Optimization of Evaluation Indicators for Driver’s Traffic Literacy: An Improved Principal Component Analysis Method

Zhuo Chen¹* and Kang Tian²*

Abstract
The traditional traffic concept seems to be unable to adapt to the traffic problems brought by cities’ rapid development. People must cultivate new modern traffic literacy to deal with traffic problems. Based on traffic literacy, this paper constructs a traffic literacy evaluation indicator system including 13 evaluation indicators such as traffic rules and mechanical knowledge by summarizing relevant literature. We propose an Improved Principal Component Analysis (I-PCA) method, introduce the concept of information contribution sensitivity, and optimize and empower the traffic literacy indicator system. The primary research is to construct a traffic literacy evaluation indicator system including 13 evaluation indicators such as traffic rules and mechanical knowledge. The top 10 indicators that satisfy the cumulative information contribution rate value greater than 90% are retained, and the three indicators with low contribution rate are excluded. The optimization method can retain the indicator with a relatively large information contribution rate so that the indicator’s weight can genuinely reflect the information content of the corresponding indicator. The optimization method can retain the indicator with a relatively large information contribution rate so that the indicator’s weight can genuinely reflect the information content of the corresponding indicator.

Keywords
traffic literacy, evaluation indicator, principal component analysis, optimization, empowerment

Introduction
With the rapid development of the economy and the continuous improvement of people’s living standards, the urban traffic demand has caused changes in the overall traffic layout: roads become wider, traffic networks are denser, and vehicle traffic surges. This has led to more and more severe road congestion, frequent traffic accidents, and numerous violations of road traffic laws and regulations, which have seriously affected the development of cities and the quality of life of residents (Aijaz, 2019). The traditional concept of transportation in the past seems to be unable to adapt to this fast-developing era, and large-scale transportation facilities alone cannot fundamentally alleviate urban problems. People must improve their literacy to solve traffic problems in the new era. As the main transportation participants, motor vehicle drivers should improve their transportation literacy and jointly cope with transportation problems (S. Liu et al., 2022; Wang & Tian, 2021). From the perspective of drivers (drivers of all types of vehicles, including but not limited to two-wheeler drivers or four-wheeler drivers, or heavy vehicle drivers), this paper establishes and optimizes the evaluation indicator system of traffic literacy to lay the foundation for subsequent research.

Scholars have paid less attention to the research on traffic literacy, and most of the literature aims to explore the factors that affect traffic accidents (e.g., Lu et al., 2019; Nævestad et al., 2015; Yan et al., 2020), traffic safety (Han et al., 2010), and Transportation Health Literacy (Sargent-Cox et al., 2011). Some scholars believe that one of the factors to improve urban traffic safety is the development of urban safety infrastructure, such as the design of roundabouts (Macioszek, 2020), the design of traffic calming devices (Solowczuk, 2021), and the intelligent city network (Keyvan-Ekbatani et al., 2012). The practice has proved that the critical factor for traffic safety, ensuring health, and avoiding traffic accidents still lies in the ability of traffic participants to reflect and process real-time traffic conditions (Mahmud et al., 2019). Therefore, to

¹North China University of Water Resources and Electric Power, Zhengzhou, China
²Henan University of Economics and Law, Zhengzhou, China
*These authors contributed equally to this work.

Corresponding Author:
Kang Tian, School of Business Administration, Henan University of Economics and Law, No.180, Jinsishu Road, Zhengzhou, 450046, Henan, China.
Email: tian6039@126.com
fundamentally solve the traffic problems caused by human factors, it is necessary to improve the literacy of drivers themselves, which we call traffic literacy.

Referring to the definition of traffic literacy in existing research (S. Liu et al., 2022), the formation of traffic literacy starts from the individual’s mastery of traffic knowledge and then turns traffic knowledge into a skill and uses it to guide actions, cultivate, and form a good traffic safety awareness. The level of traffic literacy is the embodiment of the driver’s quality, and under the combined influence of acquiring Knowledge, mastering skills, safety awareness, and driving safety behavior, it gradually reduces the probability of traffic safety accidents. From the perspective of drivers, this study reviews the research related to traffic literacy, such as traffic safety evaluation indicators (Guo et al., 2011; Y. Zhang et al., 2015), green traffic evaluation indicators (G. X. Liu, 2019; Long & Gao, 2017), systematically summarizes various factors that affect drivers’ traffic literacy and preliminarily proposes an evaluation indicator system for driver’s traffic literacy.

It should be noted that the rationality of the driver’s traffic literacy evaluation index system is crucial to the evaluation results. Therefore, we need to optimize the indicator. Indicator selection is a standard method for indicator optimization. The existing indicator selection methods mainly focus on the relative importance of indicators and information overlap. The selection methods of relative importance mainly include the coefficient of variation (Meng & Chi, 2018), optimal variance method (Ahmad et al., 2017; Rezaei et al., 2018), and Bayesian longitudinal model (Fouskakis et al., 2020), etc. However, Relative importance can only indicate that a specific indicator is more important to the evaluation results, and it is impossible to judge whether there is information overlap between the indicators (Chen, 2021). The selection methods of overlapping information indicators mainly include correlation analysis (Kazemi, 2020; Ma et al., 2017), cluster analysis methods (Chen & Chi, 2015), and Support Vector Regression (Xiong et al., 2022). The effectiveness of these methods has been verified in numerous studies, but they cannot achieve the empowerment of indicators. This study aims to eliminate information redundancy indicators and introduce information contribution rates based on the I-PCA method to optimize and empower driver traffic literacy evaluation indicators.

The rest of this paper is structured as follows: Section 2 is the theoretical part of constructing the traffic literacy evaluation indicator system, and the indicator optimization method is introduced in Section 3. In Section 4, the data sources and indicator selection results are analyzed. Sections 5 and 6 are discussion and conclusion, respectively.

Theoretical Foundations

Concepts of Driver Traffic Literacy

Drawing on previous research experience on scientific literacy and environmental literacy, we introduce the concept of literacy into the research on driver traffic safety and regard current traffic problems as “human behavior.” The driver’s traffic literacy is a kind of professional literacy, and it is also a category of ability. It is the traffic behavior or tendency that traffic participants gradually develop through their understanding, attitude, and control ability to drive activities during long-term driving activities. It reflects the comprehensive quality of the driver’s traffic safety awareness in dealing with complex road traffic conditions and the implementation of traffic safety behaviors when people interact with the external traffic environment.

There are usually two identities in the natural traffic environment: traffic participants and traffic managers (Z. Li et al., 2011). However, we all need to master basic Knowledge such as mechanical common sense, traffic rules, traffic laws, and regulations, no matter what kind of identity. As a driver, he/she must have skills such as vehicle driving and emergency measures. A qualified driver’s Knowledge, experience, and method skills are insufficient. When the driver drives on the road, he has already entered the traffic environment and is faced with interaction with other vehicles, passing pedestrians, and passengers. At this time, it is necessary to have the ability to correctly judge whether to drive illegally, whether to comply with traffic orders, and whether to cooperate with traffic management. Drivers also need to have the social and moral concepts of “safety first” and “being kind to others” and cultivate the awareness of “driving carefully” and “respecting others,” which will prompt drivers to actively obey traffic rules and be courteous to other vehicles or pedestrians.

Since there are few studies on transportation literacy, this study mainly explores the evaluation dimension based on previous scholars’ experiences, conclusions, and research results. It mainly draws on more mature evaluation dimensions such as scientific and environmental literacy to find the evaluation dimension that conforms to the driver’s traffic literacy. American scholar Miller first proposed a “three-dimensional model” to define scientific quality, including understanding important scientific terms and concepts, scientific methods and research processes, and cognition and understanding of the social impact of science and technology (Miller, 1983). The International Student Assessment Project implemented by the OECD defines scientific literacy as a core competency and constructs an evaluation system for scientific literacy from four aspects: ability, Knowledge, attitude, and context (Bybee et al., 2009). After decades of development and precipitation, the viewpoints on the structure of scientific literacy have not yet formed a unified model but generally include two significant parts: scientific Knowledge and scientific ability.

Hungerford and Peyton (1976) first proposed the environmental literacy assessment framework, including ecological Knowledge, understanding of problems, concepts, values, attitudes, attribution judgments, environmental sensitivity, and action strategies (Hungerford & Peyton, 1976). With the
deepening of research, experts such as Hungerford H., Wilke R., and McKeown-Ice R. formed the Environmental Literacy Assessment Consortium. Based on the connotation, evaluation, and research conclusions of environmental literacy, they proposed four dimensions of the environmental literacy assessment framework (Simmons & Koenig, 1995; Wilke, 1995): (1) cognition (knowledge and skills); (2) emotion; (3) other factors that determine responsible environmental behavior; and (4) individual or collective participation in responsible environmental behavior (Tian, 2022). Another is the environmental literacy assessment framework proposed by Stapp and other scholars in the Tbilisi Declaration issued by UNESCO (1978): (1) knowledge; (2) emotion; (3) skills; and (4) behavior (Stapp, 1978).

Based on the evaluation system of scientific literacy and environmental literacy, this study defines the concept of traffic literacy as the sum of the Knowledge and experience that drivers have in traffic reserves, the mastery of reasonable methods and skills, the concept of people-oriented consciousness, and the implementation of self-control behaviors. By learning traffic knowledge, drivers can master traffic skills proficiently, establish good traffic awareness, and use it to guide their actions.

**Construction of Driver Traffic Literacy Indicator System**

According to the concept of traffic literacy defined above, we construct an indicator evaluation system from four dimensions: knowledge, awareness, skills, and behavior. Based on reviewing the existing research, the next-level indicators of each dimension are further explored. Finally, 13 evaluation indicators such as Traffic rules, Mechanical common sense, and Knowledge of laws and regulations are formed; see the third column of Table 1. The corresponding indicator observation points and references are shown in the fourth and fifth columns of Table 1.

**Method**

To realize the validity and scientificity of the measurement indicator, we propose an I-PCA method, which introduces the concept of information contribution rate to optimize the evaluation indicator. The sum of the difference between the retained principal component and an indicator and the cumulative contribution ratio of the corresponding principal component variance represents the indicator’s information contribution rate of the original indicator system. The indicator’s Information contribution rate reflects the indicator’s information contribution to the overall evaluation of traffic literacy and can be used as an essential basis for indicator optimization. This study draws on the method of Wang and Tian (2021), and the specific optimization steps are as follows.

1. Preliminarily process the data obtained based on the evaluation indicators, eliminate the measurement questionnaires that do not meet the measurement standards, and at the same time perform \( Z \) standardization processing on the preliminarily screened data:

   \[
   x_{ij} = \frac{y_{ij} - E(Y_i)}{s(Y_i)}
   \]

   where \( x_{ij} \) is the normalized value of the \( i \) th indicator for the \( J \) th sample, \( y_{ij} \) represents the value before normalization, and \( E(Y_i) \) and \( s(Y_i) \) represent the expectation and variance of the \( i \) th indicator’s original data, respectively; \( i = 1, 2, \ldots, m \), \( m \) indicates the number of indicators; \( j = 1, 2, \ldots, n \), \( n \) indicates the total sample size. The purpose of standardization is to prevent the data from affecting the analysis results during the PCA, which is not equivalent to the unified dimension of the indicators.

2. Solve the correlation coefficient matrix \( R \) based on the standardized data:

   \[
   R = \left[ r_{ij} \right]_{m \times m} = X^T X
   \]

   where \( r_{ij} \) is the correlation coefficient between the evaluation indicators, \( X \) is the standardized data matrix, and \( X^T \) is the transpose of \( X \).

3. Bring the matrix \( R \) into the equation:

   \[
   \begin{bmatrix} R - \lambda_i E_n \end{bmatrix} = 0
   \]

   where \( E_n \) represent an identity matrix of order \( n \), yielding the eigenvalues \( \lambda_i \). Bring the resulting eigenvalue \( \lambda_i \) into the equation:

   \[
   \omega_i = \frac{\lambda_i}{\sum_{j=1}^{n} \lambda_j}
   \]

   Further calculate the variance contribution rate \( Z_i \) (represent the \( i \) th principal component, \( i = 1, 2, \ldots, k \), \( k \) represent the number of retained principal components) of the principal component \( \omega_i \). \( \omega_i \) represent the proportion of the \( i \) th principal component \( Z_i \) explaining the original indicators set’s variation. The larger the value of \( \omega_i \), the more original information content of the indicator set carried by the principal component \( Z_i \). According to the variance contribution rate of the principal components, calculate the cumulative variance contribution rate \( U_k \), and the calculation equation is:

   \[
   U_k = \sum_{i=1}^{k} \omega_i
   \]
Table 1. Traffic Literacy Evaluation Indicators, Key Points, and References.

| Evaluation goals | Evaluation dimension | Evaluation indicators | Key points | References |
|------------------|----------------------|------------------------|------------|------------|
| Traffic literacy | Traffic knowledge    | Traffic rule $X_i$     | Road traffic markings; Traffic lights; Traffic signs; Knowledge of traffic safety instructions, etc. | Arthur and Doverspike (2001), Hardini and Indryss (2018), and Mirzeai et al. (2014) |
|                  | Mechanical common sense $X_i$ | The vehicle’s basic structure; Master the Knowledge of the correct operation of means of transportation; Understand each component’s performance | L. Li et al. (2018) |
|                  | Knowledge of laws and regulations $X_i$ | Traffic safety laws; Administrative regulations, and departmental rules and regulations | Machin and Sankey (2008), Zelikov et al. (2019), and Touahmia (2018) |
| Traffic awareness | Safety awareness $X_i$ | Alertness to the external driving environment; Defensive driving awareness | Ovedo-Trespalacios et al. (2019) and Vogelpohl et al. (2019) |
|                  | Values $X_i$ | Correct concept of road traffic; Attitude to traffic laws and regulations; Traffic risk perception ability | Martinussen et al. (2013) and Yao et al. (2019) |
|                  | Kind to others $X_i$ | Safety and equality; Personal ethics | Safarpour et al. (2020), Di Ciommo and Shihtan (2017), and Ori (2020) |
| Traffic skills   | Situation foresight $X_i$ | Predict possible situations while driving; Make scientific predictions for complex driving environments | Kaçan et al. (2019) and Mannering et al. (2020) |
|                  | Vehicle control $X_i$ | Master driving skills; Proficient in handling vehicles | McKenna and Horswill (2006) and Uzuncuoğlu et al. (2020) |
|                  | Precautionary measures $X_i$ | Perception of driving environment and risk; Ability to predict possible accidents; Prediction of possible driving behaviors | Lajunen and Summala (1995) and Lobanova and Evtiukov (2020) |
|                  | Emergency measures $X_i$ | Emergency Response Ability to Accidents; Emergency Handling Methods | Lobanova and Evtiukov (2020) and Kroll et al. (2020) |
| Traffic behavior | Driving control behavior $X_i$ | Driving control operating sensitivity; Driving control activity intensity; Driving control state stability | Seppelt and Lee, (2019), L. Zhang et al. (2019), and Treffner et al. (2002) |
|                  | Self-management behavior $X_i$ | Self-management of driver psychology; Self-reported driving behavior; Self-controlled driving behavior | Chai et al. (2020) |
|                  | Safe civilized behavior $X_i$ | Safe and civilized driving; Calmly face complex traffic conditions; Avoid illegal driving behavior | Deffenbacher (2016), Zepf et al. (2019), and Shukri et al. (2022) |

Source: Wang and Tian (2021).

The research shows that several principal components with a cumulative variance contribution rate of more than 60% can carry most of the information of the original indicator set (Wang, 2010). This study defined this threshold at 65%, but it can be adjusted according to the actual calculation situation to meet the research needs.

4. The difference function $\frac{Z_i^* - Z_i}{X_i^* - X_i}$ represents the influence degree of the principal component $Z_i$ by the change of the indicator $X_i$. The difference reflects a change between discrete quantities, indicating that based on the constant size of other indicators, the change in the size of the $j$th indicator $X_j$ leads to the change in the information of the $i$th principal component $Z_i$. $Z_i^*$ and $X_i^*$ are the changed values. By the equation:

$$Z_i = p_{i1}X_1 + p_{i2}X_2 + \cdots + p_{ij}X_j + \cdots + p_{im}X_m$$

the difference value can be calculated. The eigenvectors of the orthogonally normalized correlation coefficient matrix are $p_j^T = (p_{j1}, p_{j2}, \cdots, p_{jm})$. Next, calculate the degree $\alpha_{ij}$ of the $i$th principal component affected by the change of the $j$th indicator:

$$\alpha_{ij} = \left| \frac{Z_i^* - Z_i}{X_i^* - X_i} \right| = \left| p_{ij} \right|$$  \hspace{1cm} (7)

Where $\left| \frac{Z_i^* - Z_i}{X_i^* - X_i} \right|$ only reflects the degree to which the principal component $Z_i$ is affected by the change of the indicator $X_j$ and does not reflect the influence of the indicator on the information of the original indicator set. The result after the difference operation is approximately equal to the value after the eigenvectors are orthogonally normalized, and the absolute value is taken.

5. $\frac{Z_i^* - Z_i}{X_i^* - X_i}$ indicates that the information of all original indicators is only reflected by the $i$th
principal component and is affected by the \(i^{th}\) indicator’s change, which is recorded as information contribution degree \(\beta_i\):

\[
\beta_i = \omega_i \left| \frac{Z_i - Z_i^*}{X_i - X_i^*} \right| = \omega_i \left| p_{ii} \right| \quad (8)
\]

6. Denote the information contribution degree of the indicator \(X_i\) change to the original indicator set information by \(\sum_{j=1}^{k} \omega_j \left| \frac{Z_j - Z_j^*}{X_j - X_j^*} \right| \), which is called the information contribution rate of \(X_i\), denoted by \(\beta_i\):

\[
\beta_i = \sum_{j=1}^{k} \omega_j \left| \frac{Z_j - Z_j^*}{X_j - X_j^*} \right| = \sum_{j=1}^{k} \omega_j \left| p_{ij} \right| \quad (9)
\]

7. Arrange the obtained \(\beta_i\) values in descending order. The ratio of the sum of the \(\beta_i\) values of the top \(l\) indicators to the sum of the \(\beta_i\) values of all the indicators indicates that the selected indicators carry the information amount of the original indicator set, which is called the cumulative information contribution rate, denoted by \(\gamma_m\):

\[
\gamma_m = \frac{\sum_{j=1}^{l} \beta_j}{\sum_{j=1}^{m} \beta_j} \quad (10)
\]

PCA theory believes that when the cumulative information contribution rate is between 70% and 90%, the first \(n\) principal components can carry most of the information of the original indicators and should be retained (Jolliffe & Cadima, 2016). To ensure that the retained index can carry the maximum amount of information of the original indicator set, we retain the first \(L\) indicators that satisfy the value of \(\gamma_m\) greater than 90%.

8. The weighting of indicators is calculated by the ratio of the cumulative information contribution rate of one indicator to the cumulative information contribution rate of all indicators, denoted by \(\chi_i\):

\[
\chi_i = \frac{\beta_i}{\sum_{j=1}^{m} \beta_j} \quad (11)
\]

The indicator weights calculated in the process of indicator screening in this study were not excluded. After the indicator screening is completed, the problem corresponding to the excluded indicator should be deleted, and the remaining indicator should be re-weighted, which is also calculated using equation (11).

Data Sources and Results

Data Sources and Analysis

We designed a questionnaire based on the evaluation indicator system of driver traffic literacy constructed in Table 1. Respondents’ responses to the questionnaire were categorized in the form of a five-level Likert scale. The response levels are “Completely understanding, more understanding, somewhat understanding, Not quite understand, completely do not understand,” and the corresponding values are 5, 4, 3, 2, and 1. The questionnaire is divided into demographic characteristics and a traffic literacy survey scale. Individuals’ essential characteristics include 7 questions, including gender, age, education, occupation, place of residence, average monthly income, and years of driving experience; the traffic literacy survey scale includes 17 questions. Four indicators covering a wide range of information, such as Traffic rules, Safety awareness, Vehicle control, and Self-management behavior, are set with two questions, and the other indicators are set with one question.

The survey data all came from offline surveys. Drivers of different ages were randomly selected from different administrative regions in Zhengzhou to fill in the questionnaires, and a total of 550 questionnaires were distributed. Invalid questionnaires such as incomplete questionnaires and more than 80% of the questions answered the same option were excluded, and the preliminary screening of the questionnaires was completed. Finally, 393 valid questionnaires were obtained, and the effective sample rate was 71.33%. Studies have demonstrated that when the sample size is approximately 300, the consistency of sample loadings relative to population loadings greater than .9 can be obtained, which is an accepted value in the behavioral sciences for establishing an equivalence between two sets of loadings (Lorenzo-Seva & Ten Berge, 2006; Saccenti & Timmerman, 2016). In addition, in the process of PCA, the sample size should be more than 10 times the number of variables. If the study wants to get ideal results, it is better to be more than 25 times the number of variables (W. Zhang & Dong, 2013). We can conclude that the sample size obtained in this study meets the basic needs of the analysis.

When conducting PCA, it is necessary to ensure that the reliability and validity of the questionnaire meet the basic requirements of the analysis. SPSS 24.0 was used to test the reliability and validity of the questionnaire data. The primary reference indicators were the internal consistency coefficient (Cronbach’s \(\alpha\)), Kaiser-Meyer-Olkin (KMO) test statistic, and Bartlett’s sphericity test. The results of data analysis showed that Cronbach’s \(\alpha\) coefficient was .821, and the Cronbach’s alpha coefficient based on standardized items was .815, indicating that the overall reliability of the questionnaire was good. The value of the KMO test statistic is .854, and the Bartlett sphericity test has a significant \(p = .000\),
and the cumulative variance contribution rate of the common factor is 63.311%, only one-factor loading under each item is more significant than .4, and the other factor loading values are less than .4. It shows that the validity of the questionnaire is high.

### Empirical Analysis Results

Driver’s traffic literacy evaluation indicators are screened and optimized based on the I-PCA indicator optimization method. To obtain a more credible evaluation system, the indicators with relatively low cumulative contribution rates are deleted.

1. The four indicators of Traffic rules, Safety awareness, Vehicle control, and Self-management behavior correspond to two problems. Therefore, it is necessary to perform weighted average processing on the corresponding two sets of data to convert them into one set of data. According to equations (1) and (2), solve the matrix $R$ of the $Z$ normalized data, see Table 2.

2. Bring the correlation matrix coefficient result into equation (3) to get the eigenvalue $\lambda_i$, and arrange the eigenvalues in descending order, see the second column of Table 3. Bring the eigenvalues into equation (4), calculate the $Z_i$ of the principal component $\omega_i$, see the third column of Table 3. Calculate the $U_k$ according to equation (5), see the last column of Table 3. From the results, if the cumulative variance contribution rate of the first three principal components is greater than 65%, the top three principal components with the most significant $\omega_i$ are retained.

3. The eigenvector $p_i$ of the matrix $R$ is obtained from the solution of the equation: $(R - \lambda_i E) x = 0$. Substituting the initial matrix and eigenvalues into this equation computes the eigenvectors preserving the principal components. Then, according to equations (6) and (7), the degree $\alpha_{ij}$ is calculated, that is, the value of $[P_{ij}]$, as shown in columns 3 to 5 of Table 4. According to equations (8) and (9), the information contribution rate $\beta_j$ of the $j$ th indicator is calculated, and the calculation results are listed in the last column of Table 4.

4. Arrange the obtained $\beta_j$ values in descending order and place them in the third column of Table 4. Calculate the cumulative information contribution rate $\gamma_m$ according to equation (10); see the fourth column of Table 5. Therefore, the first nine indicators are retained, and the last three indicators are removed. According to equation (11), weights are assigned to each indicator; see the last column of Table 5.

### Discussion

This study draws on scientific, environmental, and other research foundations related to “literacy,” combined with the existing research related to traffic literacy. The traffic literacy indicator system we constructed from the four dimensions of Knowledge, skills, awareness, and behavior can reflect the overall quality of drivers, from basic Knowledge to behavior implementation in the driving process. The drivers are the most direct participant in traffic, and they not only need to consider the safety of passengers and the safety of other people and economic property around

### Table 2. Coefficient Matrix for Normalized Data.

| Indicator | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $\ldots$ | $X_{13}$ |
|-----------|-------|-------|-------|-------|-------|----------|---------|
| $X_1$     | 1.0000 | 0.457 | 0.612 | 0.526 | 0.320 | $\ldots$ | 0.222   |
| $X_2$     | 0.457  | 1.000 | 0.589 | 0.527 | 0.275 | $\ldots$ | 0.127   |
| $X_3$     | 0.612  | 0.589 | 1.000 | 0.628 | 0.387 | $\ldots$ | 0.237   |
| $X_4$     | 0.526  | 0.447 | 0.527 | 1.000 | 0.628 | $\ldots$ | 0.234   |
| $X_5$     | 0.320  | 0.275 | 0.387 | 0.628 | 1.000 | $\ldots$ | 0.218   |
| $\ldots$ | $\ldots$ | $\ldots$ | $\ldots$ | $\ldots$ | $\ldots$ | $\ldots$ | $\ldots$ |
| $X_{13}$  | 0.222  | 0.127 | 0.237 | 0.234 | 0.218 | $\ldots$ | 1.000   |

Source: Wang and Tian (2021).

### Table 3. Eigenvalues and Contribution Rates of Principal Components.

| Number/item | $\lambda_i$ | $\omega_i$ (%) | $U_k$ (%) |
|-------------|-------------|---------------|-----------|
| 1           | 6.697       | 51.518        | 51.518    |
| 2           | 1.173       | 9.025         | 60.543    |
| 3           | 1.074       | 7.494         | 68.037    |
| 4           | 0.839       | 6.455         | 74.492    |
| 5           | 0.608       | 4.680         | 79.172    |
| 6           | 0.531       | 4.084         | 83.256    |
| 7           | 0.469       | 3.605         | 86.861    |
| 8           | 0.395       | 3.035         | 89.896    |
| 9           | 0.345       | 2.655         | 92.551    |
| 10          | 0.329       | 2.528         | 95.078    |
| 11          | 0.264       | 2.033         | 97.112    |
| 12          | 0.243       | 1.871         | 98.983    |
| 13          | 0.132       | 1.017         | 100.000   |

Source: Wang and Tian (2021).
the driving environment (S. Liu et al., 2022). Improving the driver’s traffic literacy is conducive to improving the driver’s control over the entire driving environment, effectively controlling his driving behavior, taking emergency avoidance behaviors, etc. (Vaa, 2007; Zaidan et al., 2022). This will directly or indirectly reduce traffic accidents and reduce the maintenance cost of traffic facilities so that the functions of traffic facilities will play a more significant role. Drivers with higher traffic literacy will have higher safety awareness (S. Liu et al., 2022) and can effectively improve the culture of traffic safety behavior. Under cultural constraints, guidance and incentives, traffic order can be better improved, and the probability of traffic accidents is reduced (Sujon & Dai, 2021; Wishart et al., 2019).

In the era of informatization and intelligence, the dimensions of the indicator system for evaluation of different complex systems are becoming more and more complex, and information overlap between indicators will inevitably occur. Some scholars use dimensionality reduction methods to reduce information overlap (Chen et al., 2012; Uğuz, 2012), but there may be information loss, and the weight of indicators cannot be determined. The index optimization method based on relative importance is mainly based on the importance of the amount of information carried by the indicator

### Table 4. Optimal Results of Indicators.

| Serial number | Indicator | Orthogonal normalization of eigenvectors to take their absolute values | Information contribution sensitivity |
|---------------|-----------|-------------------------------------------------------------------|-------------------------------------|
|               |           | $|p_j|$ | $|p_j|$ | $|p_j|$ | $\beta_j$ |
| 1             | $X_1$     | .2280 | .5124 | .0231 | .1654 |
| 2             | $X_2$     | .3917 | .1167 | .0328 | .2148 |
| 3             | $X_3$     | .0996 | .0876 | .0342 | .0618 |
| 4             | $X_4$     | .3678 | .2321 | .0239 | .2122 |
| 5             | $X_5$     | .6317 | .4582 | .0625 | .3715 |
| 6             | $X_6$     | .1565 | .0287 | .0375 | .0860 |
| 7             | $X_7$     | .2302 | .5372 | .0096 | .1678 |
| 8             | $X_8$     | .1289 | .2220 | .1862 | .1004 |
| 9             | $X_9$     | .1537 | .1769 | .6955 | .1473 |
| 10            | $X_{10}$  | .1748 | .2200 | .6737 | .1604 |
| 11            | $X_{11}$  | .1772 | .1047 | .1287 | .1104 |
| 12            | $X_{12}$  | .0896 | .0666 | .0484 | .0504 |
| 13            | $X_{13}$  | .2502 | .1513 | .0033 | .1428 |

*Source. Wang and Tian (2021).*

### Table 5. Indicator Screening and Weighting Based on Cumulative Information Contribution Sensitivity.

| Serial number (1) | Indicators (2) | Indicators sorted by $\beta_j$ value | Cumulative information contribution sensitivity $\gamma_m$ (%) | Retained and removed indicators | Weights |
|-------------------|----------------|-------------------------------------|-------------------------------------------------------------|--------------------------------|---------|
| 1                 | $X_1$          | $X_3$ (0.3715)                     | 18.66                                                       | Retained $X_3$                 | 0.0831  |
| 2                 | $X_2$          | $X_2$ (0.2148)                     | 29.44                                                       | Retained $X_2$                 | 0.1079  |
| 3                 | $X_3$          | $X_4$ (0.2122)                     | 40.10                                                       | Retained $X_4$                 | 0.0310  |
| 4                 | $X_4$          | $X_7$ (0.1678)                     | 48.53                                                       | Retained $X_7$                 | 0.1066  |
| 5                 | $X_5$          | $X_1$ (0.1654)                     | 56.84                                                       | Retained $X_1$                 | 0.1866  |
| 6                 | $X_6$          | $X_{10}$ (0.1604)                  | 64.89                                                       | Retained $X_{10}$              | 0.0432  |
| 7                 | $X_7$          | $X_9$ (0.1473)                     | 72.29                                                       | Retained $X_9$                 | 0.0843  |
| 8                 | $X_8$          | $X_{13}$ (0.1428)                  | 79.46                                                       | Retained $X_{13}$              | 0.0504  |
| 9                 | $X_9$          | $X_{11}$ (0.1104)                  | 85.00                                                       | Retained $X_{11}$              | 0.0740  |
| 10                | $X_{10}$       | $X_8$ (0.1004)                     | 90.05                                                       | Retained $X_8$                 | 0.0806  |
| 11                | $X_{11}$       | $X_6$ (0.086)                      | 94.37                                                       | Removed $X_6$                  | 0.0554  |
| 12                | $X_{12}$       | $X_3$ (0.0618)                     | 97.47                                                       | Removed $X_3$                  | 0.0253  |
| 13                | $X_{13}$       | $X_{12}$ (0.0504)                  | 100.00                                                      | Removed $X_{12}$               | 0.0717  |

*Source. Wang and Tian (2021).*

*Note. The weights in the last column correspond to the indicators in the second column.*
relative to the evaluation results (Fouskakis et al., 2020; Kazemi, 2020), and the problem of information overlap cannot be avoided. This study proposes an I-PCA method, which calculates the indicators’ information contribution rate and judges the retention or deletion of indicators according to the set threshold. The research results show that the optimized index of this method can ensure the maximum information-carrying, avoid information overlap between indicators, and obtain the weight of each indicator. This method overcomes the disadvantage that the traditional PCA method cannot give weight and expands the method system of objective weighting.

It can be seen from the indicator selection results that we have deleted “Knowledge of laws and regulations,” “Kind to others,” and “Driving control behavior.” The information contribution rates of these three indicators are 4.32%, 3.1%, and 2.23%, respectively. The original indicator set has less information or overlaps with the information carried by other indicators. In actual traffic activities, the “traffic rules” include most of the “knowledge of traffic laws and regulations,” and the information between the two overlaps to a high degree, so the sensitivity of their information contribution is low. Similarly, the low sensitivity of the indicator of “Kind to others” to the information contribution of traffic literacy may be because the driver’s values are influenced by his social background and personal experience (Kaçan et al., 2019), which have formed the emotional intensity and behavioral standards for treating others. Driving control behavior occurs during the entire driving process, and its importance is self-evident. However, with the advancement of technology, vehicles are becoming increasingly automated (Hancock et al., 2020); it is easier for the driver to control the vehicle. In addition, the driving test standards in different countries and regions have been improved, and drivers who pass the test can master driving skills proficiently (Beanland et al., 2013; Williams, 2017).

These are the reasons for the relatively low degree of impact of the indicator on traffic literacy. It can be seen from Table 4 that the weights of Mechanical common sense, Safety awareness, and Values are .1079, .1066, and .1866, respectively. The mastery of “mechanical common sense” reflects the driver’s understanding of the vehicle, including the performance and operating instructions of the vehicle (Koo et al., 2015), which is a prerequisite to ensuring the driver’s safe driving. Safety awareness and values belong to traffic awareness. Research shows that traffic attitudes are critical to preventing traffic accidents (Martinussen et al., 2013; Yao et al., 2019). Safety awareness and values are part of a traffic attitude. When drivers have safety awareness and good values, they will take timely measures to avoid possible traffic accidents. Values exist at different levels of an individual and are dynamically changing with ground and personal experience (Kaçan et al., 2019), which serves as the theoretical foundation for the research on evaluating driver traffic literacy. The traffic management department can formulate a targeted driver traffic literacy improvement strategy for an indicator with a high information-sensitive contribution rate. It can quickly improve the driver’s traffic literacy level, activate the driver’s subjective initiative to effectively avoid traffic accidents, and promote the road traffic system to maintain a safe state.

The research aims to eliminate information redundancy indicators, and based on the I-PCA method, the concept of Information contribution rate is proposed, and the indicators are optimized and weighted. Firstly, a questionnaire is designed to obtain the original data based on the constructed traffic literacy indicator. Secondly, according to the calculation results, three indicators with low sensitivity to information contribution, such as Knowledge of laws and regulations, Kind to others, and Driving control behavior, are excluded, and nine indicators, such as Traffic rules that can reflect most information on traffic literacy are retained. Finally, all indicators are weighted according to the calculation results of the information contribution rate to reflect the importance of the indicators.

Acknowledgments

The author would like to thank the No. 6 Traffic Police Brigade of Zhengzhou Public Security Bureau for their support in obtaining data for this study. Thank you to the drivers surveyed for their responses to the questionnaire. Thanks to the authors’ colleagues who assisted in collecting the questionnaires during the paper’s writing process.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.
References

Ahmad, W. N. K. W., Rezaei, J., Sadaghiani, S., & Tavasszy, L. A. (2017). Evaluation of the external forces affecting the sustainability of oil and gas supply chain using best worst method. *Journal of Cleaner Production*, 153, 242–252. https://doi.org/10.1016/j.jclepro.2017.03.166

Aijaz, R. U. M. I. (2019). *India’s peri-urban regions: The need for policy and the challenges of governance* (ORF Issue Brief No. 285, March 2019). Observer Research Foundation. Retrieved March 22, 2022, from https://enrd.ec.europa.eu/sites/enrd/files/enrd_publications/publi-enrd-rr-26-2018-en.pdf

Arthur, W., Jr., & Doverspike, D. (2001). Predicting motor vehicle crash involvement from a personality measure and a driving knowledge test. *Journal of Prevention & Intervention in the Community*, 22(1), 35–42. https://doi.org/10.1080/1088230109511209

Beanland, V., Goode, N., Salmon, P. M., & Lenné, M. G. (2013). Is there a case for driver training? A review of the efficacy of pre-and post-licence driver training. *Safety Science*, 51(1), 127–137. https://doi.org/10.1016/j.ssci.2012.06.021

Bybee, R., McCrae, B., & Laurie, R. (2009). *PISA 2006: An assessment of scientific literacy*. *Journal of Research in Science Teaching*, 46(8), 865–883. https://doi.org/10.1002/tea.20333

Chai, C., Zhou, Z., Yun, Y., Chen, C., & Yang, L. (2020, September). Assessing the relationship between self-reported driving behavior, psychology and risky driving based on GPS trajectory data from car-hailing apps [Conference session]. 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), Rhodes, Greece (pp. 1–7). [Online]. IEEE. https://doi.org/10.1109/ITSC45102.2020.9294472

Chen, H. (2021). Method of screening evaluation indicators based on anti-image correlation matrix [Online]. *China Management Science*, 1–10. (In Chinese). https://doi.org/10.16381/j.cnki.issn1003-207x.2020.0969

Chen, H., & Chi, G. (2015). Urban green development evaluation indicator system model based on clustering-rough set and application. *ICIC Express Letters. Part B: Applications: An International Journal of Research and Surveys*, 6(10), 2649–2654.

Chen, S. F., Wang, S., & Chen, C. Y. (2012). A simulation study using EFA and CFA programs based the impact of missing data on test dimensionality. *Expert Systems with Applications*, 39(4), 4026–4031. https://doi.org/10.1016/j.eswa.2011.09.085

Deffenbacher, J. L. (2016). A review of interventions for the reduction of driving anger. *Transportation Research Part F: Traffic Psychology and Behaviour*, 42, 411–421. https://doi.org/10.1016/j.traf.2015.10.024

Di Ciommo, F., & Shiftan, Y. (2017). *Transport equity analysis*. *Transport Reviews*, 37(2), 139–151. https://doi.org/10.1080/01441647.2017.1278647

Fouskasik, D., Petarakos, G., & Rotous, I. (2020). A Bayesian longitudinal model for quantifying students’ preferences regarding teaching quality indicators. *Metron*, 78(2), 255–270. https://doi.org/10.1007/s40300-020-00175-5

Guo, Z., Shang, X., & Li, H. (2011). AHP-based safety assessment model for rail transit system. *Zhongguo Tiedao Kexue*, 32(3), 123–125.

Han, J., Singh, M., & Zhao, D. (2010). Road safety literacy for speakers of English as a foreign language: Educating novice drivers for the public’s health. *Literacy and Numeracy Studies*, 18(1), 52–66. https://doi.org/10.5130/lns.v18i1.1427

Hancock, R. A., Kajaks, T., Caird, J. K., Chignell, M. H., Mizobuchi, S., Burns, P. C., Feng, J., Fernie, G. R., Lavallière, M., Noy, I. Y., Redelmeier, D. A., & Vrkijan, B. H. (2020). Challenges to human drivers in increasingly automated vehicles. *Human Factors*, 62(2), 310–328. https://doi.org/10.1177/0018720819900402

Hungerford, H. R., & Peyton, R. B. (1976). *Teaching environmental education*. J. Weston Walton Publisher.

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2015), 20150202. https://doi.org/10.1098/rsta.2015.0202

Kaçan, B., Fındık, G., Üzümçioğlu, Y., Azık, D., Solmazer, G., Ersan, Ö., Özkan, T., Lajunen, T., Öz, B., Pashkevich, A., Pashkevich, M., Danielli-Mlyona, V., Georgioanni, D., Krasniqi, E. B., Krasniqi, M., Makris, E., Shubenkova, K., & Xheladini, G. (2019). Driver profiles based on values and traffic safety climate and their relationships with driver behaviors. *Transportation Research Part F: Traffic Psychology and Behaviour*, 64, 246–259. https://doi.org/10.1016/j.trf.2019.05.010

Kazemi, M. (2020). Partial correlation screening for varying coefficient models. *Journal of Mathematical Modeling*, 8(4), 363–376. https://doi.org/10.22124/jmmm.2020.15692.1379

Keyvan-Ekbatani, M., Kouvelas, A., Papamichail, I., & Papageorgiou, M. (2012). Exploiting the fundamental diagram of urban networks for feedback-based gating. *Transportation Research Part B: Methodological*, 46(10), 1393–1403. https://doi.org/10.1016/j.trb.2012.06.008

Koo, J., Kwac, J., Ju, W., Steinert, M., Leifer, L., & Nass, C. (2015). Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, 9(4), 269–275. https://doi.org/10.1007/s12008-014-0227-2

Kroll, V., Mackenzie, A. K., Goodge, T., Hill, R., Davies, R., & Crundall, D. (2020). Creating a hazard-based training and assessment tool for emergency response drivers. *Accident Analysis & Prevention*, 144, 105607. https://doi.org/10.1016/j.aap.2020.105607

Lajunen, T., & Summalu, H. (1995). Driving experience, personality, and skill and safety-motive dimensions in drivers’ self-assessments. *Personality and Individual Differences*, 19(3), 307–318. https://doi.org/10.1016/0191-8869(95)00068-I

Li, L., Lin, Y. L., Zheng, N. N., Wang, F. Y., Liu, Y., Cao, D., Wang, K., & Huang, W. L. (2018). Artificial intelligence test: A case study of intelligent vehicles. *Artificial Intelligence Review*, 50(3), 441–465. https://doi.org/10.1007/s10462-018-9631-5

Li, Z., Chen, C., & Wang, K. (2011). Cloud computing for agent-based urban transportation systems. *IEEE Intelligent Systems*, 26(1), 73–79. https://doi.org/10.1109/MIS.2011.10

Liu, G., & Zhao, D. (2020). Challenges to human drivers in increasingly automated vehicles. *Human Factors*, 62(2), 310–328. https://doi.org/10.1177/0018720819900402

Mizobuchi, S., Burns, P. C., Feng, J., Fernie, G. R., Lavallière, M., Noy, I. Y., Redelmeier, D. A., & Vrkijan, B. H. (2020). Challenges to human drivers in increasingly automated vehicles. *Human Factors*, 62(2), 310–328. https://doi.org/10.1177/0018720819900402

Xheladini, G. (2019). Driver profiles based on values and traffic safety climate and their relationships with driver behaviors. *Transportation Research Part F: Traffic Psychology and Behaviour*, 64, 246–259. https://doi.org/10.1016/j.trf.2019.05.010
Sujon, M., & Dai, F. (2021). Social media mining for understanding traffic safety culture in Washington state using Twitter data. *Journal of Computing in Civil Engineering, 35*(1), 04020059. https://doi.org/10.1061/(asce)cp.1943-5487.0000943

Tian, K. (2022). Construction and empirical analysis of citizens’ water literacy evaluation index system: A structural equation model. *Water Resources Management, 36*(4), 1393–1411. https://doi.org/10.1007/s11269-022-03089-1

Touahmia, M. (2018). Identification of risk factors influencing road traffic accidents. *Engineering, Technology & Applied Science Research, 8*(1), 2417–2421. https://doi.org/10.3846/16484142.2014.915581

Treffner, P., Barrett, R., & Petersen, A. (2002). Stability and skill in driving. *Human Movement Science, 21*(5–6), 749–784. https://doi.org/10.1016/j.humov.2001.03.013

Uğuz, H. (2012). A hybrid system based on information gain and principal component analysis for the classification of transcranial Doppler signals. *Computer Methods and Programs in Biomedicine, 107*(3), 598–609. https://doi.org/10.1016/j.cmpb.2011.03.013

United Nations Educational, Scientific, and Cultural Organization (UNESCO), Paris (France). (1978). Intergovernmental Conference on Environmental Education Organized by UNESCO in Co-operation with UNEP (Tbilisi, USSR, 14–26 October 1977). Final Report. UNIPUB.

Üzümcüoğlu, Y., Özkan, T., Wu, C., & Zhang, H. (2020). Traffic climate and driver behaviors: The moderating role of driving skills in Turkey and China. *Journal of Safety Research, 75*, 87–98. https://doi.org/10.1016/j.jsr.2020.08.004

Vaa, T. (2007). Modelling driver behaviour on basis of emotions and feelings: Intelligent transport systems and behavioural adaptations. In P. C. Cacciabue (Ed.), *Modelling driver behaviour in automotive environments* (pp. 208–232). Springer.

Vogelpohl, T., Kühn, M., Hummel, T., & Vollrath, M. (2019). Asleep at the automated wheel: Sleepiness and fatigue during highly automated driving. *Accident Analysis & Prevention, 126*, 70–84. https://doi.org/10.1016/j.aap.2018.03.013

Wang, L. B. (2010). *Multivariate Statistical Analysis: Model Cases and Application of SPSS*. Economic Science Press. (In Chinese)

Wang, L., & Tian, K. (2021). Optimisation of driver’s traffic literacy evaluation index from the perspective of information contribution rate. *Journal of Mathematics, 2021, 9503037*. https://doi.org/10.1155/2021/9503037

Wilke, R. (1995). Environmental literacy and the college curriculum-colleges and universities have a challenge to meet. *EPA Journal, 21*(2), 28–30.

Williams, A. F. (2017). Graduated driver licensing (GDL) in the United States in 2016: A literature review and commentary. *Journal of Safety Research, 63*, 29–41. https://doi.org/10.1016/j.jsr.2017.08.010

Wishart, D., Rowland, B., & Somoray, K. (2019). Safety citizenship behavior: A complementary paradigm to improving safety culture within the organizational driving setting. In N. J. Ward, B. Watson, & K. Fleming-Vogl (Eds.), *Traffic safety culture* (pp. 145–171). Emerald Publishing Limited.

Xiong, H., Fan, C., Chen, H., Yang, Y., Antwi, C. O., & Fan, X. (2022). A novel approach to air passenger index prediction: Based on mutual information principle and support vector regression blended model. *SAGE Open, 12*(1), 1–11. https://doi.org/10.1177/21582440211071102

Yan, Z., Lu, X., & Hu, W. (2020). *Analysis of factors affecting traffic accident severity based on heteroskedasticity ordinal Logit* [Conference session]. ICTE 2019, Chengdu, China (pp. 422–435). American Society of Civil Engineers. https://doi.org/10.1061/9780784482742.048

Yao, Y., Carsten, O., Hibberd, D., & Li, P. (2019). Exploring the relationship between risk perception, speed limit credibility and speed limit compliance. *Transportation Research Part F: Traffic Psychology and Behaviour, 62*, 575–586. https://doi.org/10.1016/j.trf.2019.02.012

Zaidan, R. A., Alamoodi, A. H., Zaidan, B. B., Zaidan, A. A., Albahri, O. S., Talal, M., Garfan, S., Sulaiman, S., Mohammed, A., Hashim, Z., Malik, R. Q., & Ameen, H. A. (2022). Comprehensive driver behaviour review: Taxonomy, issues and challenges, motivations and research direction towards achieving a smart transportation environment. *Engineering Applications of Artificial Intelligence, 111*, 104745. https://doi.org/10.1016/j.engappai.2022.104745

Zelikov, V. A., Denisov, G. A., Dorokhin, S. V., Razgonyaeva, V. V., & Zelikova, N. V. (2019). Improvement of the current version of road traffic regulations of the Russian federation as a promising approach to road safety. In E. Popkova (Ed.), *Ubiquitous computing and the internet of things: Prerequisites for the development of ICT* (pp. 1081–1088). Springer. https://doi.org/10.1007/978-3-030-13397-9_111

Zepf, S., Dittrich, M., Hernandez, J., & Schmitt, A. (2019, May). *Towards empathetic car interfaces: Emotional triggers while driving* [Conference session]. Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, Scotland (pp. 1–6). https://doi.org/10.1145/3290607.3312883

Zhang, L., Cui, B., Yang, M., Guo, F., & Wang, J. (2019). Effect of using mobile phones on driver’s control behavior based on naturalistic driving data. *International Journal of Environmental Research and Public Health, 16*(8), 1464. https://doi.org/10.3390/ijerph16081464

Zhang, W., & Dong, W. (2013). *SPSS statistical analysis advanced course*. Higher Education Press. (In Chinese)

Zhang, Y., Chen, Z. W., & Zhang, Q. S. (2015). Building of assessment indicator system of safe transport. *Journal of Transportation Research, 1*(6), 8–13. (In Chinese)