standard.	□
	Self-service ticket or recharge needs help, but no one can be found.	□
	There are too few ticket windows and the waiting time is long.	□
	Ticket window staff are not proficient in business and slow.	□
	The staff at the ticket window had a negative attitude and a stiff tone.	□
C2 Equipment and facilities	The number of self-service ticket vending (recharging) machines is insufficient, resulting in queuing.	□
	The self-service ticket vending (recharging) machine often breaks down and is not maintained in time, resulting in queuing.	□
	The interface of the self-service ticket machine is not clear and easy to operate.	□
	There are few channels for self-service ticket sales and it is not convenient to purchase tickets.	□
C3 Environmental sanitation	The sanitation condition of ticket window and surrounding area is poor.	□
	The appearance of the self-service ticket (recharge) machine and the sanitation of the surrounding areas are poor.	□
D. Card swiping		
D1 Staff	When swiping in and out of the station, you need help but no one can be found.	□
	Staff's attitude is not positive and negative.	□
	The staff were blunt and impatient.	□
D2 Equipment and facilities	The number of card swiping machines is insufficient, which affects the entry and exit of the station.	□
	The card swiping machine often breaks down, and the maintenance is not timely, which affects the entry and exit of the station.	□
	When swiping the card, the mobile phone is insensitive and has many faults.	□
D3 Environmental sanitation	The appearance of the card swiping machine is in poor sanitary condition.	□
	The sanitation condition around the card swiping machine is poor.	□
E. Waiting for boarding		
E1 Staff	Staff appearance is not standard.	□
	The staff were blunt and impatient when asking questions.	□
	In the peak period of daily life, the personnel are not timely and passive.	□
	In case of emergency (train delay, etc.), the response measures are not timely.	□
E2 Guidance sign	The query information of subway network route is lack and hard to find.	□
	The guide signs (ride and transfer) in the station are missing, the location is unreasonable and hard to see.	□
	The information of guide signs (ride and transfer) in the station is not clear and easy to understand.	□
	The surrounding information (public transport, building) outside the station is	□

	missing, the location is unreasonable and hard to see.	
	The peripheral information (bus, building) outside the station is not clear and easy to understand.	□
	Service facility signs (elevators, toilets, etc.) are missing, incoherent or unclear.	□
E3 Equipment and facilities	Service facilities (elevators, toilets, etc.) often fail.	□
	The train doors and screen doors often fail.	□
	The station video and broadcast information are too frequent, and the content is not prominent.	□
	The volume of station video and broadcast information is too large or small.	□
	The number of waiting seats in the station is insufficient.	□
	The types and quantity of commercial service facilities are relatively small.	□
	The vending machine or ATM machine in the station often breaks down and the maintenance is not timely.	□
	The number of ATMs in the station is insufficient and the bank coverage is too small.	□
	The train is too old and uncomfortable.	□
	The food in the vending machine was not replaced in time.	□
	The number of bins in the station is insufficient.	□
E4 Environmental sanitation	The sanitation condition in the station (ground, seat, etc.) is poor.	□
	The temperature in the station is high or low, uncomfortable.	□
	Sometimes the communication signal in the station is not good.	□
	The waiting environment is simple and monotonous.	□
	There is a lot of noise in the station.	□
E5 Toilet	Toilets often fail to function.	□
	The equipment (faucet, hook, flushing, etc.) in the toilet cannot be used, and there is no maintenance.	□
	There is no toilet paper or hand sanitizer in the toilet.	□
	The sanitary condition of the toilet floor or wall is poor, and the cleaning is not timely.	□
	There are few toilet facilities and long waiting time.	□
	The toilet has poor ventilation and strong peculiar smell.	□
F. In vehicle experience		
F1 Train operation	The first and last train departure times are unreasonable, too late or too early.	□
	The train interval (waiting time) is too long in off peak period.	□
	Due to the congestion, the waiting time is too long.	□
	Information release of emergencies (such as faults, weather, etc.) is not timely, and there are few channels.	□
	The train runs unevenly and brakes hard frequently.	□
F2 Staff	In case of emergency, the driver's broadcasting is not timely and clear, and the tone is stiff.	□
	In case of emergency, the driver's broadcast volume is too large or too small.	□
	The stewards in the carriage are not active and passive.	□
F3 Equipment and facilities	The information such as indicator lights and signs in the carriage is not easy to see or clear.	□
	The equipment, facilities and signs of the old and the young in the carriage are missing or hard to see.	□
	The volume of broadcast information in the carriage is too high or low.	□
	The broadcast information in the carriage is too frequent and the content is not prominent.	□
	The emergency safety equipment (safety hammer, safety call, etc.) in the carriage	□

	is missing.	
	The emergency safety signs in the carriage are not clear.	□
	The electronic display equipment in the carriage often fails.	□
F4 Environmental sanitation	The sanitary condition of seats or ground in the carriage is poor.	□
	The temperature in the carriage is too high or too low, uncomfortable.	□
	The noise of train operation is large.	□
	The carriage video volume is unreasonable.	□
	The video content of the carriage is boring, and the update is too slow.	□
	Sometimes the communication signal in the carriage is not good.	□
	The content or form of station advertisement (including public service advertisement) is not appropriate.	□
	There are often uncivilized behaviors such as loud noise in the carriage.	□
	There is a peculiar smell in the carriage.	□
	There are often fights, eating and other uncivilized behaviors in the carriage.	□
	There are often beggars and entertainers in the carriage.	□
G. Transfer		
G1 Staff	During the peak period, the transfer and facilitation measures are unreasonable, and congestion often occurs.	□
	In case of emergency (falling, crowding, etc.), the response measures are not timely.	□
G2 Equipment and facilities	The transfer process is too long.	□
	The transfer facilities are few and inconvenient.	□
	The transfer channel is narrow and crowded.	□
G3 Guidance sign	The transfer guide signs (ride and transfer) are missing, the location is unreasonable and hard to see.	□
	The information of transfer guide signs (ride and transfer) is not clear and easy to understand.	□
G4 Environmental sanitation	The sanitary condition of the floor or wall of the transfer passage is poor.	□
	The temperature in the transfer passage is too high or too low, uncomfortable.	□
	There is a peculiar smell in the transfer passage.	□

REFERENCES

[1] S. Sadhukhan, U. K. Banerjee, and B. Maitra, "Commuters's willingness-to-pay for improvement of transfer facilities in and around metro stations—A case study in Kolkata," *Transp. Res. A, Policy Pract.*, vol. 92, pp. 43–58, 2016.

[2] X.-Y. Xu, J. Liu, H.-Y. Li, and M. Jiang, "Capacity-oriented passenger flow control under uncertain demand: Algorithm development and real-world case study," *Transp. Res. E, Logistics Transp. Rev.*, vol. 87, pp. 130–148, Mar. 2016.

[3] X. Xu, H. Li, J. Liu, B. Ran, and L. Qin, "Passenger flow control with multi-station coordination in subway networks: Algorithm development and real-world case study," *Transportmetrica B, Transp. Dyn.*, vol. 7, no. 1, pp. 446–472, Dec. 2019.

[4] X.-Y. Xu, J. Liu, H.-Y. Li, and J.-Q. Hu, "Analysis of subway station capacity with the use of queueing theory," *Transp. Res. C, Emerg. Technol.*, vol. 38, pp. 28–43, Jan. 2014.

[5] M. Friman, S. Fujii, D. Ettema, T. Gärling, and L. E. Olsson, "Psychometric analysis of the satisfaction with travel scale," *Transp. Res. A, Policy Pract.*, vol. 48, pp. 132–145, Feb. 2013.

[6] D.-T. Le-Klähn, M. Hall, and R. Gerike, "Analysis of visitor satisfaction with public transport in Munich," *J. Public Transp.*, vol. 17, no. 3, pp. 68–85, Mar. 2015.

[7] J. De Oña, R. De Oña, and G. López, "Transit service quality analysis using cluster analysis and decision trees: A step forward to personalized marketing in public transportation," *Transportation*, vol. 43, no. 5, pp. 725–747, Sep. 2016.

[8] C. Zhang, Y. Liu, W. Lu, and G. Xiao, "Evaluating passenger satisfaction index based on PLS-SEM model: Evidence from Chinese public transport service," *Transp. Res. A, Policy Pract.*, vol. 120, pp. 149–164, Feb. 2019.

[9] F. Díez-Mesa, R. De Oña, and J. De Oña, "Bayesian networks and structural equation modelling to develop service quality models: Metro of Seville case study," *Transp. Res. A, Policy Pract.*, vol. 118, pp. 1–13, Dec. 2018.

[10] T. F. Golob, "Structural equation modeling for travel behavior research," *Transp. Res. B, Methodol.*, vol. 37, no. 1, pp. 1–25, Jan. 2003.

[11] L. Eboli and G. Mazzulla, "Service quality attributes affecting customer satisfaction for bus transit," *J. Public Transp.*, vol. 10, no. 3, pp. 21–34, Apr. 2015.

[12] J. De Oña, R. O. Mujalli, and F. J. Calvo, "Analysis of traffic accident injury severity on Spanish rural highways using Bayesian networks," *Accident Anal. Prevention*, vol. 43, no. 1, pp. 402–411, Jan. 2011.

[13] J. De Oña, R. De Oña, L. Eboli, and G. Mazzulla, "Perceived service quality in bus transit service: A structural equation approach," *Transp. Policy*, vol. 29, pp. 219–226, Sep. 2013.

[14] V. Yilmaz and E. Ari, "The effects of service quality, image, and customer satisfaction on customer complaints and loyalty in high-speed rail service in Turkey: A proposal of the structural equation model," *Transportmetrica A, Transp. Sci.*, vol. 13, no. 1, pp. 67–90, Jan. 2017.

[15] F. Cugnata, R. S. Kenett, and S. Salini, "Bayesian networks in survey data: Robustness and sensitivity issues," *J. Qual. Technol.*, vol. 48, no. 3, pp. 253–264, Jul. 2016.

[16] J. Roos, G. Gavin, and S. Bonnevay, "A dynamic Bayesian network approach to forecast short-term urban rail passenger flows with incomplete data," *Transp. Res. Procedia*, vol. 26, pp. 53–61, 2017.

[17] X. Li, H. Li, and X. Xu, "A Bayesian network modeling for departure time choice: A case study of Beijing subway," *PROMET-Traffic Transp.*, vol. 30, no. 5, pp. 579–587, Nov. 2018.

[18] Z. Zhu, X. Chen, C. Xiong, and L. Zhang, "A mixed Bayesian network for two-dimensional decision modeling of departure time and mode choice," *Transportation*, vol. 45, no. 5, pp. 1499–1522, Sep. 2018.

[19] C. W. Duarte, Y. C. Klimentidis, J. J. Harris, M. Cardel, and J. R. Fernandez, "A hybrid Bayesian network/structural equation modeling (BN/SEM) approach for detecting physiological networks for obesity-related genetic variants," in *Proc. IEEE Int. Conf. Bioinf. Biomed. Workshops (BIBMW)*, Nov. 2011, pp. 696–702.

[20] F. Trentini, M. Malgaroli, A. L. Camerini, C. Di Serio, and P. Schulz, "Multivariate determinants of self-management in health care: Assessing health empowerment model by comparison between structural equation and graphical models approaches," *Epidemiol., Biostatist. Public Health*, vol. 12, no. 1, pp. 1–13, 2015.

[21] W. Shen, W. Xiao, and X. Wang, "Passenger satisfaction evaluation model for Urban rail transit: A structural equation modeling based on partial least squares," *Transp. Policy*, vol. 46, pp. 20–31, Feb. 2016.

[22] L. Eboli and G. Mazzulla, "A methodology for evaluating transit service quality based on subjective and objective measures from the passenger's point of view," *Transp. Policy*, vol. 18, no. 1, pp. 172–181, 2011.

[23] *Transportation—Logistics and Services—Public Passenger Transport—Service Quality Definition, Targeting and Measurement*, Standard EN 13816, CEN, 2002, pp. 1–37.

[24] J. Li, X. Xu, Z. Yao, and Y. Lu, "Improving service quality with the fuzzy TOPSIS method: A case study of the Beijing rail transit system," *IEEE Access*, vol. 7, pp. 114271–114284, 2019.

[25] J. Pearl, *Probabilistic Reasoning in Intelligent Systems*. Amsterdam, The Netherlands: Elsevier, 1988.

[26] K. Bollen, *Structural Equations With Latent Variables*. New York, NY, USA: Wiley, 1989.

[27] D. Hooper, J. Coughlan, and M. Mullen, "Structural equation modelling: Guidelines for determining model fit. Electronic journal of business research methods," *Electron. J. Bus. Res. Methods*, vol. 6, no. 1, pp. 53–59, 2008.

[28] R. P. Mcdonald and M.-H.-R. Ho, "Principles and practice in reporting structural equation analyses," *Psychol. Methods*, vol. 7, no. 1, pp. 64–82, Oct. 2005.

[29] J. Miles and M. Shevlin, "A time and a place for incremental fit indices," *Personality Individual Differences*, vol. 42, no. 5, pp. 869–874, May 2007.

[30] S.-O. Leung, "A comparison of psychometric properties and normality in 4-, 5-, 6-, and 11-point likert scales," *J. Social Service Res.*, vol. 37, no. 4, pp. 412–421, Jul. 2011.

[31] *Metro Satisfaction Survey*, Sina MicroBlog, Beijing, China, 2018.

[32] Y.-X. Zhou, L. Shi, Y.-L. Gao, Y. Liang, and H.-Z. Wang, "Factor analysis on validity of perceived control in Internet of health care with Cronbach's reliability test," *Cogn. Syst. Res.*, vol. 52, pp. 596–602, Dec. 2018.

[33] *Beijing Municipal Bureau of Statistics and NBS Survey Office in Beijing*, Beijing Municipal Bureau of Statistics, Beijing, China, 2018.

[34] H. Li, X. Li, X. Xu, J. Liu, and B. Ran, "Modeling departure time choice of metro passengers with a smart corrected mixed logit model—A case study in Beijing," *Transp. Policy*, vol. 69, pp. 106–121, Oct. 2018.

[35] X. Zheng, H. Li, and X. Chen, "Analysis of the service time of security check in subway station and its influence factors," *J. Beijing Jiaotong Univ.*, vol. 42, pp. 1–6, 2018.

[36] Y. Hamdouch, H. Ho, A. Sumalee, and G. Wang, "Schedule-based transit assignment model with vehicle capacity and seat availability," *Transp. Res. B, Methodol.*, vol. 45, no. 10, pp. 1805–1830, Dec. 2011.

[37] J. Zhao, F. Zhang, L. Tu, C. Xu, D. Shen, C. Tian, X.-Y. Li, and Z. Li, "Estimation of passenger route choice pattern using smart card data for complex metro systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 4, pp. 790–801, Apr. 2017.

[38] Y. Zhu, H. N. Koutsopoulos, and N. H. Wilson, "Inferring left behind passengers in congested metro systems from automated data," *Transp. Res. Procedia*, vol. 23, pp. 362–379, 2017.

[39] X. Xu, L. Xie, H. Li, and L. Qin, "Learning the route choice behavior of subway passengers from AFC data," *Expert Syst. Appl.*, vol. 95, pp. 324–332, Apr. 2018.

[40] "Beijing metro will be able to swipe QR codes across its entire line of stations for all smartphones," The Beijing News, 2018.

[41] "Guangzhou metro pushes APP QR code through the gate," Guangzhou Daily, 2018.

[42] Z. Zhang, L. Jia, and Y. Qin, "Optimal number and location planning of evacuation signage in public space," *Saf. Sci.*, vol. 91, pp. 132–147, Jan. 2017.

[43] J. Qi, L. Yang, Y. Gao, and Z. Di, "Service-oriented train timetabling problem with consideration of women-only passenger cars," *Comput. Ind. Eng.*, vol. 127, pp. 1164–1182, Jan. 2019.

[44] *Women-Only' Cars*, Tokyo Metro, Tokyo, Japan, 2019.

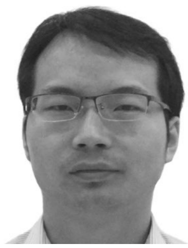
[45] *The Beijing Metro has Launched Trial Electronic One-Day Ticket*, Xinhua Net, Beijing, China, 2019.

[46] S. Raveau, Z. Guo, J. C. Muñoz, and N. H. Wilson, "A behavioural comparison of route choice on metro networks: Time, transfers, crowding, topology and socio-demographics," *Transp. Res. A, Policy Pract.*, vol. 66, pp. 185–195, Aug. 2014.

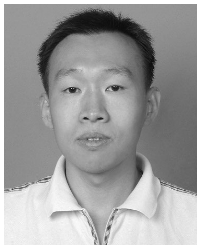
[47] Y. Zhu, H. N. Koutsopoulos, and N. H. Wilson, "A probabilistic passenger-to-train assignment model based on automated data," *Transp. Res. B, Methodol.*, vol. 104, pp. 522–542, Oct. 2017.

[48] J. Allen, L. Eboli, G. Mazzulla, and J. de D. Ortúzar, "Effect of critical incidents on public transport satisfaction and loyalty: An ordinal probit SEM-MIMIC approach," *Transportation*, pp. 1–37, 2018.

[49] J. Allen, L. Eboli, C. Forciniti, G. Mazzulla, and J. D. D. Ortúzar, "The role of critical incidents and involvement in transit satisfaction and loyalty," *Transp. Policy*, vol. 75, pp. 57–69, Mar. 2019.

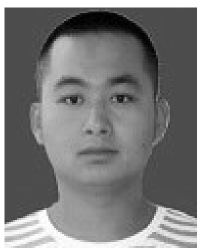


XINYUE XU received the Ph.D. degree from Beijing Jiaotong University, Beijing, China. He is currently an Associate Professor with the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing. His research interests include behavioral modeling, capacity analysis of subway stations, passenger flow control at crowding stations, and intelligent transportation systems.

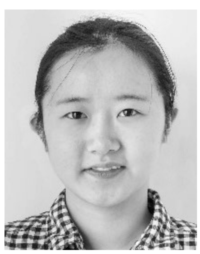


YI LU received the M.S. degree from Beijing Jiaotong University. He is currently a Senior Traffic and Transport Manager with the Subway Operation Technology Centre, Mass Transit Railway Operation Corporation Ltd., and with the Beijing Key Laboratory of Subway Operation Safety Technology. His research has focused on transport planning, traffic state estimation, and service quality evaluation.

YINGXIANG WANG received the B.S. degree from Beijing Jiaotong University, where he is currently a Graduate Student with the School of Traffic and Transportation. His research interests include passenger behaviors analysis and the evaluation of service quality.



JIANMIN LI received the B.S. degree from Inner Mongolia Agricultural University. He is currently pursuing the Ph.D. degree with the School of Traffic and Transportation, Beijing Jiaotong University. His research interests include capacity analysis and the evaluation of service quality.



HANYU ZHANG received the B.A. degree from Oberlin College, OH, USA, in 2018. She is currently a Graduate Student with the Georgia Institute of Technology, Atlanta, USA.

Her current research interests include travel behavior and passenger flow prediction.