Accident Analysis and Prevention 153 (2021) 106038

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

The mediating effect of driver characteristics on risky driving behaviors moderated by gender, and the classification model of driver's driving risk

Xiaolin Song [a], Yangang Yin [a], Haotian Cao [a,*], Song Zhao [b], Mingjun Li [a], Binlin Yi [a]

[a] State Key Laboratory of Advanced Design and Manufacturing for the Vehicle Body, Hunan University, No.2 Lushan South Rd, Yuelu District, Changsha, Hunan 410082, People's Republic of China

[b] Department of Mechanical & Mechatronics Engineering, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada

ARTICLE INFO

Keywords:
Driver driving risk
Sensation seeking
Risk perception
Risky driving behaviors
Structural equation model
SHRP 2

ABSTRACT

High-risk drivers are more likely to be involved in traffic accidents, and the driving risk level of drivers could be affected by many potential factors, such as demographics and personality traits. Based on the Structural Equation Model (SEM), this study involves a sample of 3150 drivers from the Strategic Highway Research Program 2 (SHRP 2), to explore the relationships among drivers' demographic characteristics (gender, age, and cumulative driving years), sensation seeking, risk perception, and risky driving behaviors. More specifically, the mediation model of driver characteristics on risky driving behaviors moderated by gender is constructed by the SEM. The results show that the effects of driving experience on risky driving behaviors are partially mediated by sensation seeking and risk perception for male drivers, while those are completely mediated by sensation seeking and risk perception for female drivers. Moreover, the development trend of risky driving behavior engagements declines greater with the growing of driving experience for female drivers than male drivers. Finally, a classification model of the driver's driving risk is proposed by the Random Forest classifier, in which the driving risk level of the driver evaluated by the crash and near-crash rate could be classified through the driver's self-reported demographics, sensation seeking, risk perception, and risky driving behaviors. The classification accuracy achieves up to 90 percent, which offers an alternative approach to identifying potential high-risk drivers to reduce property losses, injuries, and death caused by traffic accidents.

1. Introduction

Traffic accidents lead to numerous economic losses, body injuries, and even death. According to the report by the World Health Organization (2018), it is the 8th leading cause of death for people of all ages. The number of road traffic deaths remains unacceptably high. More than 1.35 million people, involving pedestrians, cyclists, and motorcyclists, died in traffic accidents every year. It should be noted that the Sustainable Development Goals (SDG) section of 3.6 to halve road deaths and injuries by 2020 will not be met without drastic action. Therefore, there is a lot of work to do in improving traffic safety. A high incidence of traffic accidents can be attributed to road infrastructure construction, traffic flow density, weather conditions, and the behavior of drivers, etc. (Juhnke et al., 1995; World Health Organization, 2018). Furthermore, characteristics of the driver, highway, and vehicle also contribute to injury severity in traffic accidents (de Oña et al., 2011).

It should be noted that human and behavioral factors play key roles in road traffic accidents (Lawton et al., 1997). Existing reports indicate that the percentage of crashes involving certain types of driving errors by the driver was as high as 94 % (Singh, 2015). More than 90 % of the fatal traffic accidents were caused by human errors (World Health Organization, 2018), and there is a strong correlation between driver violation and reported traffic accidents (Parker et al., 1995). Moreover, 78 % of crashes in the *U.S. 100-vehicle study* database were related to the driver's inattention (Dingus et al., 2006). And, aggressive driving maneuvers contribute to more severe injuries, which were involved in more than 55 % of all fatal accidents (Du et al., 2018; Haleem and Gan, 2015).

Nevertheless, different types of driving behavior occurrences may be related to psychological factors. Previous studies have linked drivers' personality traits, psychological factors, driving behaviors, and crash participation (Ma et al., 2010; Ram and Chand, 2016; Zhao et al., 2019). Drivers' deliberate driving violations and unintentional mistakes are

* Corresponding author.
E-mail addresses: jqysxl@hnu.edu.cn (X. Song), yinyangang@hnu.edu.cn (Y. Yin), caohaotian@hnu.edu.cn (H. Cao), song.zhao@uwaterloo.ca (S. Zhao), mingjunl@hnu.edu.cn (M. Li), yibinlin@hnu.edu.cn (B. Yi).

https://doi.org/10.1016/j.aap.2021.106038
Received 15 March 2020; Received in revised form 14 January 2021; Accepted 6 February 2021
Available online 22 February 2021

respectively associated with the more unemotional and more impulsive aspects of psychologies (Panayiotou, 2015). Drivers with higher levels of anger traits are more likely to be provoked and tend to engage in more aggressive behaviors (Bogdan et al., 2016).

Therefore, the driver's demographics, personality traits, and driving behaviors play crucial roles in traffic safety. Exploring relationships among driver's demographics, personality traits and risky driving behaviors could be helpful to identify specific factors related to driving safety. Meanwhile, these factors can be further utilized to identify high-risk drivers. Such that, it is expected to reduce property losses, injuries, and even death caused by traffic accidents.

1.1. Risk perception

Perception of risk is a subjective judgment of the severity of a particular risk in traffic psychology, which can influence decision-based behaviors (Deery, 1999) (e.g. assessing the participant's associated risk with running a red light). Risk perception depends on various factors such as driving training, driving experience, age, etc. For instance, driving training and experience increases the risk perception of drivers (Deery, 1999; Rosenbloom et al., 2008). Moreover, risk perception can be affected by anger and fear moderated by reappraisals (Lu et al., 2013).

Some scholars have discussed the relationship between the risky driving attitude and risk perception. Attitudes can be defined as "tendencies to evaluate an entity with some degree of favor or disfavor, ordinarily expressed in cognitive, affective and behavioral responses'' (Eagly and Chaiken, 1993), while risk-taking attitude reflects the driver's tendency towards risky driving intention, which will then affect the driver's risky driving behaviors and be regulated by risk perception. Ram and Chand (2016) conclude that risk perception could inhibit risky driving attitude (e.g. the tendency to engage in risky driving behaviors) directly, as well as mediate the relationship between perception of road task and risky driving attitude.

Besides, risk perception is found to be negatively correlated with risky driving behaviors. Drivers with a higher level of risk perception are less likely to be engaged in risky driving behaviors (Machin and Sankey, 2008; Mills et al., 2008). Meanwhile, risk perception mediates the effect of age and gender on risky driving behaviors (Rhodes and Pivik, 2011). Moreover, Ma et al. (2010) have found that risk perception seems not to have direct effects on risky driving behaviors, while two kinds of risk perception scales, the likelihood of crash and concern (e.g. how concerned are you about traffic risks that will result in others' victimization), have indirect effects on rule violations and speeding through drivers' attitudes. However, there are still some controversies on this conclusion. For instance, there is no significant causal relationship that risk perception negatively affects risky driving behaviors when controlling drivers' attitudes, and opposite conclusions have also been found in other studies (Cohn et al., 1995; Ulleberg and Rundmo, 2003). One explanation for these phenomena is that the risk perception ability of the crash may be a consequence of risky driving behaviors (Horvath and Zuckerman, 1993), which cause those risk driving behaviors to be positively correlated with risk perception.

1.2. Sensation seeking

The need for stimulation will be presumably manifested in many aspects of behaviors, including sensation, social contact, and thrill-seeking activities. The sensation-seeking scale (SSS) was developed in an early attempt to assess individual differences in optimal levels of stimulation (Zuckerman, 1971), which evaluates the tendency of people to seek experiences with strong feelings and take multiple risks (Burns and Wilde, 1995). Sensation-seeking can be decomposed into four representative factors including *Thrill and Adventure Seeking (TAS,* a desire to engage in outdoor sports or other activities*)*, *Experience Seeking (ES,* whose essence is "experience for its own sake."*)*, *Disinhibition (DIS,* the loss of social inhibitions*)*, and *Boredom Susceptibility (BS,* a dislike of repetition of experience*)*. It should be noted that sensation seeking shares some conceptual similarities with excitement seeking. As a sub-factor of Extraversion, excitement seeking is a secondary trait from the perspective of the Big Five Model (BFM) (Costa and McCrae, 1985), and is characterized by a need for thrills, risk-taking, and strong stimulation (Piedmont, 1998). For clarification purposes, we adopt the concept of sensation seeking defined as the desire and participation in various, novel, complex, arousing sensations and experiences by Zuckerman (1984) as the target secondary personality trait in this study, which has been proven to be associated with risky driving behaviors in some empirical studies (Jonah, 1997).

In terms of the impact of sensation seeking on driving safety, a case-control study comparing drivers convicted and not convicted of offenses such as speeding or reckless driving also yielded significant differences in sensation-seeking measures (Furnham and Saipe, 1993). Sensation seeking is also one of the personality traits that are potentially related to risky driving behaviors. The stronger the sensation-seeking tendency, the more likely the driver is to pursue stimulation, which ultimately leads to the adoption of risky driving behaviors (Zuckerman, 1994). Jonah et al. (2001) have found that drivers with high sensation-seeking were more likely to engage in speeding, drinking, and reported aggressive driving habits, etc. Furthermore, young drivers' sensation-seeking is positively associated with risky driving intention, attitude, and speeding driving behavior (Cestac et al., 2011). Bachoo et al. (2013) have investigated 306 post-graduate students in South Africa to report the correlation between self-reported sensation-seeking and targeted risky driving behaviors, which indicates that seeking stimulation is associated with risky driving behaviors of the young driver group.

However, the effect of sensation-seeking on risky driving seems to be indirect. The direct effects of personality traits, especially sensation seeking, sensitivity to punishment, reward, and impulsivity on driving outcomes, were few. Still, these personality traits have significant correlations with aberrant driving behaviors, which shows that sensation seeking is a distal but important predictor of negative driving outcomes (Constantinou et al., 2011). Sensation-seeking also has a strong correlation with risk perception (Rundmo and Iversen, 2004). Doran et al. (2011) have found a negative relationship between sensation seeking and risk perception in adolescent smoking. However, studies focus on the effect of these two factors in the area of traffic safety are still few. Thus, this study will further explore the relationship between sensation seeking and risk perception regarding traffic safety.

1.3. Demographic characteristics

Drivers' demographical factors are usually treated as a series of important independent or cooperative variables that affect driving behaviors in many ways. Risky driving behaviors and crash accidents are related to demographical factors such as age, gender, driving experience, and education level, with gender and age being the most significant factors (Evans and Wasielewski, 1982). For example, male drivers and teenage drivers are reported to be engaged in risky driving behaviors more frequently than female drivers and adult drivers (Rhodes and Pivik, 2011). Moreover, female drivers commit fewer violations but more errors than male drivers (De Winter and Dodou, 2010). Wu et al. (2014) also have found that young drivers are more likely to be involved in crashes or collision-related events.

Some researchers have explored relationships among different demographics and certain psychological factors. They believe that the driver's demographic characteristics, such as gender and age, can significantly affect the personality traits (Cestac et al., 2011; Harbeck and Glendon, 2018), which is helpful to understand the role of demographical characteristics in driving safety. For instance, the time since license impairs the occurrence of accidents in general. However, when the type of driver's license changes, the accident rate will increase all of a sudden, which may be explained by the driver's psychological factors, for example, showing off (Harbeck and Glendon, 2018). Rhodes and Pivik (2011) also have concluded that positive affect and risk perception mediated the relationships of age and gender on risky driving behaviors independently.

Moreover, Santamariña-Rubio et al. (2014) have used the Poisson regression model to show that male drivers are more likely to suffer serious injuries among younger driver groups, while female drivers show a higher level of risk in the old group. This potential phenomenon may be attributed to an interaction between gender and age in road traffic injury risk. In the study of Rhodes and Pivik (2011), interactions of positive affect and perceived risk with gender and age have indicated that positive affect could be more likely to predict risky driving behaviors for teenage and male drivers, other than for adults and female drivers. It suggests that the influence of potential factors on risky driving behaviors could be influenced by age and gender. Nevertheless, this finding involves an inevitable problem that the effects of driver's age on risky driving behaviors could be different between genders, it may draw wrong conclusions regardless of the age interaction. Overall, it is necessary to consider the interaction of other demographic variables while studying the impact of specific demographical characteristics on risky driving behaviors.

1.4. Risky driving behavior

Drivers who often engage in risky driving behaviors pose a great danger to themselves, as well as other road users. General unsafe driving actions include active, conscious rule violations, as well as errors due to inexperience, momentary mistakes, or inattention. Therefore, it is urgent and necessary to develop a methodology that can measure the frequency of these behaviors they committed, and determine which actions are the most likely to predict traffic collision involvement (Cordazzo et al., 2014).

The driving behavior questionnaire (DBQ) developed by the Manchester driver behavior group is an often-used toolbox to measure abnormal driving behaviors (Reason et al., 1990). So far, the self-reported Manchester DBQ is one of the most widely used approaches for measuring behavior and bad habits in daily driving (af Wåhlberg et al., 2011). Especially, the three-component structure of the DBQ, which includes *Harmless lapse*, *Dangerous erro*r, and *Violation*, is widely used in earlier studies. Violation is defined as deliberate deviations from those practices believed necessary to maintain the safe operation of a potentially hazardous system (e.g. How often the participant has disregarded the speed limits). Lapse and error are defined as the failure of planned actions to achieve their intended consequences. Lapse is the unwitting deviation of action from intention (e.g. How often the participant has forgotten where they left their car in a parking lot), while the error is the departure of planned actions from some satisfactory path towards a desired goal (e.g. How often the participant has underestimated the speed of an oncoming vehicle) (Parker et al., 1995). Moreover, subsequent researchers developed a four-component structure. They have subdivided the *Violation* into *Ordinary violation* and *Aggressive violation* that has emotional components (e.g. How often the participant has got involved in unofficial 'races' with other drivers) (Lawton et al. (1997). Lajunen et al. (2004) also have analyzed the DBQ in Finland and the Netherlands through the four-factor structures (*Aggressive violation, Ordinary violation, Error,* and *Lapse*). However, the results of the four-factor structures show that the structure of the DBQ found in Finland and the Netherlands were congruent but not perfect with the target structure found in Britain, which indicates that the structure of the DBQ varies from different areas and even different times. Therefore, the structure should be analyzed and carefully chosen according to practical situations to ensure the validity of the scale when using the DBQ.

It is found that the DBQ score was correlated with objectively measured risky driving behaviors to certain extents (Zhao et al., 2012). Besides, previous studies have shown that the DBQ score is correlated with driving crashes. For example, violation scores are often reported to be positively correlated with crash rates (Parker et al., 1995). While in a recent study of the Strategic Highway Research Program 2 (SHRP 2), 22.4 % of participants who were classified into the 'violating unsafe drivers' group were involved in a recorded traffic offense. This percentage is 1.8 times higher than the average of the other groups, but the violation scores of the DBQ only exhibit small correlations with recorded crashes, as well as small-to-moderate correlations with recorded near-crashes (De Winter et al., 2018). Furthermore, the combination of error scores could predict crashes (af Wåhlberg et al., 2011). Wang et al. (2019) have divided the risk while driving into three levels by the crash and near-crash (CNC) rate and found that high-risk drivers are more likely to engage in inattention errors and ordinary violations. There also seems to be a strong correlation between age and risky driving behaviors, and young drivers are more likely to be engaged in risky driving behaviors in all categories of the DBQ (Blockey and Hartley, 1995).

1.5. Crash and near-crash

One possible way to identify high-risk drivers is to detect a crash involving kinetic energy transfer or dissipation. However, the crash is a small probability event in daily driving, so it is difficult to distinguish the risk level of drivers through the crash and find its correlation with driver's characteristics and behaviors. Researchers need alternative indicators of driver risk with a higher frequency of occurrence. As described by Bagdadi (2013), there is a correlation between the frequency of critical braking events and crashes. Similarly, the speed at the beginning of braking had a strong relationship with near-crash events (Zheng et al., 2014). Therefore, the near-crash evaluated by high breaking acceleration and low time-to-collision (TTC) is developed to solve the problem of low collision occurrences. Near-crash is defined as "a conflict situation that requires a rapid, severe evasive maneuver to avoid a crash" and has been proven to have a correlation and similar kinematic characteristics with the crash (Dingus et al., 2006).

Compared with self-reported collision events, the usage of the crash and near-crash (CNC) of naturalistic driving data has two advantages. First, the CNC of naturalistic driving data is more accurate than self-reported collision records because it is not generally affected by the driver's subjective evaluation criteria. Second, video records of naturalistic driving data can be repeatedly verified and analyzed. Such that, the CNC is widely used in the research of driving safety (Wang and Xu, 2019). Dingus et al. (2006) have explored the correlation between driver inattention and the CNC events, through the naturalistic driving study by using the 100-car naturalistic driving data. Moreover, the CNC rate can also be adopted into the prediction of high-risk drivers, for example, Guo and Fang (2013) have utilized the CNC rate as the main criterion for evaluating the overall risk of individual drivers and extract multiple factors to predict the risk level of drivers.

1.6. Research motivations, goals, and main contributions

Most previous studies have used linear regression or multi-group correlation analysis in the relationships among demographics, traffic psychology, and driving safety. However, there are still some problems that cannot be solved through these methods. On the one hand, it is hard to measure the latent psychological and behavioral factors. On the other hand, it is also difficult to investigate complex relationships among driver's demographics, sensation seeking, risk perception, and risky driving behaviors in-depth. Previous studies have mainly focused on the effects of age or driving experience on personality traits associated with risky driving behaviors, or the effects of age or driving experience on risky driving behaviors solely, rather than incorporating them into a combined model.

Therefore, the method of the Structural Equation Model (SEM) is adopted into the current study. SEM is a tool that can handle a large number of observed variables and specified latent variables by their linear combination and has been widely applied in various fields such as sociology, psychology, political science, tourism behavior research (Golob, 2003), as well as in traffic psychology and safety (Hamdar et al., 2008; Ram et al., 2016). SEM consists of multiple measurement models and a structural model. The unobservable latent factors can be represented by multiple observable indicators in measurement models, while the structural model can explore the complex relationship among latent variables. Therefore, the SEM absorbs the advantages of many other methods (e.g. factor analysis and regression analysis) and has its unique advantages. For example, there might exist certain factors (e.g. psychological factors) in the structural model that cannot be directly observed, while we can use the combination of multiple observable indicators to explain the unobservable latent variables. Also, the SEM is generally used to investigate complex relationships among various endogenous variables and exogenous variables (Lee et al., 2008). For example, suppose factor A has a certain influence on B, it might be interesting to investigate if there is any possibility that A has an indirect influence on B through a mediating factor C. Such conjectures can be explicitly verified in the SEM. Nevertheless, it should be noted that, since the SEM is a "confirmatory rather than exploratory" tool, analysts should specify the model that properly represents the real situation (Hamdar et al., 2008). Therefore, when it comes to building a model, almost all paths need full scrutiny and justification to improve the effectiveness of the model.

It is an undeniable fact that insufficient driving experience can lead to many driving safety problems, such as insufficient driving skills (Olsen et al., 2007) and poor ability to handle driving tasks (Sarkar and Andreas, 2004). However, considering that aggressive driving is responsible for a large proportion of accidents, driving experience might not be the direct cause of risky driving behaviors, it seems to be more of a remote factor, which is mediated by certain psychological factors (Rhodes and Pivik, 2011; Constantinou et al., 2011). Rhodes and Pivik (2011) have found that teenage male drivers are more likely to report enjoying these risky driving behaviors and perceiving them as less risky than their adults and female counterparts. Therefore, studying psychological factors related to driving experience will be helpful to understand the inhibitory relationship between driving experience and risky driving behaviors. As sensation-seeking is one of the very typical personality traits in traffic psychology, it is worthy to investigate its relationship with driving experience. Also, lower sensation-seeking scores were associated with the increase in age (Nordfjærn et al., 2010). While risk perception is another important psychological factor associated with driving experience, as novice drivers generally cannot judge driving risks accurately (Deery, 1999), and they tend to overestimate their driving skills (Rothbart et al., 2001), or underestimate the serious consequences of risky driving behaviors (Delhomme et al., 2009). Therefore, sensation seeking and risk perception are included as mediators in the mediation model, to explore their roles in the influence of driving experience on risky driving behaviors in this study.

One of the reasons for focusing on the drivers' personality traits (e.g. sensation seeking) is that they are usually correlated with abnormal driving behaviors, which can be important influence factors of risky driving behaviors. For young people, risk-related personality traits may be a significant factor implicated proximally in their accident risk. Besides, there are small positive correlations with *Thrill*, *Adventure Seeking*, and *Disinhibition* when regarding traffic offenses (Constantinou et al., 2011). Moreover, sensation seeking is an important personality trait that significantly affects driving behaviors, for example, impulse and sensation seeking are positively correlated with the DBQ scores (Jonah et al., 2001). Since driving is not only a means of transportation but also a way to mark their independence, young drivers tend to drive more dangerously and aggressively to demonstrate their masculinity, fearlessness, and competence when staying with other male companions (Jonah, 1986). Therefore, risky and aggressive driving appears to be the dominant human factor that places those young drivers at risk (Reason et al., 1990), which are affected by personality traits such as sensation seeking. However, sensation-seeking also seems not to completely affect risky driving behaviors directly, many researchers have found that personality traits also indirectly affect accident involvement through potential mediators, such as the attitude to driving safety and risk perception (Ulleberg and Rundmo, 2003). Risk perception and perceived benefits mediate the relationship between sensation-seeking and adolescent's risk-taking behavior (Zhang et al., 2016).

Therefore, several hypotheses are proposed that driving experience is a distant influence factor on risky driving behaviors, and it is mediated by sensation-seeking and risk perception. Meanwhile, sensation seeking is also a distant influence factor on risky driving behaviors, where risk perception plays a mediator role. Eventually, an interpretable path of the influence of driving experience on risky driving behavior mediated by sensation-seeking and risk perception is proposed. And the corresponding remote mediation model is constructed. Most of the previous studies usually treat driving experience and personality traits as isolated independent variables, while we will incorporate them into the SEM as dependent factors, to explore their relationships with risky driving behaviors.

More specifically, as depicted in Fig. 1, the hypotheses of this study via the SEM are listed as follows:

H1: Driving experience (DE) significantly affects sensation seeking (SS) and risk perception (RP) of the driver.

H2: Sensation seeking (SS) and risk perception (RP) significantly affect risky driving behaviors (RDB).

H3: The influences of driving experience (DE) on risky driving behaviors (RDB) are mediated by sensation seeking (SS) and risk perception (RP) indirectly.

It should be noted that gender could be another important demographical factor that may significantly influence the risky driving behaviors of the driver. It has been shown that male drivers typically report more violations (Reason et al., 1990), and engage in more risky driving behaviors (Rhodes and Pivik, 2011) than female drivers. Moreover, young males are known to have an overall higher crash risk than females of the same age (Bingham and Ehsani, 2012). However, we also notice that the previous results on risky driving behaviors between different genders are not entirely consistent after taking the interaction of age into account. For example, Beck et al. (2007) have reported that male drivers were at higher risk than female users regardless of age. Nevertheless, al-Balbissi (2003) has observed a higher risk of severe injuries among male drivers than their female counterparts in all age groups except the elderly. While another study conducted by Williams (2003) has shown that the incidence was higher in females than that in males in all age groups.

Besides, gender is also found to be related to personality traits, which might be an effective factor to explain the difference in risky driving behaviors between different genders. The strong influence of gender on accident propensity can be attributed in part to personality differences between male and female drivers (Norris et al., 2000). For instance, compared with female drivers, male drivers are more likely to be impulsive and high sensation-seeking (Arnett, 1994), especially in younger age groups. Ulleberg (2002) also has identified a subgroup of dangerous drivers, most of whom are male, characterized by low anxiety, low altruism, high aggression, and a sense of pursuit. Young male drivers may be more prone to accidents precisely because they have higher sensation seeking, lower sensitivity to punishment, while female drivers are less prone to risky driving because of a potential protective factor of sensitivity to punishment (Constantinou et al., 2011). Drivers' personality differences between different genders are also reflected in risky driving behaviors. There is a significant correlation between personality traits and abnormal driving behaviors, and the high-risk characteristic seems to reach a peak among young male drivers because they have certain personality traits that cause them to underestimate risk and take greater risks (Constantinou et al., 2011). According to the aforementioned studies, it can be found that personality traits that are

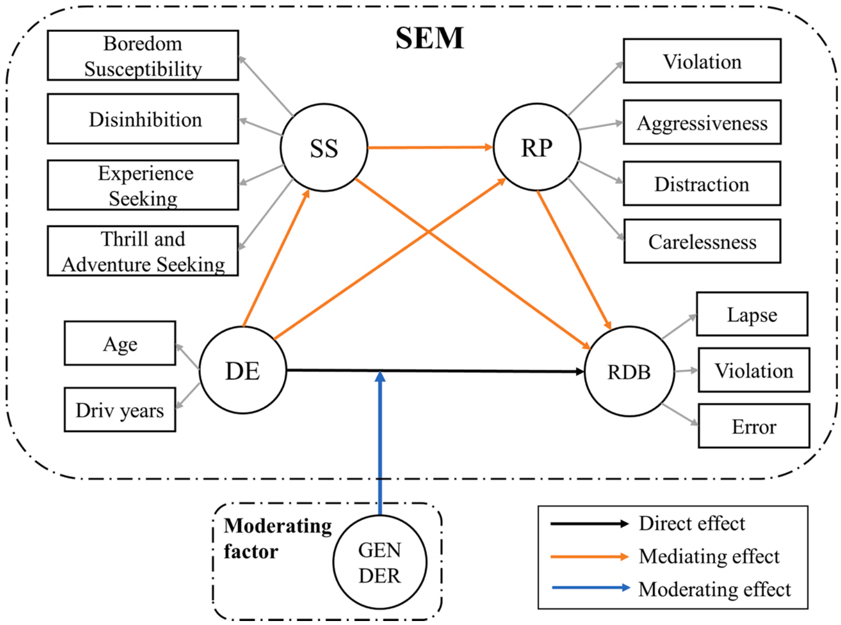

Fig. 1. The hypothesized SEM including mediating effects and moderating effects.
Note: Blackline represents the direct effect; Orange line represents the mediating effect; the blue line represents the moderating effect; DE: driving experience, SS: sensation seeking, RP: risk perception, RDB: risky driving behaviors; moderating factor: gender.

pronounced in male drivers are also frequently put forward in young drivers. As a result, the influence of drivers' driving experience on their personality traits, as well as personality traits on risky driving behaviors, might be related to gender. Therefore, gender, as a moderating variable, will be adopted in this study to explore its interaction with driving experience through sensation seeking.

As far as we know, most of the published studies usually regarded the factor of age and gender as descriptive covariates in statistical analyses, and they only focused on variations in personality traits and risky driving behaviors between different genders at a limited range of age or driving experience. Even if age or driving experience were considered, they are often divided into limited groups (e.g. novice drivers/experienced drivers or young/middle-aged/elderly drivers). However, the conclusion of personality traits and risky driving behaviors obtained at limited age groups may change when considering all age groups. Especially, the comparison of personality traits and risky driving behaviors between several age groups might be influenced by the grouping threshold. For instance, in the study of Constantinou et al. (2011), only drivers in their mid-20 s are considered to explore the relationship between personality traits, risky driving, and driving outcomes. Except for inhibition of sensation seeking, there is no strong relationship between driving experience and personality traits. Thus, they also suggest that future findings should be further investigated in other age groups to take the role of increased exposure and experience into consideration, and to explore the development trends in these traits and behaviors for each gender. Therefore, driving experience regarded as a continuous variable covers the drivers of all ages in this study, which includes the young people who have just got their licenses, as well as the elder people with sufficient driving experience. In brief, we will explore the development trends of risky driving behaviors with the change of sensation seeking, as well as the increase of driving experience between different genders. Finally, the following hypothesis is further proposed:

H4: Development trends of risky driving behavior engagements with the increase of driving experience are distinct between different genders.

Overall, the main purpose of this study is to take the driving experience as a continuous variable and explore its relationships among sensation seeking, risk perception, as well as risky driving behaviors. Incorporated with large-scale questionnaire surveys in the SHRP 2, a meditation model of driver's characteristics on risky driving behaviors is constructed via the SEM to verify the proposed hypothesis. After taking gender as an interactive variable, a moderated mediation model is constructed to explore the differences in driving experience on risky driving behaviors between different genders, which is also depicted in Fig. 1. Furthermore, we build a classification model of driver's driving risk by using the Random Forest (RF) method, in which the CNC rate is adopted as a training label and clustered into three risk levels including the low-risk, the middle-risk, and the high-risk, by the K-means clustering method. While drivers' demographics (age, cumulative driving years, gender), sensation seeking, risk perception, and risky driving behaviors are the input variables of the classification model. Thus, the driving risk level of drivers is expected to be recognized in the RF classifier.

A quick flow-chart of this study is depicted in Fig. 2, and the main contributions summarized in brief as follows,

1) The influences of driving experience on risky driving behaviors mediated by sensation seeking and risk perception are investigated.
2) The influence trends of driving experience on risky driving behaviors between different genders are further explored and analyzed.
3) A classification model of the driver's driving risk is proposed based on the self-reported questionnaires of the driver's demographics, sensation seeking, risk perception, and risky driving behaviors by using the RF algorithm.

2. Materials and methods

2.1. Data

The SHRP 2 Naturalistic Driving Study, which is one of the largest naturalistic driving-behavior studies to date, monitored over 3500 participants from 2010 to 2013 in the United States, in which the data were collected from 6 sites, including Seattle, Washington; Tampa, Florida; Buffalo, New York; Durham, North Carolina; State College, Pennsylvania; and Bloomington, Indiana. One way to provide the public with access to SHRP 2 data is the SHRP 2 Insight website (https://insight.shrp2nds.us/), which allows researchers to browse de-identified driving data and build queries to search what they are interested in.

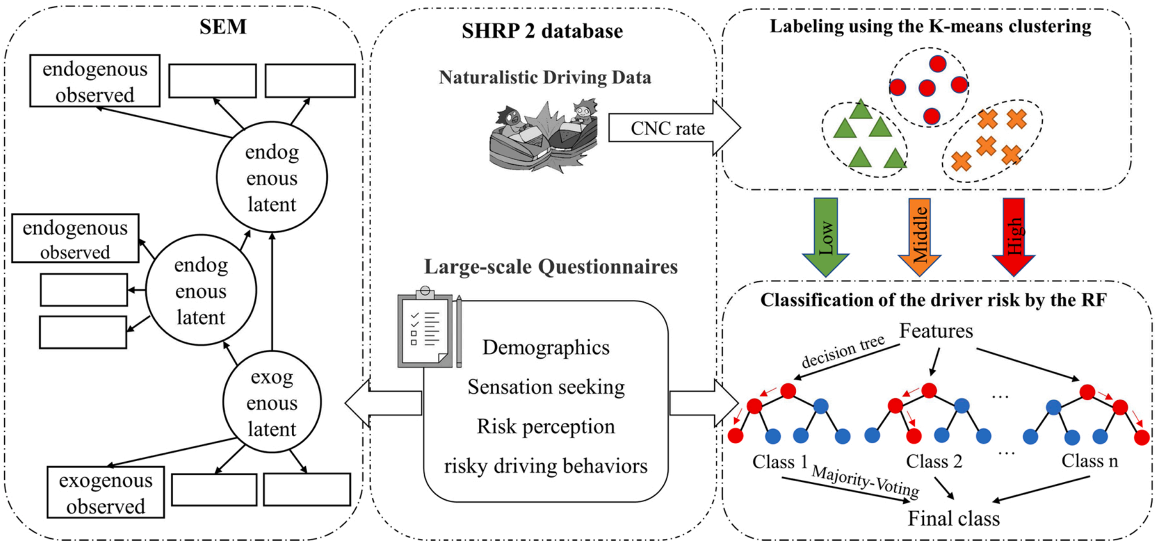

Fig. 2. The brief flow of the SEM and the Random Forest classification model in this study.
Notes: CNC: crash and near-crash.

The data in this study were extracted from self-reported questionnaires in the driver branch and naturalistic driving data in the event branch of the SHRP 2. Before collecting the naturalistic driving data, drivers were asked to fill out a series of questionnaires, such as the drivers' demographical questionnaire, sensation seeking scale (SSS), risk perception questionnaire (RPQ), and driving behavior questionnaire (DBQ). The naturalistic driving data includes the crash, near-crash, and baseline records. Questionnaire survey is one of the mainstream methods to collect drivers' data at a low cost. It is possible to predict drivers' accidents using the self-reported questionnaire of violations (Hatakka et al., 1997). Also, aggressive violations and ordinary violations collected by the questionnaire are positively related to the probability of being a high-risk driver (Wang and Xu, 2019).

However, large samples are needed to improve the accuracy and reliability of data and obtain a reliable solid conclusion. Thus, a large-scale naturalistic driving experiment (e.g. the SHRP 2) offers an opportunity. The naturalistic driving data (NDS) were obtained using the data acquisition system (DAS) from key-on to key-off for every trip. The onboard DAS collected four video messages (i.e. driver's face, driver's hand, front and rear road), vehicle network information (i.e. speed, brake, accelerator position), and additional sensors from the DAS (including global positioning system (GPS), forward radar, accelerometer). The Technical Coordination and Quality Control were conducted by the Virginia Tech Transportation Institute (VTTI) at Virginia Polytechnic Institute and State University (Dingus et al., 2016). The CNC records are extracted by analyzing the naturalistic driving data. The CNC rates in this study are calculated by using the CNC records and baseline records, which are necessary for comparisons with crashes and near-crashes, for example, to calculate risk estimates (Hankey et al., 2016). These baselines in the SHRP 2 were randomly selected with a goal of 20,000 baselines and a minimum of 1-baseline per driver. The number of baselines per driver is proportional to the total driving time larger than 8.064 km/h (5-mph). Therefore, the driver's CNC rate could be represented by the number of crash and near-crash records per baseline records. In summary, the variables involved in this study are driver's demographics, including driving experience (represented by age and cumulative driving years) and gender; sensation seeking and risk perception; risky driving behaviors; the CNC rate.

2.2. Participants

Participants comprise approximately half female (1701) and half male (1499) drivers, ages range from 16 to 99, and cumulative driving years range from 0 to 76. After deleting the drivers whose deletion of items in questionnaires are more than 20 %, the samples of the remaining 3150 drivers are included in the analysis, where the number of female drivers is 1499 which accounts for 47.5 % of the total number, while the number of male drivers is 1651 which accounts for 52.4 % of the total number. On account most American drivers are used to getting their licenses at age 16, there could be a strong correlation between age and cumulative driving years. Indeed, age is positively and significantly related to the cumulative driving years ($r = 0.94, p < 0.01$) according to the SHRP 2 data, which indicates they are related but not entirely overlapping. Therefore, similar to the study of Constantinou et al. (2011), age and cumulative driving years are regarded as two observed variables for representing the latent variable "driving experience" in the SEM. Besides, it should be noted that the age and cumulative driving years would be taken as continuous variables in the SEM according to the original SHRP 2 Insight data. Nevertheless, for the convenience of presenting a descriptive statistic, the statistical summary of drivers' demographic factors including the age groups, and cumulative driving years, is shown in Table 1.

Besides, considering small samples will easily lead to convergence failure, inappropriate solutions (illegal estimates), low parameter estimates, and incorrect standard errors, Bentler and Chou (1987) suggest that the sample size should be at least 5 times the estimated parameter (in the case of normal, no omission value and no extreme value, otherwise the sample size should be at least 15 times). Kline (2015) also recommend that the number of samples should up to 20 times. As incorporating with the large-scale data of the SHRP 2 Insight, the sample size of this study meets those requirements.

2.3. Methodologies

Generally, the SEM is composed of the structural model and measurement model. Measurement models use specific observed variables to characterize potential variables. The structural model explores the relationships between potential variables. Measurement models are normally specified in two sets of equations. The first set (the exogenous measurement model) is represented as follows:

$$\boldsymbol{X} = \boldsymbol{\Lambda}_{\mathbf{X}}\boldsymbol{\xi} + \boldsymbol{\delta} \tag{1}$$

where $\boldsymbol{X}$ denotes the vector of observed exogenous variables; $\boldsymbol{\Lambda}_{\mathbf{X}}$ denotes the matrix of structural coefficients for latent exogenous variables to

Table 1
Statistical table of driver demographics (involving 3150 drivers from the SHRP 2 Insight).

Variables	Category	Frequency	Percent	Variables	Category	Frequency	Percent
	16–19	529	0.168		0–4	776	0.246
	20–24	721	0.229		5–9	603	0.191
	25–34	405	0.129		10–19	339	0.108
Age	35–49	351	0.111	Cumulative driving years	20–34	338	0.107
	49–64	389	0.123		34–49	393	0.125
	65–79	553	0.176		49–64	551	0.175
	79-	188	0.060		64-	136	0.043

their observed indicator variables; $\boldsymbol{\xi}$ denotes the vector of latent exogenous variables; $\boldsymbol{\delta}$ denotes the vector of measurement error terms for observed variables.

The second (endogenous measurement model) is a set of equations which are summarized as follows,

$$\boldsymbol{Y} = \boldsymbol{\Lambda}_{\mathbf{Y}}\boldsymbol{\eta} + \boldsymbol{\varepsilon} \tag{2}$$

where Y denotes the vector of observed endogenous variables; $\boldsymbol{\Lambda}_{\mathbf{Y}}$ denotes the matrix of structural coefficients for latent endogenous variables to their observed indicator variables; $\boldsymbol{\eta}$ denotes the vector of latent exogenous variables; $\boldsymbol{\varepsilon}$ denotes the vector of measurement error terms for observed endogenous variables.

A structural model relating the exogenous latent variables and endogenous latent variables can be expressed as:

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta} \tag{3}$$

where $\boldsymbol{\eta}$ denotes the vector of the latent endogenous variable; $\mathbf{B}$ denotes the matrix of structural coefficients between endogenous latent variables; $\boldsymbol{\xi}$ denotes the vector of latent exogenous variables; $\boldsymbol{\Gamma}$ denotes the matrix of structural coefficients for exogenous latent variables to endogenous latent variables; $\boldsymbol{\zeta}$ denotes the unexplainable part of latent variables contained in the model.

In this study, observed variables of the driver's age and cumulative driving years are denoted by $\boldsymbol{X}$, observed variables of the questionnaires (SSS, RPQ, and DBQ) after dimensionality reduction are denoted by $\boldsymbol{Y}$. As for latent variables, the driving experience is an exogenous latent variable represented by $\boldsymbol{\xi}$. Sensation seeking, risk perception, and risky driving behaviors are endogenous latent variables represented by $\boldsymbol{\eta}$. A more detailed description of each variable in the SEM can be found in Table 2.

Moreover, the K-means clustering (MacQueen, 1967) and the Random Forest (RF) classifier (Breiman, 1996) are well-known unsupervised and supervised machine learning methods, respectively. The K-means clustering method is adopted to categorize the CNC rates into three groups, namely, the high-risk, the middle-risk, and the low-risk groups of drivers, the sample means of each cluster will be calculated as the new clustering center during the clustering, and the algorithm will end if the cluster center no longer changes or the number of iterations reached. While the RF classifier is used to classify the driver risk through drivers' demographics, sensation seeking, risk perception, and risky driving behaviors in the current study.

Finally, the K-means clustering and the RF algorithm are analyzed with the help of Python 3.6, while the results of the SEM are obtained by AMOS® 26.0. All statistical significance levels are tested at $p < 0.05$ in this study.

Table 2
The description of endogenous and exogenous variables in the structural model and measurement models.

Latent variables	Observed variables	Description
Exogenous variables		
ξ_1 (DE)		Driving experience
	X_1 (Age)	Drivers' age
	X_2 (Driving Years)	Drivers' cumulative driving years
Endogenous variables		
η_1 (SS)		Sensation seeking
	Y_3 (BS)	Boredom Susceptibility: a dislike of repetition of experience
	Y_4 (DIS)	Disinhibition: the loss of social inhibitions
	Y_5 (ES)	Experience Seeking: whose essence is "experience for its own sake".
	Y_6 (TAS)	Thrill and Adventure Seeking: a desire to engage in outdoor sports or other activities.
η_2 (RP)		Risk perception
	Y_7 (Violation)	the perception of taking deliberate illegal traffic rules of behavior
	Y_8 (Aggressiveness)	the perception of engaging in aggressive driving behavior
	Y_9 (Distraction)	the perception of taking the second driving task in the process of driving
	Y_{10} (Careless)	the perception of careless driving habits
η_3 (RDB)		Risky driving behaviors
	Y_{11} (Error)	the departure of planned actions from some satisfactory path towards a desired goal
	Y_{12} (Violation)	deliberate deviations from those practices believed necessary to maintain the safe operation of a potentially hazardous system
	Y_{13} (Aggressiveness)	aggressive violation-closely related to the aggressive personality factors
	Y_{14} (Lapse)	the unwitting deviation of action from the intention

3. Results

3.1. Factor analysis for the RP, SS, and RDB

According to the dimensionality reduction, these three questionnaires of the DBQ, SSS, and RPQ are divided into several sub-dimensions respectively, which are used as observed variables of potential variables. The SSS is divided into four sub-dimension, namely, the *BS* (boredom susceptibility, a dislike of repetition of experience), *DIS* (disinhibition, the loss of social inhibitions), *ES* (experience-seeking, its essence is "experience for its own sake.") and the *TAS* (thrill and adventure seeking, a desire to engage in outdoor sports or other activities.). The RPQ is divided into four sub-dimensions, including the *Violation* (the perception of taking deliberate illegal traffic rules of behavior), *Aggressiveness* (the perception of engaging in aggressive driving behavior), *Distraction* (the perception of taking the second driving task in the process of driving) and *Carelessness* (the perception of careless driving habits). Since the three-component and four-component structures are commonly used in the DBQ study, they will be verified separately to select the most suitable structure for current samples. Subsequently, questionnaires after the dimensionality reduction are verified by the confirmatory factor analysis (CFA) to test the reliability and validity of the sub-dimension. The CFA is conducted to test whether the RDB, RP, and SS could explain their observed variables effectively, such that to correct observed variables that do not belong to these latent factors.

The performance indexes can be found in Table 3, as we can observe, standardized factor loadings in the SS with four observed variables range from 0.534 (*BS*) to 0.76 (*DIS*), standardized loadings in the RP

Table 3
Confirmatory factor analysis, including composite reliability and convergent validity.

Latent variables	Observed variables	Parameters of significant test				STD. loading	Item reliability	Composite reliability	Convergent validity (AVE)
		Estimate	S.E.	EST/S.E.	p-value				
	BY							.704	.378
	BS	1				.534	.285		
SS	DIS	1.922	.086	22.451	***	.76	.578		
	ES	1.253	.058	21.436	***	.586	.343		
	TAS	1.777	.086	20.749	***	.552	.305		
	BY							.906	.708
	Violation	1				.959	.920		
RP	Aggressiveness	.791	.010	78.339	***	.880	.774		
	Distraction	.887	.016	55.172	***	.743	.552		
	Carelessness	.949	.016	58.248	***	.765	.585		
	BY							.716	.463
RDB (three-component)	Error	1				.811	.658		
	Violation	0.893	.052	17.102	***	.550	.303		
	Lapse	1.127	.063	17.992	***	.654	.428		
	BY							.705	.381
RDB (four-component)	Error	1				.752	.566		
	Violation	1.354	.056	24.350	***	.593	.352		
	Aggressiveness	.690	.034	20.285	***	.458	.210		
	Lapse	1.293	.052	24.976	***	.630	.397		

Note: ***: p-value < 0.001; S.E.: standard error; STD: standardized; EST: estimate; DE: driving experience; SS: sensation seeking; RP: risk perception; RDB: risky driving behaviors; BS: boredom susceptibility: DIS: disinhibition; ES: experience seeking; TAS: thrill and adventure-seeking. The bold factor of the RDB with a four-component structure indicates that it is abandoned in the following study.

range from 0.743 (*Distraction*) to 0.959 (*Violation*), while loadings in the RDB with four-component structure range from 0.458 (*Aggressiveness*) to 0.752 (*Error*). Since the factor loading of *Aggressiveness* in the RDB with a four-component structure is only 0.458, which indicates that its explanatory ability with a four-structure structure is unsatisfactory. Thus, the three-component structure of the RDB is adopted in the following analysis, whose loadings range from 0.550 (*Violation*) to 0.811 (*Error*),

Among them, the minimum factor loading is the *Violation* in the RDB (.550), while the maximum one is the *Violation* in the RP (.959). As most of the factor loadings are larger than 0.6, it indicates that these observed variables can be explained by the latent variables well. However, a few factor loadings range between 0.5 to 0.6, which indicates that the explanatory ability of potential variables to the observed variables is not that perfect but still acceptable.

The composite reliabilities of the three questionnaires are 0.704 (SS), 0.906 (RP), and 0.716 (RDB) respectively, which shows that the consistency of observed variables in the three questionnaires is ideal. The convergent validities of the three questionnaires are 0.378 (SS), 0.708 (RP), and 0.463 (RDB), respectively, which demonstrates that the observed variables could be highly explained by the latent variables in the RPQ, while the interpretation proportion of observed variables in the SS is slightly worse. Moreover, Table 4 reports the discriminant validity among potential variables in this study. A large correlation between potential variables means multiple potential variables jointly explaining one factor, which will lead to the problem of multi-collinearity.

As shown in Table 4, it can be found that the square root of Average

Table 4
Analysis of discriminant validity.

Variables	Convergent validity	Discriminant validity			
	AVE	DE	SS	RP	RDB
DE	.962	**.981**			
SS	.378	−.495	**.615**		
RP	.708	.438	−.372	**.841**	
RDB	.463	−.184	.326	−.251	**.680**

Note: The bold diagonal is the square root of Average variance extraction (AVE), and the lower triangle is the Pearson correlation;
DE: driving experience; SS: sensation seeking; RP: risk perception; RDB: risky driving behaviors.

Variance Extraction (AVE) of each potential variable is greater than the corresponding Pearson correlation, which means the correlation between potential variables is acceptable. Thus, each potential variable can represent different factors and the problem of multicollinearity would be not our concern.

3.2. Results of the SEM

After the preliminary analysis of those questionnaires, the SEM is then derived from the theoretical basis and proposed hypothesis described above. The structural model depicts the relationships among the driving experience, sensation seeking, risk perception, and risky driving behaviors. In this model, there are 13 observed variables and 4 latent variables in total. The standardized path coefficients of the SEM are shown in Fig. 3. To assess the goodness-of-fit of the hypothesized model, the χ^2 which measures the discrepancy between the hypothesized model and the data is used. An insignificant result indicates that the measurement model is consistent with the hypothesized model. However, the χ^2 value is very sensitive to the sample size, which is not a very practical fitness index. When the sample size is more than 200, the p-value of almost all studies is significant (Newsom, 2012), which in our case are, $N = 3150, \chi^2(59) = 1162.364, \chi^2/\text{df} = 19.701, P < 0.001$. Instead, alternative global fitting indexes such as the Goodness of Fit Index (*GFI*), Adjusted Goodness of Fit Index (*AGFI*), Normative Fit Index (*NFI*), Comparative Fit Index (*CFI*), and the Tucker and Lewis Index (*TLI*) are adopted. Contrary to the χ^2 goodness-of-fit, they are not easily affected by the sample size, which varies between 0 (no fit) to 1 (perfect fit). While values larger than 0.9 would generally indicate an adequate fit, which in our case are, $GFI = 0.943, AGFI = 0.913, NFI = 0.947, CFI = 0.950, TLI = 0.933$ respectively.

Also, the Root Mean Square Error of Approximation (*RMSEA*) and the Standardized Root Mean Square (*SRMR*) are included to assess the goodness-of-fit of the proposed model, which both vary between 0 (perfect fit) to 1 (no fit). While the RMSEA and SRMR values are less than 0.08 would generally indicate an adequate fit (Byrne, 2001), which in our case are, $RMSEA = 0.076, SRMR = 0.063$, respectively. Therefore, the goodness-of-fit of the hypothesized model is acceptable.

As shown in Fig. 3 and Table 5, standardized regression weights, and the significance of the structural model can be acknowledged as follows:

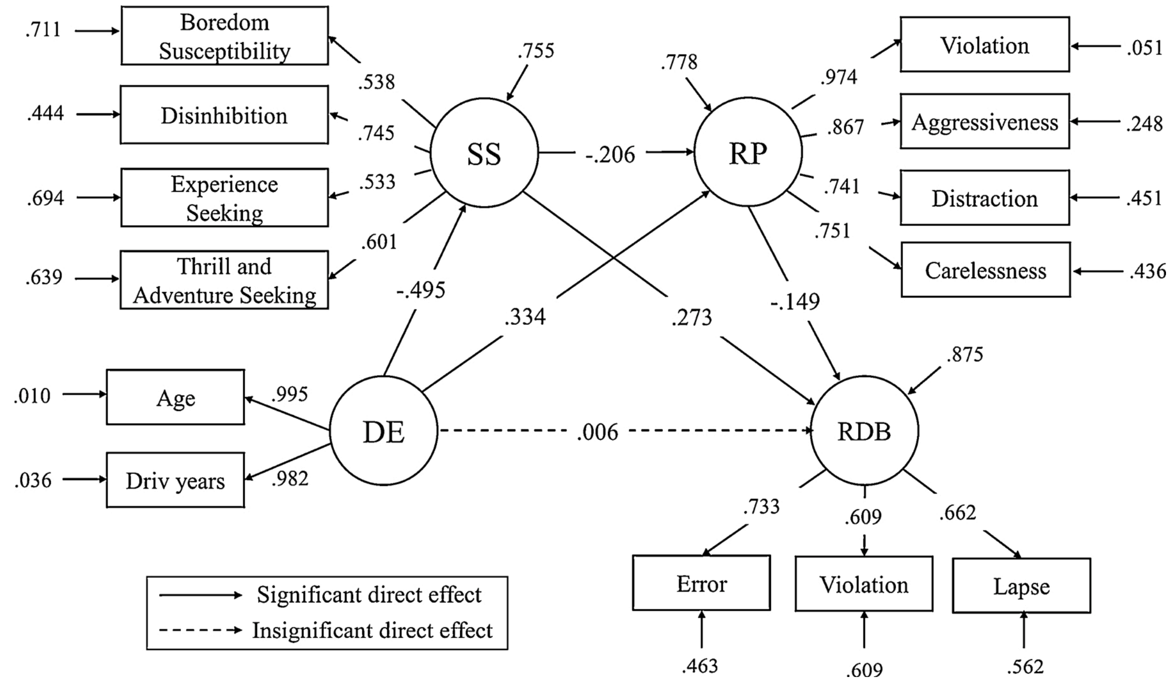

Fig. 3. Standardized path coefficients of the SEM.
Note: the dashed line indicates an insignificant direct effect, while the solid line indicates a significant effect; DE: driving experience; SS: sensation seeking; RP: risk perception; RDB: risky driving behaviors.

Table 5
Regression coefficients of the structural model and the model hypotheses (hypothesis: H1-H2).

Dependent	Independent	Estimate	S.E.	EST./S.E.	p-Value	Hypothesis
SS	DE(H1)	−.495	.015	−32.923	***	Supported
RP	DE(H1)	.334	.018	18.250	***	Supported
	SS	−.206	.023	−8.876	***	
RDB	DE	.006	.024	0.238	0.812	
	SS(H2)	.273	.028	9.645	***	Supported
	RP(H2)	−.149	.026	−5.773	***	Supported

Note: ***: p-value < 0.001; S.E.: standard error, EST: estimate; DE: driving experience; SS: sensation seeking; RP: risk perception; RDB: risky driving behaviors.

1 The standardized factor loadings of driving experience on the SS and the RP are -0.495 ($p < 0.001$), 0.334 ($p < 0.001$), respectively, which indicates that driving experience has a significant positive effect on risk perception and a significant negative effect on sensation seeking. Hypothesis H1 is supported.
2 The standardized factor loading of the SS on the RDB is 0.273 ($p < 0.001$), and that of the RP on the RDB is -0.149 ($p < 0.001$), indicating that sensation seeking positively affects risky behaviors significantly, while risk perception negatively affects risky driving behaviors significantly. Hypothesis H2 is also supported.
3 Moreover, the standardized factor loading of the SS on the RP is − 0.206 ($p < 0.001$), indicating that sensation seeking has a significant negative effect on risk perception.

Besides, the contribution of *age* and *driving years* can be obtained in the results of the SEM, which shows that both observed variables take up an almost equal proportion in the potential variable of driving experience (path regression estimations as 0.995 of driving experience on *age*, and 0.982 of driving experience on *driving years*).

In summary, hypotheses of H1 and H2 are supported, while the direct effect of driving experience on risky driving behaviors is not significant ($p = 0.812$). However, the effect of driving experience on risky driving behaviors might be mediated by sensation seeking and risk perception. Therefore, hypothesis H3 will be analyzed in the following mediation model.

3.3. Mediation model

According to the above analysis of the SEM, the direct effect of driving experience on risky driving behaviors is not significant. However, the effect could be indirect, where driving experience affects risky driving behaviors through sensation seeking and risk perception. Table 6a shows the unstandardized result of the mediation model. The Bootstrap estimation method which resamples the data to obtain more accurate results is further adopted in the study, and estimations through Percentile and Bias Corrected are obtained.

As shown in Table 6a, the direct effect of driving experience on the RDB without mediation is insignificant ($p = 0.839$). However, when sensation seeking and risk perception are included in the model as mediators, the total effect is significant ($p < 0.001$), where the regression weight decreased to -0.009. Moreover, all the indirect effects of driving experience on the RDB is significant while mediated by the SS (the regression weight as − 0.006 ,$p < 0.001$), RP (the regression weight as − 0.002, $p < 0.001$) respectively, or both of the SS and RP (the regression weight as − 0.001, $p < 0.001$), which indicate the case of completely mediated effects. Therefore, driving experience affects risky driving behaviors completely and indirectly through sensation-seeking and risk perception. Hypothesis H3 is proven to be supported.

Besides, given the opposite relation between risky driving behavior and risk perception found in some previous studies, it is speculated that risk perception may be a result of risky driving behavior, rather than an

Table 6a
Mediating effects of driving experience on risky driving behaviors (hypothesis: H3).

Path	Product of Coefficients				Bootstrapping 1000 Times 95 % CI			
					Percentile		Bias Corrected	
	Estimate	S.E.	EST./S.E.	p-Value	Lower 2.5 %	Upper 2.5 %	Lower 2.5 %	Upper 2.5 %
				Indirect effects				
DE→SS→RDB	−.006	.001	−8.248	***	−.005	-.008	−.005	-.008
DE→RP→RDB	−.002	.000	−5.057	***	−.001	−.003	−.001	−.003
DE→SS→RP→RDB	−.001	.000	−4.235	***	.000	−.001	.000	−.001
Total indirect effects	−.009	.001	−10.624	***	−.008	−.011	−.008	−.011
				Direct effect				
DE→RDB	.000	.001	.203	0.839	−.002	.003	−.002	.003
				Total effect				
Total effects	−.009	.003	−7.824	***	−.007	−.011	−.007	−.011

Note: ***: p-value < 0.001; mediating factors: sensation seeking; risk perception; S.E.: standard error; EST: estimate; CI: confidence interval; DE: driving experience; SS: sensation seeking; RP: risk perception; RDB: risky driving behaviors.

influencing factor (Horvath and Zuckerman, 1993). Especially, more frequent engagements in risky driving behaviors may lead to an increase in the cognitive ability to drive risks, and drivers may consider certain driving behaviors could be more dangerous. Therefore, based on the above theory, an alternative mediation model is constructed in the study. While in this model, the roles of risky driving behaviors and risk perception are reversed, that is, risk perception is the terminal dependent variable, while the risky driving behavior is the mediating factor of driving experience on risk perception.

As shown in Table 6b, both of the direct effect of driving experience on the RP without mediation (the regression weight as $0.086, p < 0.001$), and the total effects (the regression weight as $0.114, p < 0.001$) are significant, which are consistent with the original model. However, when the RDB is included in the model as mediators, the indirect effect of driving experience on the RP turns insignificant ($p = 0.153$). The result indicates that the RDB is not a mediating effect of driving experience on the RP. Moreover, the direct regression weight of the RDB on the RP is $-0.742(p < 0.001)$, a negative regression coefficient indicates that frequent risky driving behavior engagements will restrain the ability of risk perception, which is contradictory to the guess of Horvath and Zuckerman (1993). Therefore, the alternative mediation model will be abandoned in the subsequent study.

Moreover, the mediating effect of sensation seeking on risky driving behaviors mediated by risk perception has also been found, which can be observed in Table 7. The direct effect of the SS on the RDB without mediation is significant (the regression weight as $0.059, p < 0.001$). Moreover, when the RP is included as a mediator, the indirect effect of the SS on the RDB is also significant (the regression weight as 0.007, $p < 0.001$). And, the total effects of the SS on the RDB, including the direct and indirect effect, are significant (the regression weight as 0.065, $p < 0.001$). Therefore, the effect of sensation seeking on risky driving behaviors is a process of partial mediation, sensation seeking has a significant direct effect on risky driving behaviors, as well as an indirect effect through risk perception.

3.4. Moderated mediation model

The moderated mediation model of driving experience on risky driving behaviors regarding gender as an interaction factor is further analyzed. The different effects of driving experience on risky driving behaviors between male and female drivers are shown in Table 8 and Fig. 4.

It can be observed that the total effect of the driving experience on the RDB (including direct and indirect effects) are significant both for the male drivers (the regression weight as $-0.005, p < 0.001$) and the female drivers (the regression weight as $-0.012, p < 0.001$), respectively, which suggests that the development trend of risky driving -behavior engagements decrease with the increase of driving experience regardless of gender. And, the total effect for female drivers is significantly larger than male drivers ($p < 0.001$). However, the difference of total indirect effect is insignificant ($p = 0.153$), which indicates the indirect effects of driving experience on risky driving behaviors mediated by sensation seeking and risk perception are insignificantly distinct between different genders.

Moreover, the direct effect of the DE on the RDB is significant for male drivers (the regression weight as $0.005, p < 0.05$) while insignificant for female drivers ($p = 0.970$), and their direct effect difference is marginally significant between different genders ($p = 0.066$). The result shows that, besides the negative indirect impact of driving experience on risky driving behaviors, the direct effect is significant and positive for male drivers. We may infer other undiscussed potential factors have positive effects on risky driving behaviors, which will be further

Table 6b
Alternative mediating effects of driving experience on risk perception.

Path	Product of Coefficients				Bootstrapping 1000 Times 95 % CI			
					Percentile		Bias Corrected	
	Estimate	S.E.	EST./S.E.	p-Value	Lower 2.5 %	Upper 2.5 %	Lower 2.5 %	Upper 2.5 %
				Indirect effects				
DE→SS→RP	.022	.003	6.259	***	.014	.028	.015	.029
DE→RDB→RP	.002	.001	1.428	.153	.000	.004	.000	.004
DE→SS→RDB→RP	.005	.001	4.253	***	.003	.008	.003	.008
Total indirect effects	.028	.003	8.299	***	−.026	.021	.022	.036
				Direct effect				
DE→RP	.086	.005	15.935	***	.076	.097	.075	.096
RDB→RP	−.742	.124	−5.969	***	−1.047	0.474	−1.031	0.466
				Total effect				
Total effects	.114	.005	24.563	***	.105	.124	.105	.124

Note: ***: p-value < 0.001; mediating factors: sensation seeking; risky driving behavior; S.E.: standard error; EST: estimate; CI: confidence interval; DE: driving experience; SS: sensation seeking; RP: risk perception; RDB: risky driving behaviors.

Table 7
Mediating effects of sensation seeking on risky driving behaviors.

Path	Product of Coefficients				Bootstrapping 1000 Times 95 % CI			
					Percentile		Bias Corrected	
	Estimate	S.E.	EST./S.E.	p-Value	Lower 2.5 %	Upper 2.5 %	Lower 2.5 %	Upper 2.5 %
Indirect effect								
SS→RP→RDB	.007	.002	4.249	***	.004	.010	.004	.010
Direct effect								
SS→RDB	.059	.007	8.589	***	.045	.072	.046	.072
Total effect								
Total effects	.065	.007	9.651	***	.052	.078	.052	.079

Note: ***: p-value < 0.001; mediating factors: risk perception; S.E.: standard error; EST: estimate; CI: confidence interval; SS: sensation seeking; RP: risk perception; RDB: risky driving behaviors.

Table 8
Moderated mediation model of driving experience on risky driving behavior (Hypothesis: H4).

Path	Product of Coefficients				Bootstrapping 1000 Times 95 % CI			
					Percentile		Bias Corrected	
	Estimate	S.E.	EST./S.E.	p-Value	Lower 2.5 %	Upper 2.5 %	Lower 2.5 %	Upper 2.5 %
Male group								
Indirect effects								
DE→SS→RDB	−.006	.001	−5.082	***	−.008	-.004	−.008	-.004
DE→RP→RDB	−.003	.001	−3.713	***	−.005	−.001	−.005	−.001
DE→SS→RP→RDB	−.001	.001	−2.639	.013	−.001	.000	−.001	.000
Total indirect effect	−.010	.001	−6.985	***	−.012	−.007	−.012	−.007
Direct effect								
DE→RDB	.005	.002	2.438	.015	.001	.008	.001	.009
Total effect								
Total effect	−.005	.001	−3.361	.001	−.008	−.002	−.008	−.002
Female group								
Indirect effects								
DE→SS→RDB	−.009	.001	−7.695	***	−.011	-.007	−.011	-.007
DE→RP→RDB	−.003	.001	−3.758	***	−.004	−.001	−.004	−.001
DE→SS→RP→RDB	−.001	.000	−3.406	.001	−.001	.000	−.001	.000
Total indirect effect	−.012	.002	−8.020	***	−.015	−.010	−.015	−.010
Direct effect								
DE→RDB	.000	.002	.038	.970	−.003	.004	−.003	.004
Total effect								
Total effect	−.012	.003	−7.727	***	−.015	−.009	−.015	−.009
Difference between genders								
Total effect difference	.007	.002	3.562	***	.003	.011	.003	.011
Direct effect difference	.005	.003	1.840	.066	−.010	.001	−.010	.001
Indirect effect difference	.003	.002	1.429	.153	−.001	.006	−.001	.006

Note: ***: p-value < 0.001; moderating factor: Gender; mediating factors: sensation seeking and risk perception; S.E.: standard error; EST: estimate; CI: confidence interval; DE: driving experience; SS: sensation seeking; RP: risk perception; RDB: risky driving behaviors.

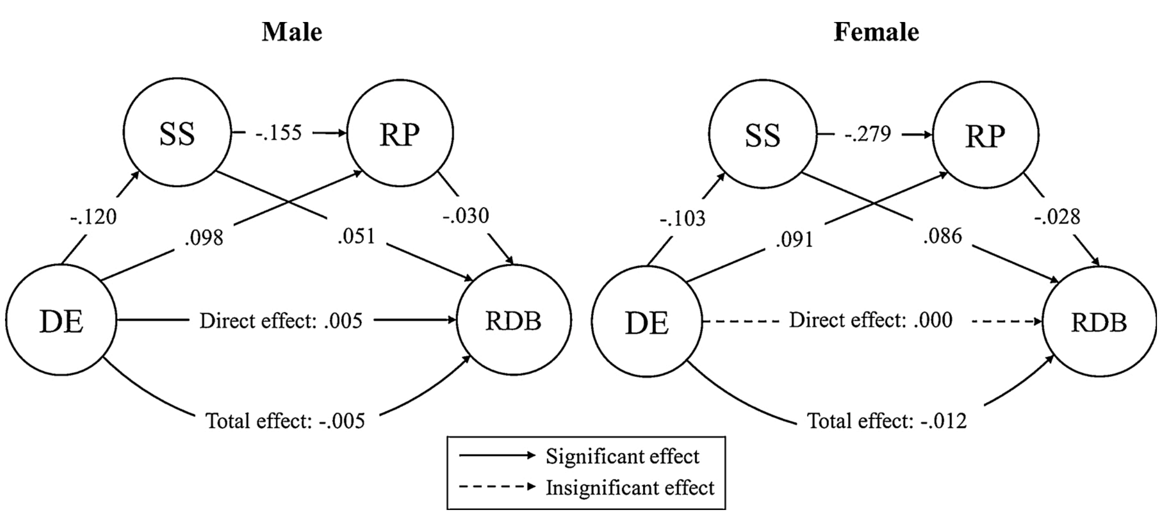

Fig. 4. The direct, indirect, and total effects of driving experience on risky driving behaviors in male and female drivers.
Notes: DE: driving experience; SS: sensation seeking; RP: risk perception; RDB: risky driving behaviors.

addressed in later discussions. On the contrary, the insignificance of the direct effect for female drivers indicates that the negative impact of driving experience on risky behaviors could be completely mediated by sensation seeking and risk perception according to the current model.

In brief, there exists a significant difference in the influence trends of driving experience on risky driving behaviors between different genders, hypothesis H4 is supported.

3.5. Classification of the driver's driving risk

A classification model of the driver risk level is further proposed in this study, which considers the driving experience, sensation seeking, risk perception, and risky driving behavior engagements of the driver. At first, the CNC rates obtained from the SHRP 2 naturalistic driving data are clustered into three risk levels by the K-mean cluster method. The risk levels of drivers are further utilized as the label in the RF classifier. As shown in Fig. 5 and Table 9 of the clustering results, the CNC rate of the low-risk group is below 0.572, the CNC rate of the middle-risk group ranges from 0.583 to 1.813, and that of the high-risk group ranges from 1.834 to 10.333, which indicates that the CNC rate of the higher risk drivers is over 10 times greater than that of the low-risk groups, and the CNC rate of the moderate risk drivers is over 4 times greater than that of the low-risk group.

Eventually, a total of 13 input variables from the driving experience, gender, sensation seeking, risk perception, and risky driving behavior, namely, *Age, Driving years* from driving experience, gender, *Boredom Susceptibility, Disinhibition, Experience Seeking, Thrill and Adventure Seeking* from sensation seeking, *Aggressiveness, Violation, Distraction, Carelessness* from risk perception, *Violation, Error, Lapse* from risky driving behaviors, are included in the RF classifier. Besides, all input variables will be normalized to ensure they are in the same dimension scale before the training process.

Moreover, according to clustering results shown in Table 9 and Fig. 5, the quantity of low-risk group is much larger than the other two groups, taking up to 73.5 % of the overall samples, while the middle-risk and high-risk groups share the rest data with a proportion of 23.5 % and 3.0 %, respectively. It can be noted that there is indeed a difference in the average age and the cumulative driving year between different risk groups, the one way ANOVA test indicates that the mean age of the middle-risk ($p < 0.001$) and high-risk drivers ($p < 0.01$) are significantly lower than that of low-risk drivers, similarly, the mean cumulative driving years of the middle-risk ($p < 0.001$) and high-risk drivers ($p < 0.001$) are also significantly lower than that of low-risk drivers. However, there is no significant difference for different genders among all groups, as well as in age and cumulative driving years between middle-risk drivers and high-risk drivers.

Furthermore, the Synthetic Minority Oversampling Technique

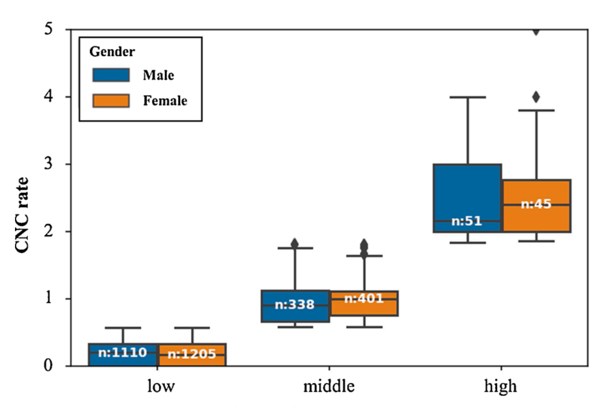

Fig. 5. The boxplot of clustering distribution of the driver risk by using K-mean cluster (clustering variable: the CNC rate).

Table 9
Characteristics of drivers at different risk levels.

risk groups	number of drivers	Mean CNC rate	Male percentage	Mean driving years	Mean age interval
Low	2315	.191	47.9	26.972	40–44
Middle	739	.965	45.7	18.8	35–39
High	96	2.688	53.1	18.1	30–34

(SMOTE) method is adopted for middle-risk and high-risk groups to achieve a reasonable balance among their proportions, as well as the satisfying diversity among the clusters. Thus, the final number of the sample included in the RF classifier is 6945 (2315 samples of low-risk, medium-risk, and high-risk levels respectively). Then, those samples are cross-validated with 10 folds to achieve more than 90.8 % average accuracy to obtain a high accuracy of classification results. The detailed classification performance in the RF classifier is illustrated by a confusion matrix as depicted in Fig. 6, which shows that the classification accuracy of the low-risk group achieves up to 90.8 %, the classification result of the middle-risk level is less satisfying, with only 83.5 percent of samples correctly classified, nevertheless, the classification performance of high-risk drivers is the best, which achieves the accuracy of up to 98.2 %. It can be observed that the classification performance for high-risk drivers is hard to be confused with low and middle-risk drivers. In a word, the overall classification performance of the proposed RF model is quite satisfying.

Fig. 7 further illustrates the estimation results of input variables' importance with the out-of-bag (OOB) estimate method. The horizontal coordinate represents the variable that is randomly disrupted, while the vertical coordinate represents the proportion of the decline in the classification accuracy of the proposed model after disrupting the input variables. The importance of input variables can be presented by comparing the proportion of accuracy decline, which shows that the most important factor is gender, then followed by *TAS* from the SS, *Violation,* and *Error* from the RDB. Both the *age* and *driving years* are important factors for the classification of the driver's driving risk because of the accuracy rate decline while regarding the *age* and *driving years* as 0.086 and 0.072, respectively. Besides, the rest sub-dimensions from the SS, RP, and the RDB also contribute to the model classification performance. Therefore, all factors discussed in the SEM are meaningful for identifying driving risk.

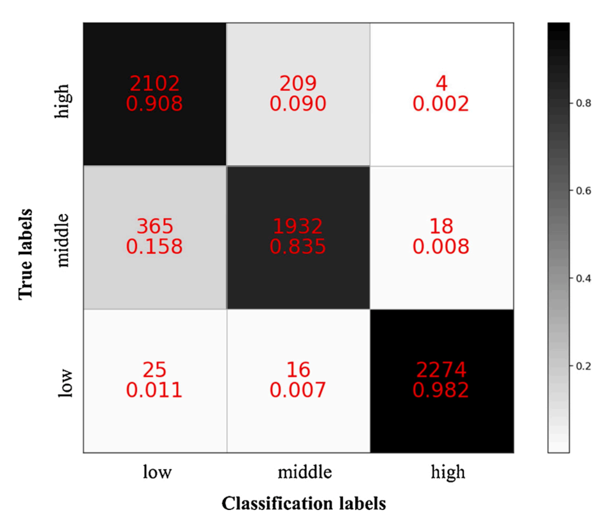

Fig. 6. The confusion matrix of classification results using the Random Forest ($N = 6945$; K-fold = 10).

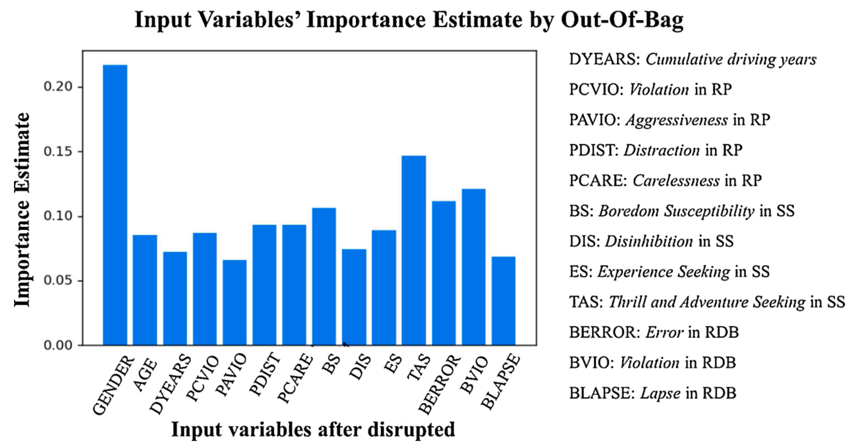

Fig. 7. The estimate of the importance of the input variables by using the OOB.
Note: DE: driving experience; SS: sensation seeking; RP: risk perception; RDB: risky driving behaviors.

4. Discussions

Complex relationships among the driver's demographics, sensation seeking, risk perception, and risky driving behaviors are established and verified in this study. Results show that the driver's driving experience, which is represented by age and cumulative driving years, negatively affects the risky driving behavior engagements. The result indicates that driving experience negatively affects risky driving behavior engagements, which are well consistent with previous age-related analyses (Blockey and Hartley, 1995; De Winter and Dodou, 2010; Rhodes and Pivik, 2011) because age plays an important role in the study while regarding driving experience.

4.1. Influence of mediators

We further analyze the reasons that driving experience inhibits risky driving behaviors through the mediation model. The results of the SEM show that driving experience seems to be a distinct factor in risky driving behaviors. More specifically, there is no significant direct influence of driving experience on risky driving behavior engagements, while there exist significant indirect influences through sensation seeking and risk perception. Among them, driving experience indeed directly affects sensation-seeking and risk perception that are related to risky driving behaviors. The current study indicates that driving experience negatively affects sensation seeking while positively affect risk perception. With the increase of driving experience, the driver has a lower tendency to seek stimulation, which is consistent with the conclusion made by Nordfjærn et al. (2010), in which lower sensation-seeking scores are associated with an incremental in age. On the contrary, an experienced driver has a higher ability to recognize the severity of risky driving behaviors, which is also verified by Delhomme et al. (2009).

Meanwhile, sensation seeking and risk perception are both highly correlated with risky driving behavior engagements. Constantinou et al. (2011) argue that there is a slight positive correlation between thrill-seeking, risk-taking, and traffic violations. Rhodes and Pivik (2011) have demonstrated that risk perception is an important influence factor for drivers in behavioral decision-making. Specifically, the results in this study indicate that the driver's high sensation-seeking will promote to pursue risky driving behaviors, which is also consistent with Jonah et al. (2001), who have found that sensation seeking is positively correlated with the DBQ scores, the stronger the driver's stimulation needs, the higher the frequency of engaging in risky driving behaviors. Conversely, with the improvement of risk perception, drivers have a lower probability to commit risky driving behaviors (Machin and Sankey, 2008; Mills et al., 2008). In summary, certain psychological factors of experienced drivers tend to improve driving safety such as lower sensation seeking and higher risk perception, as a consequence, the likelihood of drivers' risky driving behavior engagements tend to be inhibited.

As the hypothesis mentioned, we have verified that sensation seeking and risk perception are potential mediators in the influence of driving experience on risky driving behaviors in the mediation model. The result is similar to the findings in Rhodes et al., (2011), which shows that the mediating role of risk perception in the influence of age on risky driving behaviors. Moreover, just the same as the results obtained from the SEM, driving experience seems not to have a significant direct effect on risky driving behaviors when ignoring the gender factor. The driving experience affects risky driving behaviors indirectly through sensation seeking and risk perception, which indicates a completely mediating effect. These strong indirect effects might explain the reason why the direct influence of driving experience on driving behaviors in the SEM is weak when adopting sensation seeking and risk perception as mediators.

Moreover, we also verify the relationship between sensation seeking and risk perception. Theoretically, personality traits are thought to influence the individual's perception and appraisal of the environment (Costa and McCrae, 1995), and sensation seeking is generally treated as a kind of personality trait that individuals will manifest in several daily activities, including driving, while risk perception is a more of a subjective evaluation of objective driving behavior (Deery, 1999). The mediation model of sensation seeking on risky driving behaviors as shown in Table 7 indicates that sensation seeking negatively affects risk perception. Meanwhile, lower risk perception will further increase the likelihood of risky driving behavior engagements. Moreover, sensation seeking is a remote influence factor for risky driving behaviors, and risk perception plays a mediating role in the influence of sensation seeking on risky driving behaviors. Our results suggest that sensation seeking significantly affects risky driving behaviors both directly and indirectly, while risk perception mediates this kind of influence, which is also consistent with conclusions obtained by Ulleberg and Rundmo (2003).

4.2. Influence of gender interaction

The effects of driving experience on risky driving behaviors interacted by gender are further discussed in the moderated mediation model. When analyzing the moderating role of gender, the total effects of driving experience on risky driving behavior engagements are negative and significant both in male and female drivers. Moreover, the influence trends of driving experience on risky driving behavior engagements are distinct between genders, in which the development trend of risky driving behavior engagements for female drivers declines greater with the growing of driving experience. The main reason for the difference between genders in the current study is that, due to the mediation of sensation seeking and risk perception, we found that the driving experience of the male drivers has a positive direct influence on risky driving behaviors, while there is no significant direct effect for female drivers. And there exists a marginally significant difference regarding the direct effects of driving experience on risky driving behaviors between genders. The positive direct effect for males partially offsets but dose not completely conceal the negative indirect effects, thus, resulting in a weaker negative total effect than female drivers.

In other words, the effects of driving experience on risky driving behaviors are partially mediated by sensation seeking and risk perception for male drivers, while completely mediated for female drivers in the current model. However, it should be noted that the direct influence of the male drivers, may be caused by other undiscussed mediating factors that contribute to engagements of risky driving behaviors in male groups with the increase of driving experience. Considering that the effects of driving experience on sensation seeking and risk perception are almost equal for both male and female drivers, thus, it is still uncertain what factors account for the difference between genders, we will further address this issue in later discussions.

The reason might be related to the functional health of senior drivers which encompasses physical, cognitive, and psychomotor ability. Aging is accompanied by declines in cognitive, sensory, and motor abilities that can make the task of driving more difficult (Smiley, 2004). And according to the previous research based on the SHRP 2 database, senior drivers' poor functional health will promote the engagement of crash risk (Antin et al., 2017). Generally, aged drivers are aware of functional decline and then avoid travel under threatening conditions (McGwin and Brown, 1999), however, Bauer et al. (2003) have found that male drivers are less likely to reduce or stop driving under some adverse conditions than female drivers. Moreover, self-regulation in driving is generally described as the process of modifying or adjusting one's driving patterns by driving less or intentionally avoiding challenging driving (Molnar et al., 2013). D'Ambrosio et al. (2008) suggest that the degree of self-regulation both for the male and female drivers are increasing with the age, however, male drivers seem to have more confidence in driving, even in the condition of poor functional health, which may cause that the degree of self-regulation is less increased with age when compared with female drivers. Therefore, poor functional health, less self-regulation, or the lower tendency to reduce driving might contribute to more engagements of risky driving behaviors for senior male drivers other than senior female drivers. Considering that the age factor is also an important contributor to the driving experience in our study, thus, it could be reflected in the male drivers' positive direct effect of driving experience on risky driving behaviors as indicated in this study.

Nevertheless, we are fully aware that the gender difference regarding the effect of driving experience on risky driving behaviors could be controversial, for instance, Lajunen et al. (1998) have found that age is associated with a significant decrease in aggressive driving behavior engagements among male drivers, but not for female respondents. Bogdan et al. (2016) also point out that the relationship between driving anger and aggressive driving behavior is distinct between different genders and regions. More specifically, the relationship between driving anger and aggressive driving is stronger for female drivers than male drivers, which is also the strongest for participants from the United States, whereas the weakest from regions of Central Europe. Besides, there seems to exist certain differences among drivers from different regions on the sub-dimensions and scales of the DBQ, and the degree of driver characteristics affecting both the risky driving behavior and the accident rate could vary from regions (Özkan et al., 2006; Bener et al., 2008, 2013). That may lead to different conclusions on the risky driving behavior for the gender difference, therefore, we believe this issue is still open and needs more future studies from the experts of human factors and traffic safety.

4.3. The identification of the high-risk level of drivers

Previous studies have used a logistic regression method to recognize the risk level of drivers based on driver's personality traits, age, and critical-incident events (CIE) rate of the 100-Car Naturalistic Driving Study (Guo and Fang, 2013). While in our study, the input variables for the classification model include driver's demographics, sensation seeking, risk perception, and risky driving behaviors. Meanwhile, the CNC rate is adopted into the classification model to assess driving risk, which shows that individual driver risk varies substantially with three distinct risk-level groups. The cluster analysis indicates that the CNC rate of the higher risk drivers is over 10 times greater than that of the low-risk groups, and the CNC rate of the moderate risk drivers is over 4 times greater than that of the low-risk group, which is consistent with the study of Guo and Fang et al. (2013). The accuracy of the recognition of high-risk drivers achieves up to 98.2 %, which is hard to be confused with low and middle-risk drivers.

The results of the classification model suggest that the driver risk can be assessed by self-reported driver's demographics, sensation seeking, risk perception, and risky driving behaviors, instead of the analysis of the driver's daily driving data. Although high-risk drivers only account for a small proportion of the driver population, they have a substantial impact on overall traffic safety. Identifying factors associated with individual driving risk, detecting unsafe driving behaviors, and recognizing high-risk drivers will enable proper driver-behavior intervention, as well as safety countermeasures to reduce the crash likelihood of high-risk groups and improve overall driving safety (Guo and Fang, 2013). After all, it is unrealistic to monitor all collision accidents in daily driving, screen out high-risk drivers to have them guided and supervised. On the one hand, the collision is generally considered as a small probability event that is hard to be supervised. On the other hand, measures-taking after an accident is usually based on the loss of property and damage to health. If specific characteristics of drivers can be used to identify the high-risk drivers and have them supervised and guided before the occurrence of accidents, it will be beneficial to public traffic safety.

Identification of the driver's driving risk level might be also helpful for the strategy development of collaborative driving or shared driving. So far, few pieces of existing researches on collaborative driving strategies take drivers at different risk levels into account. Nevertheless, different characteristics of drivers might lead to various driving styles, such that personal cooperative driving strategies are required to be particularly designed. To avoid serious potential conflicts between the decision-making system of the intelligent vehicle and the driver himself/herself, it is promising to deepen the research of incorporating driving styles into the shared driving strategies to improve driving safety and driving comfort in the future.

4.4. Limitations

Finally, this study has several limitations reported as follows. First, the results indicate that the main influence of driving experience on risky driving behaviors is mediated by sensation seeking and risk perception, also, with significant distinctions between different genders. However, sensation-seeking is the sole adopted secondary personality trait in this study. Many shreds of evidence indicate one possibility that other undiscussed factors of personality traits are likely to affect risky driving behaviors as well. For instance, aggression is indirectly and positively linked to risky driving behaviors, and impulsivity and sensitivity to reward are related to promoting the occurrence of ordinary violations (Constantinou, 2011). In contrast, the increase in certain personality traits seems to be beneficial to traffic safety. Adolescents scoring high on altruism and trait anxiety are less likely to report risk-taking in traffic (Ulleberg and Rundmo, 2003), and Oltedal and Rundmo (2006) have shown that anxiety is significantly and negatively correlated with risky driving behaviors. These undiscussed personality traits might play a significant role in the influence of driving experience on risky driving behaviors. Besides, risk perception does not seem to be a direct influence factor for risky driving behaviors, previous evidence indicates that the influence of sensation seeking and risk perception on risky driving behaviors might be mediated by risk-taking attitude (Chen, 2009; Ram and Chand, 2016). And, attitude towards traffic safety seems to be a mediator between personality traits and risky driving behaviors, for instance, adolescents scoring high on altruism and trait anxiety are more likely to have a positive attitude towards traffic safety, and attitude towards traffic safety directly affects risky driving behavior. (Ulleberg and Rundmo, 2003). Thus, more personality traits and psychological factors should be considered as mediators in future studies.

Moreover, some studies identify drivers' driving risk based on objective driving data (Wang et al., 2018; Li et al., 2019), while subjective questionnaire results as input variables of the RF classifier are used in this study. It should be acknowledged that some drivers may be biased in their assessments or filled out questionnaires inaccurately. Therefore, we could combine subjective questionnaires with objective daily driving data to make a more accurate assessment of driver risk in the future.

5. Conclusions

This study verifies the impact of driving experience on sensation seeking, risk perception, and risky driving behaviors, meanwhile, analyses the mediating role of sensation seeking and risk perception, as well as the moderating role of gender by using the SEM, which is based on a large-scale questionnaire survey from the SHRP 2. According to the verified hypotheses, the main conclusions of this study are summarized as follows,

1) Drivers' driving experience negatively affects sensation seeking, while positively affect risk perception.
2) Sensation seeking and risk perception affect risky driving behaviors of drivers, lower sensation-seeking and higher risk perception inhibit the likelihood of risky driving behavior engagements.
3) Drivers' driving experience indirectly affects the likelihood of risky driving behavior engagements when regardless of gender, in which sensation seeking and risk perception mediate the negative effects of driving experience on risky driving behaviors.
4) There exists a significant distinction in the influence trends of driving experience on risky driving behaviors between different genders, in which the development trend of risky driving behavior engagements for female drivers declines greater with the growing of driving experience.

In summary, these findings suggest the necessity to embed issues concerning specific aspects of sensation seeking and risk perception into the planning of road safety campaigns and policies, and safety policies should pay more attention to young drivers with less driving experience, higher sensation seeking, and lower risk perception. The traditional strategy of traffic safety campaigns was using authority to tell young drivers to drive safely. But, as an alternative measure to law and authority, let drivers aware of the need for behavioral change on their initiative could be more effective (Ulleberg and Rundmo, 2003). Also, video and simulator training can significantly improve driving safety (Zhao et al., 2019). Such that, developing safety education and publicity through new media to reduce the drivers' pursuit of driving stimulation, and enhance drivers' cognitive ability for driving risks might be more effective. Further, drivers who scored higher in sensation seeking and lower in risk perception should be guided to participate in a simulation training program to improve their driving experience and awareness of driving risk, which might include but not be limited to evaluating driving ability, handling high-risk driving scenarios and training driving skill.

Besides, the driving risk level of drivers evaluated by the CNC rate could be classified through drivers' self-reported questionnaires of demographics, sensation seeking, risk perception, and risky driving behaviors. Identifying potential high-risk drivers will be beneficial to provide more targeted safety education and driving guidance for these drivers, and ultimately, reduce the crash likelihood of high-risk groups and improve overall driving safety.

CRediT authorship contribution statement

Xiaolin Song: Conceptualization, Project administration, Funding acquisition. **Yangang Yin:** Data curation, Formal analysis, Writing - original draft. **Haotian Cao:** Supervision, Methodology, Writing - review & editing, Funding acquisition. **Song Zhao:** Visualization, Investigation. **Mingjun Li:** Software, Validation. **Binlin Yi:** Writing - review & editing.

Declaration of Competing Interest

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

Acknowledgments

This work was supported by the National Natural Science Foundation of China (NSFC) under grant number 51905161 and 51975194.

We sincerely thank constructive comments and valuable suggestions from anonymous reviewers. We would also like to thank the Transportation Research Board of the National Academies of Science of the United States for the public-shared data from the SHRP 2 Insight. The findings and conclusions of this paper are those of authors and do not necessarily represent views of the SHRP 2, the Transportation Research Board, nor the National Academies of the United States.

References

af Wåhlberg, A., Dorn, L., Kline, T., 2011. The Manchester Driver Behaviour Questionnaire as a predictor of road traffic accidents. Theor. Issues Ergon. Sci. 12 (1), 66–86. https://doi.org/10.1080/14639220903023376.
Al-Balbissi, A.H., 2003. Role of gender in road accidents. Accid. Anal. Prev. 4 (1), 64–73. https://doi.org/10.1080/15389580309857.
Antin, J.F., Guo, F., Fang, Y., Dingus, T.A., Hankey, J.M., Perez, M.A., 2017. The influence of functional health on seniors' driving risk. J. Transp. Health 6, 237–244. https://doi.org/10.1016/j.jth.2017.07.003.
Arnett, J.J., 1994. Sensation seeking: a new conceptualization and a new scale. Pers. Individ. Dif. 16, 289–296. https://doi.org/10.1016/0191-8869(94)90165-1.
Bachoo, S., Bhagwanjee, A., Govender, K., 2013. The influence of anger, impulsivity, sensation seeking, and driver attitudes on risky driving behaviour among post-graduate university students in Durban, South Africa. Accid. Anal. Prev. 55, 67–76. https://doi.org/10.1016/j.aap.2013.02.021.
Bagdadi, O., 2013. Assessing safety critical braking events in naturalistic driving studies. Transp. Res. Part F Traffic Psychol. Behav. 16, 117–126. https://doi.org/10.1016/j.trf.2012.08.006.
Bauer, M.J., Adler, G., Kuskowski, M.A., Rottunda, S., 2003. The influence of age and gender on the driving patterns of older adults. J. Women Aging 15 (4), 3–16. https://doi.org/10.1300/J074v15n04_02.
Beck, L.F., Dellinger, A.M., O'Neil, M.E., 2007. Motor vehicle crash injury rates by mode of travel, United States: using exposure-based methods to quantify differences. Am. J. Epidemiol. 166 (2), 212–218. https://doi.org/10.1093/aje/kwm064.
Bener, A., Özkan, T., Lajunen, T., 2008. The driver behaviour questionnaire in Arab gulf countries: qatar and United Arab Emirates. Accid. Anal. Prev. 40 (4), 1411–1417. https://doi.org/10.1016/j.aap.2008.03.003.
Bener, A., Verjee, M., Dafeeah, E.E., Yousafzai, M.T., Mari, S., Hassib, A., et al., 2013. A cross "Ethnical" comparison of the driver behaviour questionnaire (DBQ) in an economically fast developing country: DBQ in arab gulf and south asian countries. Glob. J. Health Sci. 5 (4), 165. https://doi.org/10.5539/gjhs.v5n4p165.
Bentler, P.M., Chou, C.P., 1987. Practical issues in structural modeling. Sociol. Methods Res. 16 (1), 78–117. https://doi.org/10.1177/0049124187016001004.
Bingham, C.R., Ehsani, J.P., 2012. The relative odds of involvement in seven crash configurations by driver age and sex. J. Adolesc. Health 51, 484e90. https://doi.org/10.1016/j.jadohealth.2012.02.012.
Blockey, P.N., Hartley, L.R., 1995. Aberrant driving behaviour: errors and violations. Ergonomics 38, 1759–1771. https://doi.org/10.1080/00140139508925225.
Bogdan, S.R., Măirean, C., Havârneanu, C.E., 2016. A meta-analysis of the association between anger and aggressive driving. Transp. Res. Part F Traffic Psychol. Behav. 42, 350–364. https://doi.org/10.1016/j.trf.2016.05.009.

Breiman, L., 1996. Bagging predictors. Mach. Learn. 24 (2), 123–140. https://doi.org/10.1007/BF00058655.
Burns, P.C., Wilde, G.J., 1995. Risk taking in male taxi drivers: relationships among personality, observational data, and driver records. Pers. Individ. Dif. 18 (2), 267–278. https://doi.org/10.1016/0191-8869(94)00150-Q.
Byrne, B.M., 2001. Structural equation modeling: perspectives on the present and the future. Int. J. Test. 1 (3–4), 327–334. https://doi.org/10.1080/15305058.2001.9669479.
Cestac, J., Paran, F., Delhomme, P., 2011. Young drivers' sensation seeking, subjective norms, and perceived behavioral control and their roles in predicting speeding intention: how risk-taking motivations evolve with gender and driving experience. Saf. Sci. 49 (3), 424–432. https://doi.org/10.1016/j.ssci.2010.10.007.
Chen, C.F., 2009. Personality, safety attitudes, and risky driving behaviors–evidence from young Taiwanese motorcyclists. Accid. Anal. Prev. 41 (5), 963–968. https://doi.org/10.1016/j.aap.2009.05.013.
Cohn, L.D., Macfarlane, S., Yanez, C., Imai, W.K., 1995. Risk-perception: differences between adolescents and adults. Health Psychol. 14 (3), 217. https://doi.org/10.1037/0278-6133.14.3.217.
Constantinou, E., Panayiotou, G., Konstantinou, N., Loutsiou-Ladd, A., Kapardis, A., 2011. Risky and aggressive driving in young adults: Personality matters. Accid. Anal. Prev. 43 (4), 1323–1331. https://doi.org/10.1016/j.aap.2011.02.002.
Cordazzo, S.T., Scialfa, C.T., Bubric, K., Ross, R.J., 2014. The driver behaviour questionnaire: a North American analysis. J. Safety Res. 50, 99–107. https://doi.org/10.1016/j.jsr.2014.05.002.
Costa, P.T., McCrae, R.R., 1985. The NEO Personality Inventory. Psychological Assessment Resources, Odessa, FL. https://doi.org/10.1177/106907279500300202.
Costa Jr., P.T., McCrae, R.R., 1995. Domains and facets: hierarchical personality assessment using the revised NEO personality inventory. J. Pers. Assess. 64 (1), 21–50. https://doi.org/10.1207/s15327752jpa6401_2.
D'Ambrosio, L.A., Donorfio, L.K., Coughlin, J.F., Mohyde, M., Meyer, J., 2008. Gender differences in self-regulation patterns and attitudes toward driving among older adults. J. Women Aging 20 (3–4), 265–282. https://doi.org/10.1080/08952840801984758.
de Oña, J., Mujalli, R.O., Calvo, F.J., 2011. Analysis of traffic accident injury severity on Spanish rural highways using Bayesian networks. Accid. Anal. Prev. 43 (1), 402–411. https://doi.org/10.1016/j.aap.2010.09.010.
De Winter, J.C.F., Dodou, D., 2010. The Driver Behaviour Questionnaire as a predictor of accidents: a meta-analysis. J. Safety Res. 41 (6), 463–470. https://doi.org/10.1016/j.jsr.2010.10.007.
De Winter, J.C.F., Dreger, F.A., Huang, W., Miller, A., Soccolich, S., Machiani, S.G., Engström, J., 2018. The relationship between the Driver Behavior Questionnaire, Sensation seeking Scale, and recorded crashes: a brief comment on Martinussen et al. (2017) and new data from SHRP2. Accid. Anal. Prev. 118, 54–56. https://doi.org/10.1016/j.aap.2018.05.016.
Deery, H.A., 1999. Hazard and risk perception among young novice drivers. J. Safety Res. 30 (4), 225–236. https://doi.org/10.1016/S0022-4375(99)00018-3.
Delhomme, P., Verlhiac, J., Martha, C., 2009. Are drivers' comparative risk judgments about speeding realistic? J. Safety Res. 40 (5), 333–339. https://doi.org/10.1016/j.jsr.2009.09.003.
Dingus, T.A., Klauer, S.G., Neale, V.L., Petersen, A., Lee, S.E., Sudweeks, J., et al., 2006. The 100-car Naturalistic Driving Study, Phase II-results of the 100-car Field Experiment (No. DOT-HS-810-593). Department of Transportation. National Highway Traffic Safety Administration, The United States.
Dingus, T.A., Guo, F., Lee, S., Antin, J.F., Perez, M., Buchanan-King, M., Hankey, J., 2016. Driver crash risk factors and prevalence evaluation using naturalistic driving data. Proc. Natl. Acad. Sci. 113 (10), 2636–2641. https://doi.org/10.1073/pnas.1513271113.
Doran, N., Sanders, P.E., Bekman, N.M., Worley, M.J., Monreal, T.K., McGee, E., Brown, S.A., 2011. Mediating influences of negative affect and risk perception on the relationship between sensation seeking and adolescent cigarette smoking. Nicotine Tob. Res. 13 (6), 457–465. https://doi.org/10.1093/ntr/ntr025.
Du, X., Shen, Y., Chang, R., Ma, J., 2018. The exceptionists of Chinese roads: the effect of road situations and ethical positions on driver aggression. Transp. Res. Part F Traffic Psychol. Behav. 58, 719–729. https://doi.org/10.1016/j.trf.2018.07.008.
Eagly, A.H., Chaiken, S., 1993. The Psychology of Attitudes. Harcourt Brace, Jovanovich Fort Worth, TX.
Evans, L., Wasielewski, P., 1982. Do accident-involved drivers exhibit riskier everyday driving behavior? Accid. Anal. Prev. 14 (1), 57–64. https://doi.org/10.1016/0001-4575(82)90007-0.
Furnham, A., Saipe, J., 1993. Personality correlates of convicted drivers. Pers. Individ. Dif. 14 (2), 329–336. https://doi.org/10.1016/0191-8869(93)90131-L.
Golob, T.F., 2003. Structural equation modeling for travel behavior research. Transp. Res. Part B Methodol. 37 (1), 1–25. https://doi.org/10.1016/S0191-2615(01)00046-7.
Guo, F., Fang, Y., 2013. Individual driver risk assessment using naturalistic driving data. Accid. Anal. Prev. 61, 3–9. https://doi.org/10.1016/j.aap.2012.06.014.
Haleem, K., Gan, A., 2015. Contributing factors of crash injury severity at public highway-railroad grade crossings in the US. J. Safety Res. 53, 23–29. https://doi.org/10.1016/j.jsr.2015.03.005.
Hamdar, S.H., Mahmassani, H.S., Chen, R.B., 2008. Aggressiveness propensity index for driving behavior at signalized intersections. Accid. Anal. Prev. 40 (1), 315–326. https://doi.org/10.1016/j.aap.2007.06.013.
Hankey, J.M., Perez, M.A., McClafferty, J.A., 2016. Description of the SHRP 2 Naturalistic Database and the Crash, Near-crash, and Baseline Data Sets. Virginia Tech Transportation Institute.
Harbeck, E.L., Glendon, A.I., 2018. Driver prototypes and behavioral willingness: young driver risk perception and reported engagement in risky driving. J. Safety Res. 66, 195–204. https://doi.org/10.1016/j.jsr.2018.07.009.
Hatakka, M., Keskinen, E., Katila, A., Laapotti, S., 1997. Self-reported Driving Habits Are Valid Predictors of Violations and Accidents. Traffic and Transport Psychology. Theory and Application. http://worldcat.org/isbn/0080427863.
Horvath, P., Zuckerman, M., 1993. Sensation seeking, risk appraisal, and risky behavior. Pers. Individ. Dif. 14 (1), 41–52. https://doi.org/10.1016/0191-8869(93)90173-Z.
Jonah, B.A., 1986. Accident risk and risk-taking behaviour among young drivers. Accid. Anal. Prev. 18, 255–271. https://doi.org/10.1016/0001-4575(86)90041-2.
Jonah, B.A., 1997. Sensation seeking and risky driving: a review and synthesis of the literature. Accid. Anal. Prev. 29 (5), 651–665. https://doi.org/10.1016/S0001-4575(97)00017-1.
Jonah, B.A., Thiessen, R., Au-Yeung, E., 2001. Sensation seeking, risky driving, and behavioral adaptation. Accid. Anal. Prev. 33 (5), 679–684. https://doi.org/10.1016/S0001-4575(00)00085-3.
Juhnke, G.A., Sullivan, T.J., Harman, A.E., 1995. Attitude changes in DWI offenders: a study of a short-term treatment program. J. Addict. Offender Couns. https://doi.org/10.1002/j.2161-1874.1995.tb00098.x.
Kline, R.B., 2015. Principles and Practice of Structural Equation Modeling. Guilford publications.
Lajunen, T., Parker, D., Stradling, S.G., 1998. Dimensions of driver anger, aggressive and highway code violations and their mediation by safety orientation in UK drivers. Transp. Res. Part F Traffic Psychol. Behav. 1 (2), 107–121. https://doi.org/10.1016/S1369-8478(98)00009-6.
Lajunen, T., Parker, D., Summala, H., 2004. The Manchester driver behaviour questionnaire: a cross-cultural study. Accid. Anal. Prev. 36 (2), 231–238. https://doi.org/10.1016/S0001-4575(02)00152-5.
Lawton, R., Parker, D., Manstead, A.S., Stradling, S.G., 1997. The role of affect in predicting social behaviors: the case of road traffic violations. J. Appl. Soc. Psychol. 27 (14), 1258–1276. https://doi.org/10.1111/j.1559-1816.1997.tb01805.x.
Lee, J.Y., Chung, J.H., Son, B., 2008. Analysis of traffic accident size for Korean highway using structural equation models. Accid. Anal. Prev. 40 (6), 1955–1963. https://doi.org/10.1016/j.aap.2008.08.006.
Li, G., Zhu, F., Qu, X., Cheng, B., Li, S., Green, P., 2019. Driving style classification based on driving operational pictures. IEEE Access 7, 90180–90189. https://doi.org/10.1109/ACCESS.2019.2926494.
Lu, J., Xie, X., Zhang, R., 2013. Focusing on appraisals: how and why anger and fear influence driving risk perception. J. Safety Res. 45, 65–73. https://doi.org/10.1016/j.jsr.2013.01.009.
Ma, M., Yan, X., Huang, H., Abdel-Aty, M., 2010. Safety of public transportation occupational drivers: risk perception, attitudes, and driving behavior. Transp. Res. Rec. 2145 (1), 72–79. https://doi.org/10.3141/2145-09.
Machin, M.A., Sankey, K.S., 2008. Relationships between young drivers' personality characteristics, risk perceptions, and driving behaviour. Accid. Anal. Prev. 40 (2), 541–547. https://doi.org/10.1016/j.aap.2007.08.010.
MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability (Vol. 1, No. 14, Pp. 281-297). June.
McGwin Jr., G., Brown, D.B., 1999. Characteristics of traffic crashes among young, middle-aged, and older drivers. Accid. Anal. Prev. 31 (3), 181–198. https://doi.org/10.1016/S0001-4575(98)00061-X.
Mills, B., Reyna, V.F., Estrada, S., 2008. Explaining contradictory relations between risk perception and risk taking. Psychol. Sci. 19 (5), 429–433. https://doi.org/10.1111/j.1467-9280.2008.02104.x.
Molnar, L.J., Eby, D.W., Charlton, J.L., Langford, J., Koppel, S., Marshall, S., Man-Son-Hing, M., 2013. Reprint of "Driving avoidance by older adults: is it always self-regulation?". Accid. Anal. Prev. 61, 272–280. https://doi.org/10.1016/j.aap.2013.07.004.
Newsom, J.T., 2012. Some clarifications and recommendations on fit indices. USP 655, 123–133.
Nordfjærn, T., Jørgensen, S.H., Rundmo, T., 2010. An investigation of driver attitudes and behaviour in rural and urban areas in Norway. Saf. Sci. 48 (3), 348–356. https://doi.org/10.1016/j.ssci.2009.12.001.
Norris, F.H., Matthews, B.A., Riad, J.K., 2000. Characterological, situational, and behavioral risk factors for motor vehicle accidents: a prospective examination. Accid. Anal. Prev. 32, 505–515. https://doi.org/10.1016/S0001-4575(99)00068-8.
Olsen, E.C., Lee, S.E., Simons-Morton, B.G., 2007. Eye movement patterns for novice teen drivers: does 6 months of driving experience make a difference? Transp. Res. Rec. 2009 (1), 8–14. https://doi.org/10.3141/2009-02.
Oltedal, S., Rundmo, T., 2006. The effects of personality and gender on risky driving behaviour and accident involvement. Saf. Sci. 44 (7), 621–628. https://doi.org/10.1016/j.ssci.2005.12.003.
Özkan, T., Lajunen, T., Chliaoutakis, J.E., Parker, D., Summala, H., 2006. Cross-cultural differences in driving behaviours: a comparison of six countries. Transp. Res. Part F Traffic Psychol. Behav. 9 (3), 227–242. https://doi.org/10.1016/j.trf.2006.01.002.
Panayiotou, G., 2015. The bold and the fearless among us: elevated psychopathic traits and levels of anxiety and fear are associated with specific aberrant driving behaviors. Accid. Anal. Prev. 79, 117–125. https://doi.org/10.1016/j.aap.2015.03.007.
Parker, D., Reason, J.T., Manstead, A.S., Stradling, S.G., 1995. Driving errors, driving violations, and accident involvement. Ergonomics 38 (5), 1036–1048. https://doi.org/10.1080/00140139508925170.
Piedmont, R.L., 1998. The Revised Neo Personality Inventory: Clinical and Research Applications. Kluwer, New York.

Ram, T., Chand, K., 2016. Effect of drivers' risk perception and perception of driving tasks on road safety attitude. Transp. Res. Part F Traffic Psychol. Behav. 42, 162–176. https://doi.org/10.1016/j.trf.2016.07.012.
Reason, J., Manstead, A., Stradling, S., Baxter, J., Campbell, K., 1990. Errors and violations on the roads: a real distinction? Ergonomics 33 (10-11), 1315–1332. https://doi.org/10.1080/00140139008925335.
Rhodes, N., Pivik, K., 2011. Age and gender differences in risky driving: the roles of positive affect and risk perception. Accid. Anal. Prev. 43 (3), 923–931. https://doi.org/10.1016/j.aap.2010.11.015.
Rosenbloom, T., Shahar, A., Elharar, A., Danino, O., 2008. Risk perception of driving as a function of advanced training aimed at recognizing and handling risks in demanding driving situations. Accid. Anal. Prev. 40 (2), 697–703. https://doi.org/10.1016/j.aap.2007.09.007.
Rothbart, M.K., Ahadi, S.A., Hershey, K.L., Fisher, P., 2001. Investigations of temperament at three to seven years: the Children's Behavior Questionnaire. Child Dev. 72, 1394–1408. https://doi.org/10.1111/1467-8624.00355.
Rundmo, T., Iversen, H., 2004. Risk perception and driving behaviour among adolescents in two Norwegian counties before and after a traffic safety campaign. Saf. Sci. 42 (1), 1–21. https://doi.org/10.1016/S0925-7535(02)00047-4.
Santamariña-Rubio, E., Pérez, K., Olabarria, M., Novoa, A.M., 2014. Gender differences in road traffic injury rate using time travelled as a measure of exposure. Accid. Anal. Prev. 65, 1–7. https://doi.org/10.1016/j.aap.2013.11.015.
Sarkar, S., Andreas, M., 2004. Acceptance of and engagement in risky driving behaviors by teenagers. Adolescence 39 (156), 687.
Singh, S., 2015. Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey (No. DOT HS 812 115).
Smiley, A., 2004. Adaptive strategies of older drivers. Transportation Research Board Conference Proceedings (No. 27).
Ulleberg, P., Rundmo, T., 2003. Personality, attitudes, and risk perception as predictors of risky driving behaviour among young drivers. Saf. Sci. 41 (5), 427–443. https://doi.org/10.1016/S0925-7535(01)00077-7.
Wang, X., Xu, X., 2019. Assessing the relationship between self-reported driving behaviors and driver risk using a naturalistic driving study. Accid. Anal. Prev. 128, 8–16. https://doi.org/10.1016/j.aap.2019.03.009.
Wang, W., Xi, J., Zhao, D., 2018. Driving style analysis using primitive driving patterns with Bayesian nonparametric approaches. IEEE Trans. Intell. Transp. Syst. 20 (8), 2986–2998. https://doi.org/10.1109/TITS.2018.2870525.
Williams, A.F., 2003. Teenage drivers: patterns of risk. J. Safety Res. 34 (1), 5–15. https://doi.org/10.1016/S0022-4375(02)00075-0.
World Health Organization, 2018. Global Status Report on Road Safety 2018. World Health Organization. Available at:. WHO, Geneva https://www.who.int/violence_injury_prevention/road_safety_status/2018/en/.
Wu, K.F., Aguero-Valverde, J., Jovanis, P.P., 2014. Using naturalistic driving data to explore the association between traffic safety-related events and crash risk at driver level. Accid. Anal. Prev. 72, 210–218. https://doi.org/10.1016/j.aap.2014.07.005.
Zhang, L., Zhang, C., Shang, L., 2016. Sensation-seeking and domain-specific risk-taking behavior among adolescents: risk perceptions and expected benefits as mediators. Pers. Individ. Dif. 101, 299–305. https://doi.org/10.1016/j.paid.2016.06.002.
Zhao, N., Mehler, B., Reimer, B., D'Ambrosio, L.A., Mehler, A., Coughlin, J.F., 2012. An investigation of the relationship between the driving behavior questionnaire and objective measures of highway driving behavior. Transp. Res. Part F Traffic Psychol. Behav. 15 (6), 676–685. https://doi.org/10.1016/j.trf.2012.08.001.
Zhao, X., Xu, W., Ma, J., Li, H., Chen, Y., 2019. An analysis of the relationship between driver characteristics and driving safety using structural equation models. Transp. Res. Part F Traffic Psychol. Behav. 62, 529–545. https://doi.org/10.1016/j.trf.2019.02.004.
Zheng, L., Ismail, K., Meng, X., 2014. Freeway safety estimation using extreme value theory approaches: a comparative study. Accid. Anal. Prev. 62, 32–41. https://doi.org/10.1016/j.aap.2013.09.006.
Zuckerman, M., 1971. Dimensions of sensation seeking. J. Consult. Clin. Psychol. 36 (1), 45. https://doi.org/10.1037/h0030478.
Zuckerman, M., 1984. Experience and desire: a new format for sensation seeking scales. J. Behav. Assess. 6 (2), 101–114. https://doi.org/10.1007/BF01350166.
Zuckerman, M., 1994. Behavioral Expressions and Biosocial Bases of Sensation Seeking. Cambridge University Press.