Modeling and Analysis of Bus Satisfaction Based on improved structural equation model

Zhixin Wang, Yi Zhang*, Jinhui Li, Xiaomin Wang and Meng Zhang

Vehicle and Traffic Engineering school, Henan University of Science and Technology, Luoyang, Henan, 471003, China

*Corresponding author’s e-mail: hkdjtszy@haust. edu. cn

Abstract: In order to analyze the influencing factors of urban bus passenger satisfaction more scientifically, this paper constructs an improved structural equation model. The key point of this method is to extract principal components by principal component analysis before applying traditional structural equation model. Taking Luoyang as an example, summarize 21 indicators that affect passenger satisfaction. The SPSS software was used to analyze the principal component of the questionnaire data, and the passenger perceptible mass factor was reduced from 21 dimension to 4 dimension. On this basis, the structural equation model of potential factor correlation and passenger satisfaction evaluation was established for the extracted 4 principal components and 21 observable variables. Finally, the Amos software is used to verify it. Through the verification of the model, it is concluded that the line and the operational service factors have a positive impact on the degree of rapid response of the bus, the line factor has a positive impact on the operational service effect, the convenience factor has a positive impact on the operational service factor and the passenger satisfaction is undervaluation.

1. Introduction

The scientific evaluation of urban public transportation service quality is helpful to guide the planning and improvement of urban public transportation system. In addition, the quality of conventional bus service is an important index to evaluate the development level of urban public transport, and the quality of bus service is not determined by a single index, but by the comprehensive evaluation of multiple observable variables.

Liu Xizhen, Jia Youjun and Wang Song[1] used the scoring test and structural equation model to analyze the satisfaction of public transport service in Urumqi, and concluded that the direct positive impact on public bus satisfaction is on economic convenience and traffic safety, while the indirect positive impact is on operation order and vehicle condition. Zhang Bing, Zeng Minghua, Chen Qiyuan, etal[2] used the structural equation model to establish the relationship model between the service quality of passengers and the satisfaction and loyalty of passengers, and quantitatively analyzed the impact of various variables on passenger satisfaction and loyalty. The important factors of service quality are the transfer convenience of bus route, sanitary condition of bus and bus operation safety. Li Qiong, Ma Xinghui, Zhu Shunxing[3] put forward an evaluation model of bus passenger satisfaction based on structural equation model, and gave corresponding improvement measures for the influential factors.

At present, most bus satisfaction surveys classify observable factors in a subjective way, and then use structural equation model for correlation analysis, and rarely use principal component analysis for
scientific dimensionality reduction of observable variables. Therefore, it is more accurate to scientifically reduce the dimensionality of observable variables, and then analyze the relationship between latent variables through structural equation model and conduct passenger satisfaction analysis.

2. Systematic approach to satisfaction evaluation

2.1. Indicator setting and quantification method

2.1.1. Indicator setting basis. Compared with the bus operation systems of major cities such as Shanghai, Nanjing and Chengdu, the development of public transportation in small and medium-sized cities is in a backward stage, but the development status of buses in small and medium-sized cities is roughly the same. For the bus satisfaction evaluation of small and medium-sized cities, the setting of evaluation indicators is not completely consistent based on different cities, but it is mainly set up by the passengers' perception of speed, safety and convenience.

2.1.2. Indicator quantification standard. The Likert scale has the characteristics of simple structure and strong operability, and has higher reliability than other scales, so it is the most commonly used rating scale, suitable for questionnaires. The survey consists mainly of five parts of the statement, namely the five-level scale: 1 point is particularly dissatisfied, 2 points are dissatisfied, 3 points are general, 4 points are satisfactory, and 5 points are particularly satisfactory. Quantitative questionnaires can understand the respondents' satisfaction with the indicators, and play a vital role in the research and improvement of the problems.

2.2. PCA system analysis

PCA (Principal Component Analysis) is a method used to reduce the dimension of multivariate data. This method can compress and process data, which facilitates the analysis of data and improves the processing speed of the algorithm. It retains most of the information of the original variables, and each principal component is a linear combination of the original variables.

The principle of PCA is to project the original sample data into a new space. It can also be understood as transforming one set of coordinates into another set of new coordinates. But in the new coordinate system, it means that the original sample does not need so many variables, only need the original sample of one of the biggest linearly independent eigenvalue corresponds to the space coordinates can be.

2.3. Quantitative analysis model of bus satisfaction

Structural equation model (SEM) is a statistical method to analyze the relationship between variables based on covariance matrix. Structural equation models can deal with latent variables and their indices simultaneously.

The fitness of the structural equation model is an important indicator to test whether the research model is consistent with the original data. The overall fitting degree index is derived from the theoretical model matrix and the actual matrix. If the difference is less than the significance threshold value set by the analyst, it means the fitting degree between the hypothesis model and the actual data is good. According to international practice, five indicators in Table 1 are used to evaluate the research model.

| Fitting Index | CMIN/DF | RMR | RMSEA | AGFI | CFI |
|---------------|---------|-----|-------|------|-----|
| Reference     | >1 and  | <3  | <0.05 | <0.08| >0.9|>0.9 |

3. Case analysis

In order to verify the accuracy of the method, taking the most representative bus in Luoyang as an
example, this paper summarizes the evaluation indicators affecting the passenger satisfaction of the bus in Luoyang. After designing the questionnaire, the questionnaire is issued to the passengers who often take the bus. The main way of questionnaire survey is to distribute on-board bus and waiting area on platform. 836 questionnaires were collected, and 779 valid questionnaires were obtained after screening, and using the above methods to process and model the data.

3.1. Design of observable variables
Summarize the factors that bus travel may cause passengers' dissatisfaction, and draw 21 factors that conform to the traffic status of Luoyang and may affect the perceived quality of bus passengers in Luoyang. which are Q1 station spacing setting, Q2 departure frequency, Q3 transfer convenience, Q4 payment mode, Q5 fare setting, Q6 vehicle real-time operation awareness, Q7 route layout, Q8 line popularity, Q9 vehicle. Included passenger space, Q10 travel time, Q11 operation speed, Q12 average waiting time, Q13 punctuality rate, Q14 seat comfort, Q15 ride comfort, Q16 vehicle temperature and humidity, Q17 bus policy, Q18 driver service attitude, Q19 shape design, Q20 personal property safety performance, Q21 emergency rescue measures.

3.2. Principal component analysis of passenger perceived quality
The data obtained from the questionnaire were imported into SPSS in scv format for principal component analysis. As shown in table 2, its KMO sampling suitability quantity was 0.942, greater than 0.7, and its P value was 0, indicating that the factor analysis was valid.

|   | KMO sampling suitability quantity | Approximate chi square | Degree of freedom | Significance |
|---|----------------------------------|------------------------|-------------------|--------------|
|   | 0.942                            | 3066.055               | 210               | 0.000        |

After principal component analysis, 4 factors accounted for the main variation value, so 4 factors are selected as the best. According to the Table 3, factors with large absolute value of factor loading coefficient are classified into one category, which can be classified as follows and named: Q1, Q2, Q3, Q5, Q7 and Q8 are named as line factors; Q4 and Q6 are named as convenience factors; Q9, Q10, Q11, Q14, Q15, Q16, Q17, Q18, Q19 Q20, and Q21 are named as operational service factors; Q12 and Q13, are named as the swiftness factors.

|   |   |   |   |   |   | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|---|---|---|---|
| Q20 | 0.801 | 0.255 | 0.000 | 0.040 | Q17 | 0.700 | 0.370 | 0.141 | 0.111 |
| Q18 | 0.791 | 0.234 | -0.021 | -0.046 | Q10 | 0.658 | 0.446 | 0.247 | 0.035 |
| Q21 | 0.791 | 0.289 | -0.092 | -0.055 | Q11 | 0.652 | 0.469 | 0.196 | 0.114 |
| Q16 | 0.783 | 0.207 | 0.266 | -0.073 | Q9 | 0.632 | 0.508 | 0.074 | 0.046 |
| Q19 | 0.775 | 0.335 | 0.197 | 0.043 | Q2 | 0.258 | 0.820 | 0.162 | 0.111 |
| Q14 | 0.772 | 0.248 | 0.319 | -0.404 | Q1 | 0.276 | 0.788 | 0.138 | 0.067 |
| Q15 | 0.744 | 0.231 | 0.347 | 0.005 | Q3 | 0.297 | 0.785 | 0.210 | 0.124 |

3.3. Modeling and analysis

3.3.1. Model establishment. In order to find the direct and indirect relationships among the latent variables named in section 3.2, a model was established according to the classification results, as shown in Figure 1.

3.3.2. Analysis of potential influencing factors. AMOS software was used to establish the potential factor causality model, and 779 valid data investigated were analyzed by maximum likelihood method[9]. By analyzing the causality model of potential factors established in figure 1, the measurement index of table 4 model measurement indicators can be obtained. After the principal
component analysis, the detection indicators in the model are slightly deviated from the ideal value, which proves that AMOS is not fully recognized for the principal component extracted by SPSS, so the model needs to be revised. According to Modification Indices under AMOS program, the amendment suggestions are given in table 5.

Table 4. Model measurement indicators.

| Model    | RMR   | CFI   | AGFI  | CMIN/DF | RMSEA |
|----------|-------|-------|-------|---------|-------|
| Default model | 0.646 | 0.896 | 0.867 | 3.058   | 0.100 |

Table 5. Variables to be modified.

| M.I. | M.I. | M.I. |
|------|------|------|
| F3<--F4 | 6.766 | Q13<--F4 | 6.153 |
|       |      | Q11<--F3 | 4.255 |

According to the modified variables proposed in table 5, the model is modified to establish the potential factor relationship model as shown in figure 2:

Figure 1. Causality analysis model of potential factor.Figure 2. Modified potential factor relationship model.

The modified model was tested to obtain the non-standardized regression coefficients, the significance test in table 6, and the model measurement indicators corrected in table 7. The data in table 6 show that the C.R. values of the model are all greater than 1.96, and their P values are all less than 0.001, indicating that the model has a high level of significance. All the indexes in table 7 are within the range of ideal values, which proves that the establishment of the model is reasonable.

Table 6. Non-standardized regression coefficients and significance test.

| Estimate | S.E. | C.R. | P   | Estimate | S.E. | C.R. | P   |
|----------|------|------|-----|----------|------|------|-----|
| F2<---F1 | 1.257 | 0.124 | 10.123*** | Q18<---F4 | 0.954 | 0.084 | 11.395*** |
| F4<---F2 | 0.688 | 0.066 | 10.368*** | Q19<---F4 | 1.121 | 0.085 | 13.202*** |
| F3<---F1 | 1.164 | 0.097 | 12.048*** | Q20<---F4 | 1.012 | 0.085 | 11.954*** |
| F3<---F1 | 0.572 | 0.543 | 6.557*** | Q21<---F4 | 1.000 |       |     |
| Q6<---F2 | 1.000 |       |       | Q1<---F1 | 1.000 |       |     |
| Q4<---F2 | 1.020 | 0.095 | 10.725*** | Q8<---F1 | 1.079 | 0.091 | 11.848*** |
| Q9<---F4 | 0.966 | 0.082 | 11.823*** | Q3<---F1 | 1.178 | 0.079 | 12.174*** |
| Q10<---F4 | 1.196 | 0.094 | 12.692*** | Q5<---F1 | 0.614 | 0.084 | 7.344*** |
| Q13<---F4 | 1.137 | 0.091 | 12.529*** | Q2<---F1 | 1.114 | 0.093 | 11.942*** |
| Q14<---F4 | 1.190 | 0.092 | 12.945*** | Q7<---F1 | 1.118 | 0.090 | 12.362*** |
| Q15<---F4 | 1.227 | 0.098 | 12.553*** | Q11<----F3 | 0.942 | 0.056 | 16.878*** |
| Q16<---F4 | 1.083 | 0.089 | 12.526*** | Q12<----F3 | 1.000 |       |     |
| Q17<---F4 | 1.061 | 0.088 | 11.996*** |       |       |       |     |

Table 7. Modified model measurement indexes.

| Model    | RMR    | CFI    | AGFI   | CMIN/DF | RMSEA  |
|----------|--------|--------|--------|---------|--------|
| Default model | 0.423  | 0.906  | 0.934  | 2.875   | 0.063  |

It can be seen from the automatically corrected model of AMOS in figure 2 that: ① the line factor
has a significant direct and positive impact on the convenience factor. ② Line factors have a significant direct and positive effect on swiftness factors. ③ Convenience factors have a significant direct positive impact on operational service factors. ④ Operational service factors have a significant direct positive impact on the swiftness factor. According to the revised relationship model of potential factors in figure 2, the scoring mechanism of each potential factor can be obtained as follows:

\[ F_1 = \sum (F_1 \text{ observation index score} \times \text{ weight}) \]
\[ F_2 = \sum (\text{weight of } F_1 \times \text{measured value of } F_1 + F_2 \text{ observation index score} \times \text{weight}) \]
\[ F_3 = \sum (\text{weight of } F_4 \times \text{measured value of } F_4 + \text{weight of } F_1 \times \text{measured value of } F_1 + F_3 \text{ observation index score} \times \text{weight}) \]
\[ F_4 = \sum (\text{weight of } F_2 \times \text{measured value of } F_2 + F_4 \text{ observation index score} \times \text{weight}) \]

Taking the average of 779 survey data under each observation index as the score of each observation index, the final score of bus line factor is 2.95, convenience factor is 2.33, the swiftness factor is 2.27, and operational service factor is 3.05.

3.3.3. Comprehensive evaluation of passenger satisfaction. On the basis of the modified model in figure 2, the overall bus satisfaction in Luoyang is evaluated, and the comprehensive evaluation model shown in figure 3 is established. It can be seen that the correlation between the various potential factors were greater than 0.8, has a strong correlation between the various potential factors, among which the path coefficient of convenience factor (F2) is 0.97, and the path of swiftness factor (F3) is 0.77, which reveals that the evaluation of the bus passengers in Luoyang for the payment method of Luoyang bus, the real-time operation of the bus, the bus operation speed and the average waiting time of the bus are not very optimistic. In figure 3, “G” stands for bus satisfaction.

Figure 3. Comprehensive evaluation model of bus satisfaction.

The operation results of the bus satisfaction comprehensive evaluation model are shown in table 8. All the indexes meet the requirements of ideal values, which proves that the model has a good fitting degree.

| Model       | RMR  | CFI   | AGFI  | CMIN/DF | RMSEA |
|-------------|------|-------|-------|---------|-------|
| Default model | 0.454 | 0.910 | 0.928 | 2.698   | 0.056 |
According to the comprehensive evaluation model of bus satisfaction in figure 3, it can be concluded that the weight of bus route factor is 0.131, the weight of convenience factor is 0.385, the weight of swiftness factor is 0.306, and the weight of operation service factor is 0.178. The comprehensive evaluation score of bus satisfaction in Luoyang is 2.52, which shows that the evaluation of bus riders in Luoyang is general.

4. Conclusion
By combining principal component analysis with structural equation model, an improved structural equation model is created. The observable indexes of 21 dimensions are reduced to 4 dimensions, and the distribution relationship between observable variables and latent variables is more accurately located. The unreasonable evaluation caused by too rigid modeling caused by inaccurate subjectivity allocation is avoided. The model test results are basically consistent with the actual situation in Luoyang City, which proves the effectiveness of the method.

Acknowledgments
This research was financially supported by the National Social Science Foundation of China (Grant NO.16BJY070) and the doctoral research start-up fund of henan university of science and technology (Grant NO. 13480037).

References
[1] Liu Xizhen, Jia Youjun, Wang Song. (2016) Study on citizen satisfaction of city bus service based on SEM model——Take Urumqi Municipality for example. Journal of Harbin university of commerce(JCR-SSCI), 6:71-79.
[2] Zhang Bing, Zeng Minghua, Chen Qiuyan, etal. (2016) SEM-based research on urban bus service quality-satisfaction-loyalty. Mathematical Statistics and Management, 35(2): 198-205.
[3] Li Qiong, Ma Xinghui, Zhu Shunxing. (2016) Bus passenger satisfaction evaluation model based on SEM. logistics technology, 35(10): 78-81.
[4] Hu Linfang, Zheng Qiang, Song Xueli. (2018) Principal Component Analysis-based EWMA Control Chart for Monitoring in Short-term Trading of Single Stock. Journal of xinyang normal university (Natural Science Edition), 31(01): 11-16.
[5] Zhang Yuan, Jia Kebin. (2013) An urban expressway traffic status classification method based on video sequence. Traffic Information and Safety, 31(04): 14-20.
[6] Gabriella Conti, Sylvia Frühwirth-Schnatter, James J. Heckman, Rémi Piatek. (2014) Bayesian exploratory factor analysis. Journal of Econometrics , 183(1): 31-57.
[7] Chen Yu, Wang Ying, Nie Qiong, et al, (2017) The safety perception modeling and analysis of ride-hailing passengers based on SEM model. Traffic Information and Safety, 35(06): 85-92.
[8] Meng Xianghai, Tan Wei, Huo Xiaoyan. (2018) Data distribution characteristics of expressway traffic accidents based on statistics and hypothesis test. Journal of Transportation Engineering, 18(01): 139-149.
[9] Wen Han, Liang Yunsi. (2015) Structural equation models are commonly used to fit the essence of exponential tests. Psychological Science, 38(04): 987-994.