

Data based violated behavior analysis of taxi driver in metropolis in China

Yao, Jiao; Ni, Yiling; Zhao, Jing; Niu, Huiwei; Liu, Shanyong; Zheng, Yuhui; Wang, Jin

DOI

[10.32604/cmc.2019.06252](https://doi.org/10.32604/cmc.2019.06252)

Publication date

2019

Document Version

Final published version

Published in

Computers, Materials and Continua

Citation (APA)

Yao, J., Ni, Y., Zhao, J., Niu, H., Liu, S., Zheng, Y., & Wang, J. (2019). Data based violated behavior analysis of taxi driver in metropolis in China. *Computers, Materials and Continua*, 60(3), 1109-1122. <https://doi.org/10.32604/cmc.2019.06252>

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

Data Based Violated Behavior Analysis of Taxi Driver in Metropolis in China

Jiao Yao¹, Yiling Ni¹, Jing Zhao², Huiwei Niu¹, Shanyong Liu¹, Yuhui Zheng³ and Jin Wang^{4, 5, *}

Abstract: Violation probability of taxi drivers in metropolis is far more than that of normal drivers because they are labor-intensive, overconfident of self-driving skill, and always searching potential customers, sometimes even picking up or dropping off passengers randomly. In this paper, four types of violated behavior of taxi drivers in metropolis were first summarized, based on which corresponding scale table was initially designed with social statistical method. Furthermore, with certain samples, relative item analysis, exploratory factor analysis, validity analysis and reliability analysis were conducted to verify validity of the initial scale table, based on which some improvements were made, and we can see that the modified scale table in the paper has high fitness degree, good reliability and validity to detect violated behavior of taxi driver accurately. Finally, large area survey data of taxi driver questionnaire from Shanghai was collected with the modified scale table above, the analysis results showed that among four types of violated behavior of taxi drivers in metropolis, the probability over-speed is top to 89.57%, in which probabilities of behaviors of “driving over-speed at mid-night” and “accelerating to across the intersection during the yellow signal” are top to 64.2% and 58.2% respectively, which is meaningful for the improvement of taxi drivers’ behaviors specification and traffic safety regulation.

Keywords: Taxi driver, violated behavior, scale table design, exploratory factor analysis, validity analysis.

1 Introduction

Because of long tedious work and high labor intensity, taxi drivers easily feel tired physically and psychologically [Lu (2018)]; furthermore, emerging of car-hailing via Apps also aggravate their work pressure [Simon, Milind and Sumeet (2018)], all which result

¹ Business School, University of Shanghai for Science and Technology, Shanghai, 200093, China.

² Department of Transport & Planning, Delft University of Technology, Delft, 2628CN, Netherlands.

³ Nanjing University of Information Science and Technology, Nanjing, 210044, China.

⁴ School of Computer and Communication Engineering, Changsha University of Science and Technology, Changsha, 410114, China.

⁵ School of Information Science and Engineering, Fujian University of Technology, Fuzhou, 350118, China.

These authors contributed equally to this work.

*Corresponding Author: Jin Wang. Email: jinwang@csust.edu.cn.

to high probability of their violated behavior on road. In addition, overconfidence of self-driving skill, searching potential customers, and sometimes having to meet demand of passengers getting on or off randomly can also cause their violated behaviors.

2 Related works and motivations

According to relative survey [Zhang and Mao (2011)] there were 49.78% drivers who had record of violated behavior or accident experience. Violated behaviors of tax driver are usually caused by risk perception loss and ignorance, which is the potential safety hazard for urban traffic flow operation [Feng, Liu and Li (2012); Blankenship and Nesbit (2013); Wang, Chang and Sun (2015)]. Iversen's research [Iversen (2004)] indicated that risk attitude and behavior of taxi drivers were closely related to violated behaviors or traffic accidents. Loss of risk perception and safety attitude of taxi driver usually result to aggressive violated driving, angry violated driving, experienced violated driving and concentration missing violated driving [Li (2011)]. About aggressive driving behavior, inducing factors were founded by Luo with questionnaire test [Luo (2009)]; about angry driving behavior, survey result from Yasak indicated that it's closely related to geo-culture [Yasak and Esiyok (2009)], and pattern expression of angry driving behavior and its influence on traffic safety were studied by Lei [Lei (2011)].

Questionnaire survey was usually used to express violated behaviors, Sullman [Sullman (2013)] designed the scale of anger driving behavior to identify characteristic of general and dangerous driving behavior, Zheng et al. [Zheng, Jiang and Zhang (2014)] proposed the modified scale to explore the reason of Chinese risky driving behavior, and four factor structure model was concluded from the survey data. Li designed scale table about driving behavior, driving attitude, driving technology, and driving personality to analyze characteristics of violated behavior of Chinese drivers [Li (2012)].

In recent years, big data such as GPS data from taxi were also used to collect driving behavior information, which can be used to explore the potential violated behavior expression [Sun and McIntosh (2018)], and some intelligent algorithm were also used in its data mining analysis [Pan, Yi and Chen (2018); Zeng, Dai, Li et al. (2018); Zheng, Sun and Wang et al. (2018); Zhang, Meng and Han (2017); Zhang, Han and Li et al. (2016) ; Li, Zhang and Zhang et al. (2018)] . Huang et al. [Huang, Sun and Tang (2017)] analyzed driving styles with GPS data, and combined with road and environment variables to explore the determinant factors of over-speed, based on the time serial matching data between GPS in taxi and taximeter, Xu et al. studied the violated behavior judgment subjected to mileage, travel time and cost [Xu, Huang, Zhou et al. (2017)], Li explored the valuable information with GPS position data of taxi vehicles, to identify their abnormal driving behaviors, abnormal driving trajectory, and intentional detour [Li, (2015)]. And GPS data, combined with video data, was also collected to establish characteristic index system of status of drivers' perception, decision and manipulate and vehicle operation, all which was used to evaluate the safety of driving behaviors [Li (2010)].

From the literature review above, we can see that about the method with questionnaire survey and GPS data, some disadvantages are unavoidable, such as poor representativeness of respondents, limited sample size, and poor accuracy in expression of behavior, etc. With the scientific professional scale table design and verification, we

can find violated behavior of taxi driver simply and accurately, especially with assistant of classical statistical analysis software.

In this study, we first summarized violated behaviors of taxi driver via interview and questionnaire survey. Furthermore, initial scale table was designed, and relative item analysis, exploratory factor analysis, validity analysis and reliability analysis of scale were also conducted verify validity of the initial scale table, based on which some improvements were made. Finally, with the survey collection data from Yangpu District, Shanghai, China, some meaningful conclusions were reached with the modified scale table. The detailed research thought diagram was shown as Fig. 1.

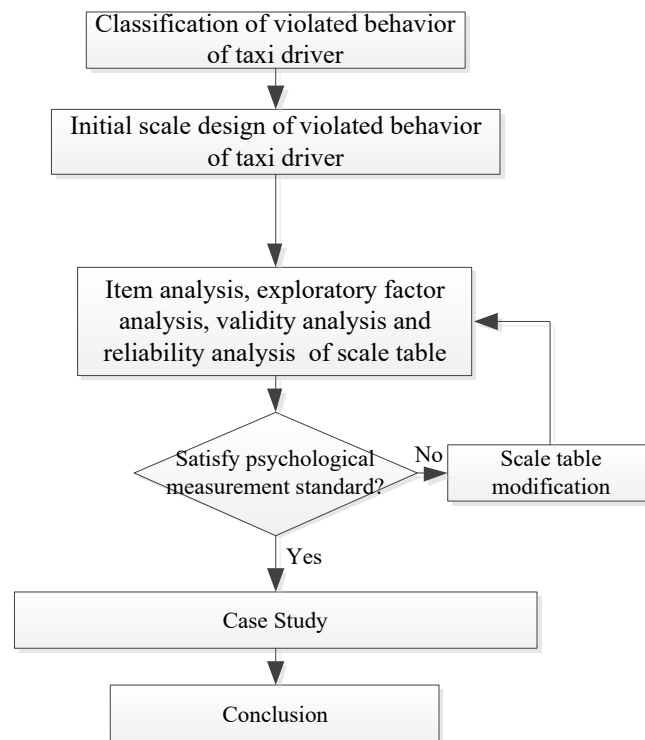


Figure 1: Research thought diagram

3 Scale table study of violated behavior of taxi driver

3.1 Initial scale table design

To understand characteristics of violated behavior of taxi driver, with relative reference scale [Yang (2007)], we designed the initial scale table with the guidance of psychological analysts. Furthermore, 10 taxi drivers were invited randomly to judge whether they can understand questions of the scale table.

After these steps, the initial scale was established which included 4 dimensions (4 classes of usual violated behaviors of taxi driver): ignorance of traffic sign and marking, dangerous car following over-speed driving, and illegal lane change and over-taking.

Every dimension also include several sub items, so there were 15 items in the initial scale table, and every item has 4 options: "Never", "Occasionally", "Frequently" and "Always", which represent 0, 1, 2, 3 points respectively. The initial scale table structure was shown as Tab. 1.

Table 1: Scale table structure of violated behavior of taxi driver

Dimension	Item Number	Item description
Ignorance of traffic sign and marking	1	Driving in wrong lane in absence of traffic police, because of unreasonable lane marking setting,
	2	Driving cross the solid line unintentionally
	3	Passenger waiting for service at unsuitable place, I have to stop illegally for boarding
	4	Red light violation at intersection without surveillance equipment in safety situation
Dangerous car-following	5	I'm an experienced driver, and I'm confident of my driving skills, so follow the car closely.
	6	I don't want other cars to cut in the platoon, so driving close to the car in front.
	7	Car in the front driving so slowly, but I cannot overtake it, so I have to follow it closely.
Over-speed driving	8	Passenger is in a hurry, I have to follow the car closely to overtake it at any time
	9	Time is money, for more orders, I am confident of my driving skills, so ignoring speed limit occasionally.
	10	There are little cars and pedestrians, so I can drive over-speed a little.
	11	The passenger is in a hurry, so driving speeding up to cross the intersection. when the signal is yellow
Illegal lane change and over-taking	12	Confident of my driving skills, I change the lane frequently to search the fast one.
	13	Somebody driving too slowly, I have to overtake it using lanes at the opposite direction.
	14	Too many cars on the road, I have to change lanes to avoid congestion.
	15	No car in the bus lane, so changing lanes to it to overtake cars in front.

3.2 Scale table analysis

3.2.1 Item analysis

Item analysis was first made to improve validity and reliability of scale, because good internal consistency between item and total scale means good discriminability of the item, and close correlation between item and scale means this item can identify samples in different levels.

If item score is high when total score high, or if is low when total score is low, we can see

that this item keep consistent with the total scale, which means discriminability of this item is very good.

Product-moment formula is selected for analysis as follow:

$$\gamma = \frac{\sum XY - (\sum X)(\sum Y)/n}{\sqrt{\sum X^2 - (\sum X)^2/n} \times \sqrt{\sum Y^2 - (\sum Y)^2/n}} \quad (1)$$

where γ is the product-moment correlation coefficient, X is the item score of sample, Y is the total score of sample, n is the total number of sample.

If the total scores of the scale table and items are relatively high, it shows that the internal consistency of items and the whole scale is better, the whole scale table is more reliable; on the other hand, high score of the item and scale table indicate that the item can distinguish items at different levels, in other words, the effectiveness of the tests is very good [Jin (2007)]. Therefore, the item identification analysis has important significance for the reliability and validity of the scale table.

In this scale table, total score of dimension is based on the total score of its related items, however, the purpose of the item analysis is to delete items that are inconsistent with total score of dimension, which means some items may score low when the total score of dimension is high, indicating that this item identification is very poor, and should be deleted. So the method of Principal Component Analysis (PCA) was used in exploratory factor analysis, in this case, there is no need to consider the correlation between total score and related items' score.

3.2.2 Exploratory factor analysis

The exploratory factor analysis allows the scale table to condense the original items into several main factors in the least loss of information, and makes the factor having named explanatory [Xue (2013)].

The main purposed of factor analysis is to reflect the majority information of original variables with less conflicting factors. The matrix form of mathematical model can be expressed as follow:

$$X = AF + \varepsilon \quad (2)$$

where F is the common factor, ε is the special factor, which means unexplained part of original variables, whose mean value is 0.

Factor analysis can be solved by the method of Principal Component Analysis, in which by means of coordinate transformation, original relative variables can be converted to another set of uncorrelated variables with linear combination.

Before exploratory factor analysis, Bartlett Test of Sphericity and Kaiser-Meyer-Olkin test have to be finished to check whether there is intensive correlation among original variables.

Original hypothesis of Sphericity in Bartlett Test is that the correlation coefficient matrix of the original variable is a unit matrix, and the test statistics can be calculated by the determinant of correlation coefficient matrix, and it approximately obey chi squared

distribution. If the observation value of this statistics is relatively large, and the probability of corresponding P value is less than the given significance level α , we will reject the original hypothesis, and believe that correlation coefficient matrix is unlikely to be the unit matrix, and the original variables are suitable for factor analysis; if we accept the original hypothesis, the original variables are not suitable for factor analysis.

The KMO test statistic is the index to compare simple correlation coefficient and partial correlation coefficient among variables, which can be expressed as follow:

$$KMO = \frac{\sum_i \sum_{j \neq i} r_{ij}^2}{\sum_i \sum_{j \neq i} r_{ij}^2 + \sum_i \sum_{j \neq i} p_{ij}^2} \quad (3)$$

where r_{ij} is the simple correlation coefficient among variables x_i and other variables x_j ; p_{ij} is the partial correlation coefficient among variable x_i and other variables x_j after controlling remaining variables. The value of KMO statistics is always between 0 and 1. When the quadratic sum of simple correlation coefficient among all the variables is far more than the quadratic sum of the partial correlation coefficient, KMO value is close to 1. Approaching 1 of KMO value means intensive correlation among variables r , and original variables are suitable for factor analysis. When KMO value approaching 0, correlation among variables becomes weaker, and original variables are not suitable for factor analysis.

3.2.3 Validity analysis

The validity is the degree that measurement results from scale table reflect the research content. If measurement results match with research contents, we can say it has high validity, or the validity is low. The validity can be divided into three types: content validity, criterion validity and construct validity.

3.2.4 Reliability analysis

Reliability is the estimation of consistency degree of measurement. Higher reliability coefficient means that the scale table test results are more consistent, stable and reliable, and it's usually use the internal consistency to show the test reliability level [Xue (2013)]. The internal consistency reliability is divided into half reliability and the homogeneity reliability.

Before solving the split half reliability, the scale is split to odd and even two parts according to the item number, and then to calculate the correlation coefficient with related formula with Pearson product moment. If variances of two parts are equal, we can adjust it with Spielman Brown formula as follow:

$$r = 2r' / (1 + r') \quad (4)$$

where r is the reliability value of the total test, r' is the correlation coefficient between the two half scores.

When homogeneity test of variances of two parts is not homogeneous, Fran Flanagan formula can be used as follow:

$$r = 2(1 - (S_1^2 + S_2^2) / S_X^2) \quad (5)$$

where S_1 and S_2 represent variation values of two parts in test score, S_X represents variation number of all participants in total score of the whole test.

In calculation of the homogeneity reliability, the Cronbach's α coefficient is usually used, and the calculation formula is as follows:

$$\alpha = \frac{k\bar{r}}{1 + (k-1)\bar{r}} \quad (6)$$

where k is the assessment number of items, \bar{r} is the mean value of correlation coefficient of k items.

4 Validity verification of scale table

4.1 Survey respondent information

According to the relevant study, ratio between sample sizes and the item of scale table should be greater than 5:1, and the total sample size is no less than 100. Therefore, 200 questionnaires was distributed in No. 26 Road East Guoshun and No. 609, Road Guohe, Yangpu District, Shanghai, China, where are rest sites of test taxi drivers, and finally 187 valid questionnaires were taken back, so the retrieve rate is 93.5%. In these drivers, numbers of male and female are 161 and 26 respectively. About ages, they are from 27 years old to 59 years old, and about driving experience, the minimal is 3 years, and the maximal is 26 years.

4.2 Statistical analysis

The sample data was analyzed in classical statistical software SPSS22.0 in four aspects: item analysis, exploratory factor analysis, validity analysis and reliability analysis.

4.2.1 Item analysis

When correlation is highly significant (P -value>0.5) and correlation coefficient is more than 0.4, it means that homogeneity between item and scale is high and the tested psychological characters are closer to the fact. The correlation analysis between the items and the total score is shown in Tab. 2, from which we can see that all correlations are highly significant (P -value<0.01) except that of Item 6 and Item 15 whose correlation coefficients are 0.331 and 0.273 respectively, therefore the other thirteen items should be retained. And it could also be summarized that items in the scale table of taxi driver behavior of traffic violation are highly effective.

Table 2: Related factors of item and total score

Item	Related factor	Item	Related factor	Item	Related factor
1	0.627**	6	0.331	11	0.601**
2	0.461**	7	0.695**	12	0.664**
3	0.505**	8	0.603**	13	0.435**
4	0.541**	9	0.485**	14	0.658**
5	0.544**	10	0.448**	15	0.273

** The correlation is significant at the level of 0.01 (two-tailed)

4.2.2 Exploratory factor analysis

After Principal Component Analysis (PCA) of the 13 items above by SPSS, we can find that *KMO* of sample data was 0.87 (more than 0.80), which means that correlation among variances is high and the original variances are suitable for Factor Analysis. The level of significance of Bartlett's Test of Sphericity is high ($p < 0.001$), which demonstrates that matrix of correlation coefficients is impossible to be a unit matrix and the variances are suitable for Factor Analysis. Through Varimax orthogonal rotation in PCA, four factors whose eigenvalues are more than 1 were chosen to analyze its characteristic roots and Serec Plot, and the results show that no items need deleting. The accumulated variance contribution rate is 64.437%, which is more than acceptable accumulated variance contribution rate 40% and applicable accumulated variance contribution rate 50% [Xue (2013)]. The four factors are named by ignorance of traffic sign and marking, dangerous car following, over-speed and lane changing illegally, and their variance rate are 18.289%, 15.491%, 15.338%, 15.319% respectively. Load capacity of all dimension and items are shown in Tab. 3. The value of load was measured by coefficients of every variance' factors, and if the value is over 0.4, the item would belong to this factor.

Table 3: Load capacity of item in 4 dimensions

Ignorance of traffic sign and marking		Dangerous car-following		Over-speed driving		Illegal lane change and over-taking	
Item	Load	Item	Load	Item	Load	Item	Load
1	0.790	5	0.797	10	0.777	12	0.864
2	0.754	8	0.737	9	0.707	13	0.690
83	0.732	7	0.715	11	0.706	14	0.602
114	0.616	-	-	-	-	-	-

4.2.3 Validity analysis

According to Section 3.2.3, taking sample data from survey, content validity and structure validity is finished in this section.

1) Structure validity analysis

Confirmatory Factor Analysis (CFA) was used to test structure validity, which was mainly applied to test if the relationship between one factor and its according tested items was consistent with theoretical structure relationship that the researcher designed. CFA

was utilized through Structural Equation Modeling. In this paper, confirmatory factor of scale about taxi drivers behavior of traffic violation was analyzed by the modeling analysis software LISREL8.70 [Byrne (2013)]. For evaluation of model's GFI (Goodness of Fit Index), the higher GFI, the better its availability. All indexes of GFI are as follow:

- (1) if χ^2 / df is less than 3, model's GFI is high;
- (2) The closer of GFI to 1, the better of the model's fitting degree. Normally, GFI is more than 0.9;
- (3) The closer of RMSEA (Root Mean Square Error of Approximation) to 0, the better of model's fitting degree. Normally, RMSEA is more than 0.1;
- (4) The other indexes like NFI (Normed Fit Index), CFI (Comparative Fit Index) and IFI (Incremental Fit Index) are the same as the index χ^2 / df , which means the closer to 1, the better of the model's fitting degree.

Analysis result show that all GFI index reached ideal standard as follow:

$$\chi^2 / f=2.13, GFI=0.91, NFI=0.95, CFI=0.97, IFI=0.97, RMSEA=0.054$$

So, we would say that structure validation of the scale table is good enough.

2) Content validity analysis

To detect whether the sample content is suitable, correlation test between every dimension and total score listed on the scale of taxi drivers 'behaviors of traffic violation was finished as Tab. 4, from which we can see that correlation coefficients in matrix are all more than 0.75 and their *p*-value is less than 0.01, which indicated that the content tested by the four dimensions is consistent with scale and the modified scale has good content validity.

Table 4: Matrix of 4 dimensions and total score

Divisor	Ignorance of traffic sign and marking	Dangerous car-following	Over-speed driving	Illegal lane change and over-taking	Total points
Ignorance of traffic sign and marking	1	—	—	—	—
Dangerous car-following	0.423**	1	—	—	—
Over-speed driving	0.453**	0.509**	1	—	—
Illegal lane change and over-taking	0.427**	0.443**	0.524**	1	—
Total points	0.776**	0.789**	0.862**	0.796**	1

** The correlation is significant at the level of 0.01 (two-tailed).

4.2.4 Reliability analysis

Reliability analysis was also finished by SPSS. The split-half reliability and Cronbach's α are shown in Tab. 5. We can see that the split-half reliability value is 0.827 and Cronbach's α is 0.836, all which are more than 0.8. Meanwhile, Cronbach's α can satisfy the requirements of reliability test even after all sub-scales and some items were deleted. Therefore, we can believe that the scale has high reliability.

Table 5: Coefficient of split-half reliability and internal consistency of 4 dimensions

Dimension	Split half reliability	Cronbach's α
Ignorance of traffic sign and marking	0.809	0.812
Dangerous car-following	0.821	0.833
Over-speed driving	0.761	0.788
Illegal lane change and over-taking	0.784	0.801

5 Case study and discussion

The modified and verified scale table was employed to conduct questionnaire analysis of taxi drivers' behavior in larger area in Shanghai. A total of 588 questionnaires were distributed and 556 were taken back, and 544 were validated, effective rate is 92.35%.

In this questionnaire survey, numbers of male and female are 498 and 46 respectively. About ages, they are from 25 years old to 58 years old, and about driving experience, the minimal is 3 years, and the maximal is 29 years.

With the statistics analysis of survey data above in SPSS, we got results shown in Tab. 6 from which we can see that among 4 types of taxi driver violation behavior, probability of over-speed occasionally or frequently is high to 89.27%, and its mean value, median and modal number are also very high.

Both item 10 of "night over-speed" and item 11 of "speeding up to cross intersection during yellow light period" have low skewness (less than 0) and higher peak value. Apart from that, "the frequent appearance" of these two violation behaviors accounted for 64.20% and 58.20% respectively, more than 50%, which means that over-speed is also one of the most serious traffic violation behavior of taxi driver, in which the night over-speed and speeding up to cross intersection during yellow light period are especially outstanding.

As for item 7 "following behind the car moving slowly" listed in "dangerous car following" and item 12 "changing lane over confidence in driving skills" listed in "illegal lane change and over-taking", "the frequent appearance" of these two violation behavior accounted for 29.70% and 26.40% respectively, which demonstrated that more than 1/4 taxi drivers have the above violation behaviors.

Additionally, "never appearance" of item 4 "red light violation at intersection without surveillance in safety situation" listed in "ignorance of traffic sign and marking" has high proportion of 93.40%, which means that most drivers would not run a red light.

Table 6: Survey data statistical analysis

	Ignorance of traffic sign and marking				Dangerous car-following		
Item No.	1	2	3	4	5	7	8
Mean value	0.55	0.94	0.70	0.09	0.63	1.11	0.76
Median	0	1	1	0	0	1	1
mode	0	1	0	0	0	1	1
Skewness	0.73	0.84	0.96	5.42	1.26	0.03	0.29
Kurtosis	0.52	2.11	0.78	28.31	1.26	-0.88	-0.62
Never (%)	50.8	21.3	44.3	93.4	53.8	25.3	35.4
Occasionally (%)	41.3	66.2	43.2	3.4	36.3	44	55.8
Often (%)	7.9	8.1	9.2	3.2	7.7	29.7	8.8
Always (%)	0	4.4	1.1	0	2.2	1	0
	Over-speed driving			Illegal lane change and over-taking			
Item No.	9	10	11	12	13	14	
Mean value	1.03	1.81	1.49	0.88	0.66	0.35	
Median	1	2	2	0	1	0	
mode	1	2	2	0	1	0	
Skewness	0.20	0.13	0.79	0.68	0.38	0.65	
Kurtosis	1.21	0.07	1.21	1.00	0.61	1.81	
Never (%)	19.3	5.2	6.8	50.8	40.3	62.1	
Occasionally (%)	48.9	29.8	33.2	16.2	50.3	37.9	
Often (%)	28.4	64.2	58.2	26.4	9.4	0	
Always (%)	3.4	0.8	1.8	6.6	0	0	

Notes: Percentages of frequency for each item in the scale about taxi driver' behavior of traffic violations are listed in the last four lines of the table.

6 Conclusion

In this paper, we first summarized 4 types of taxi driver violated behavior in metropolis, and designed the initial survey scale table, and furthermore, item analysis, exploratory factor analysis, validity analysis and reliability analysis were conducted to verify validity of the scale table, based on which some improvements were made.

Based on the work above, large area survey was finished with the modified scale table, from which we can get conclusions as follow:

1) Among four types of violated behavior of taxi driver in metropolis, the probability of over-speed is 89.57% which is the most serious problem, in which "night over-speed" and "speeding up to cross intersection during yellow light period" are the most serious phenomena.

2) Among violated behaviors of “illegal lane changes and over-taking”, “changing lane over-confidence in driving skills” accounts for 26.4%, more than 1/4, which also should be focused on.

3) Among violated behaviors of “ignorance of traffic sign and marking”, “running red light at intersection without monitor” has lower proportion with 6.6%, which indicates that most of taxi drivers are not run a red light.

It should be pointed out that case study in this paper is limited to taxi drivers in Shanghai. Validation and general applicability of the designed scale table need to be verified further in other cities. Additionally, the designed scale table in this paper is only used to test violated behaviors of taxi driver, about the interior impact factors and mechanism of violated behaviors of taxi driver, it may be an interesting research work in future.

Acknowledgment: This work was funded by MOE (Ministry of Education in China) Project of Humanities and Social Sciences (17YJCZH225). It was also funded by the humanistic and social science climbing program funding of University of Shanghai for Science and Technology (SK18PB03).

References

- Blankenship, K. L.; Nesbit, S. M.** (2013): Driving stimuli increases accessibility of aggression-related concepts in “angry” drivers. *Personality & Individual Differences*, vol. 55, no. 2, pp. 135-140.
- Feng, Z. X.; Liu, J.; Li, Y. Y.** (2012): Selected model and sensitivity analysis of aggressive driving behavior. *China Journal of Highway and Transport*, vol. 25, no. 2, pp. 106-112.
- Huang, Y. Z.; Sun, D.; Tang, J. Y.** (2017): Taxi driver speeding: who, when, where and how? A comparative study between Shanghai and New York. *Traffic Injury Prevention*, vol. 32, no. 10, pp. 574-576.
- Iversen, H.** (2004): Risk-taking attitudes and risky driving behavior. *Transportation Research Part F: Traffic Psychology & Behaviour*, vol. 7, no. 3, pp. 135-150.
- Jin, Y.** (2007): *Psychological Measurement*. East China Normal University Press.
- Lei, H.** (2011): *The Characteristics of Angry Driving Behaviors and Its Effects on Traffic Safety (Ph.D. Thesis)*. Wuhan University of Technology, China.
- Li, P. F.** (2010): *Research on Indices and Analysis of Driving Behavior (Ph.D. Thesis)*. Jilin University, China.
- Li, S. Z.** (2011): *Beijing Motor Vehicle Drivers' Risky Driving Behaviors Analysis (Ph.D. Thesis)*. Beijing Jiaotong University, China.
- Li, T.** (2012): *Research on Drivers' Aberrant Driving Behavior Styles and Their Discriminant Indices (Ph.D. Thesis)*. Tsinghua University, China.
- Li, X. Y.** (2015): *Research of Taxi Abnormal Behavior Detection based on Cloud Computing (Ph.D. Thesis)*. Wuhan University of Technology, China.

- Li, Z. T.; Zhang, J.; Zhang, K. H.; Li, Z. Y.** (2018): Visual tracking with weighted adaptive local sparse appearance model via spatio-temporal context learning. *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4478-4489.
- Lu, G. L.** (2006): *The Epidemiology of Traffic Safety among Taxi Drivers in Shanghai (Master Thesis)*. Fudan University, China.
- Luo, Y.** (2009): *Analysis and Research of Aggressive Driving Behavior at the Road Traffic (Ph.D. Thesis)*. Southwest Jiaotong University, China.
- Pan, Z. Q.; Yi, X. K.; Chen, L. M.** (2018): Motion and disparity vectors early determination for texture video in 3D-HEVC. *Multimedia Tools and Applications*, pp. 1-18.
- Simon, H.; Milind, K.; Sumeet, G.** (2016): Taxi apps, regulation, and the market for taxi journeys. *Transportation Research Part A: Policy and Practice*, vol. 88, pp. 15-25.
- Sullman, M. J.; Stephens, A. N.** (2013): A comparison of the driving anger scale and the propensity for angry driving scale. *Accident Analysis & Prevention*, vol. 58, no. 5, pp. 88-96.
- Sun, H. Y.; McIntosh, S.** (2018): Analyzing cross-domain transportation big data of new york city with semi-supervised and active learning. *Computers, Materials & Continua*, vol. 57, no. 1, pp. 1-9.
- Wang, J. L.; Chang, R. S.; Sun, L.** (2015): Reliability and validity of dangerous driving behavior scale. *Chinese Journal of Ergonomics*, vol. 21, no. 6, pp. 47-49.
- Xu, F. S.; Huang, Z. X.; Zhou, K.; Liu, Q. X.** (2017): Study on irregularity of taxi based on operational data. *Journal of Highway and Transportation Research and Development*, vol. 34, no. 5, pp. 144-149.
- Xue, W.** (2013): *SPSS Statistical Analysis Method and Its application*, Publishing House of Electronics Industry.
- Yang, L. L.** (2007): *Study on Driving Behavior and Safety of Urban Taxi (Master Thesis)*. Beijing Jiaotong University, China.
- Yasak, Y.; Esiyok, B.** (2009): Anger amongst Turkish drivers: driving anger scale and its adapted, long and short version. *Safety Science*, vol. 47, no. 1, pp. 138-144.
- Zeng, D. J.; Dai, Y.; Li, F.; Simon, S.; Wang, J.** (2018): Adversarial learning for distant supervised relation extraction. *Computers, Materials & Continua*, vol. 55, no. 1, pp. 121-136.
- Zhang, D. W.; Meng, D. Y.; Han, J. W.** (2017): Co-saliency detection via a self-paced multiple-instance learning framework. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 5, pp. 865-878.
- Zhang, D. W.; Han, J. W.; Li, C.; Wang, J. D.; Li, X. L.** (2016): Detection of co-salient objects by looking deep and wide. *International Journal of Computer Vision*, vol. 120, no. 2, pp. 215-232.
- Zhang, M. L.; Mao, H. Q.** (2011): A prevalence survey of traffic safety knowledge, attitude and behavior among taxi drivers in Xining. *Qinghai Medical Journal*, vol. 41, no. 7, pp. 2-6.

Zheng, D. P.; Jiang, Z. H.; Zhang, Q. (2014): A study on driver's risky driving behavior and its influencing factors. *Chinese Journal of Ergonomics*, vol. 20, no. 1, pp. 20-25.

Zheng, Y. H.; Sun, L.; Wang, S. F.; Zhang, J. W.; Ning, J. F. (2018): Spatially regularized structural support vector machine for robust visual tracking. *IEEE Transactions on Neural Networks and Learning Systems*.

<https://ieeexplore.ieee.org/document/8428432>.