Risk Awareness and Perception Training using VR Headsets: The Validation of VR Headsets to Measure Hazard Anticipation Behaviors

SAFETY RESEARCH USING SIMULATION

UNIVERSITY TRANSPORTATION CENTER

Cole Fitzpatrick, Ph.D.
Research Assistant Professor
Department of Civil & Environmental Engineering
University of Massachusetts, Amherst

Ganesh Pai Mangalore
Graduate Research Assistant
Department of Mechanical & Industrial Engineering
University of Massachusetts, Amherst

Risk Awareness and Perception Training Using VR Headsets:
The Validation of VR Headsets to Measure Hazard Anticipation Behaviors

Cole Fitzpatrick, Ph.D., PI

Research Assistant Professor

Department of Civil and Environmental Engineering

University of Massachusetts, Amherst

https://orcid.org/0000-0002-9873-1391

Michael Knodler Jr., Ph.D.

Professor

Department of Civil and Environmental Engineering

University of Massachusetts, Amherst

https://orcid.org/0000-0002-6517-4066

Ganesh Pai Mangalore

Graduate Research Assistant

Department of Mechanical and Industrial Engineering

University of Massachusetts, Amherst

https://orcid.org/0000-0002-5151-9860

Yalda Ebadi

Graduate Research Assistant

Department of Mechanical and Industrial Engineering

University of Massachusetts, Amherst

https://orcid.org/0000-0002-0871-5225

Siby Samuel, Ph.D.

Assistant Professor

Systems Design Engineering

University of Waterloo, Canada

https://orcid.org/0000-0002-2168-8479

Donald Fisher, Ph.D.

Emeritus Professor

Department of Mechanical and Industrial Engineering

University of Massachusetts, Amherst

https://orcid.org/0000-0001-5994-3079

A Report on Research Sponsored by

SAFER-SIM University Transportation Center

Federal Grant No: 69A3551747131

March 2019

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

Table of Contents

List of Figures

List of Tables

Abstract

The objective of the current study is to evaluate the use of virtual reality (VR) headsets to measure driving performance for driving simulation studies and training. This is desirable because they are several orders of magnitude less expensive and, if validated, could greatly extend the powers of simulation. Out of several possible measures of performance that could be considered for evaluating VR headsets, the current study specifically examines drivers' latent hazard anticipation behavior both because it has been linked to crashes and because it has been shown to be significantly poorer in young drivers compared to their experienced counterparts in traditional driving simulators and in open road studies. The total time middle-aged drivers spend glancing at a latent hazard and the average duration of each glance were also compared to these same times for younger drivers using a VR headset and fixed-base driving simulator. In a between-subjects design, 48 participants were equally and randomly assigned to one of four experimental conditions: two young driver cohorts (18-21 years) and two middle-aged driver cohorts (30-55 years) navigating either a fixed-base driving simulator or a VR-headset-based simulator. All participants navigated six unique scenarios while their eyes were continually tracked. The proportion of latent hazards anticipated by participants, which constituted the primary dependent measure, was found to be greater for middle-aged drivers than for young drivers across both platforms. Results also indicate that the middle-aged participants glanced longer than their younger counterparts on both platforms at latent hazards, as measured by the total glance duration; however, there was no difference when measured by the average glance duration. Moreover, the magnitude of the difference between middle-aged and younger drivers was the same across the two platforms. The study provides some justification for the use of VR headsets as a way of understanding drivers' hazard anticipation behavior and using this understanding to create a risk awareness training program.

1 Introduction

1.1 Overview of Driving Simulators and Virtual Reality Headsets

Over the years, driving simulators have been extensively used for various transportation, human factors, and behavioral studies [1]. Their increased level of safety and their ability to simulate real-world scenarios with a high sense of immersion have made them useful tools for studying drivers' behavior and performance in low- and high-risk scenarios, for evaluating alternative in-vehicle interface designs, and for conceptualizing and designing training programs [2, 3, 4]. The realism in the simulation of these virtual environments is particularly useful as the simulator tests can be used as a precursor to open road evaluations, thereby minimizing research expenditures and increasing the level of safety (5, 6, 7).

In the past few years, the market has been saturated with a wide variety of virtual reality (VR) headsets such as Oculus Rift, Nintendo Wii U VR, and HTC Vive, among others, which have been used for research, training and educational purposes [8, 9, 10]. The ambiguity of a 3D environment is eliminated in VR headsets, and the true experience of that 3D environment that cannot be achieved in a non-VR-headset platform is possible [11, 12]. Virtual reality headsets allow the user to experience virtual worlds with higher-resolution graphical quality and a more regulated visual flow with a higher sense of realism than is possible in environments presented on conventional driving simulators. Additionally, VR headsets have greater flexibility and portability.

With VR headsets, users can use the engaging, immersive virtual worlds to learn rich and complex content while enhancing their technical, creative, and problem-solving skills [13]. By executing optimized, intelligent designs with systematic delivery, a user can grasp more complex concepts [14]. This greatly enhances training programs aimed at drivers and pilots by not only making it possible to measure participants' behavioral responses more effectively, but by also making considerable cost reductions on infrastructure, equipment, and their accompanying technical support [15, 7].

Virtual reality headsets can also be used to review certain expensive designs and concepts more effectively; for example, by combining 3D models with VR headsets, an architect or contractor can walk through a simulated virtual space of a structural design before the expensive real-life construction of that structure begins [16]. However, VR headsets do have their own disadvantages. They are known to cause a phenomenon called the *screen door effect*, which can be described as a black grid over the original image while displaying a virtual world. The Oculus Rift headset, when worn close to the eyes of the user, has been known to cause a screen door effect. It is unclear whether the VR headset used in this research (the HTC Vive) causes this phenomenon. Ghosting is another phenomenon, where faded trails appear behind moving objects. This has again been detected during the use of the Oculus Rift [17]. Prolonged use of VR headsets could also cause physical discomfort, which may affect the user's experience of the virtual environment. This may lead to the user developing a negative attitude towards VR use in general. It should not be assumed that physical ergonomics are simply due to poor design of VR peripherals because VR peripherals are developing fast. However, it is worth noting that sophisticated models may not be cost effective [18]. In a study utilizing the Oculus Rift VR headset, simulator sickness was a strong factor in modulating people's gaming experiences using the Rift, though it was found that simulator sickness did not always significantly diminish the participants' immersive experiences. With that in mind, it is pivotal to consider the effect of simulator sickness during the development of virtual worlds [19].

In the driving safety research domain, such headsets have been used to train hazard anticipation behavior in young drivers [20]. It was found that when VR headsets were used in the mediation stage of the Risk Awareness and Perception Training (RAPT) program, novice drivers were able detect a high proportion of latent hazards in the mastery stage when compared to traditional mediation stages that involved PowerPoint slides or other PC-based applications. The aim of the current study is to perform an initial validation of VR headsets as a platform for driving simulation since they offer better immersion [21], additional portability, and much lower costs while maintaining the level of safety provided by traditional simulators. This would make it possible to design a RAPT program on a standalone VR-headset-based driving simulator. Additionally, they could greatly extend the use of simulators in science and engineering, possibly making the study of hundreds of drivers in mixed-traffic environments a real possibility. However, at least two concerns stand in the way. First, there is a lack of documented research that specifically examines the ability of these headsets to measure driving performance and to do so as well as traditional fixed-base driving simulators. Second, there is a concern that VR headsets can lead to simulator sickness [22, 23].

1.2 Hazard Anticipation

With respect to the first concern about VR headsets, several aspects of driver performance could have been examined for such a validation study. In this experiment, we specifically focused on a higher-order cognitive skill—latent hazard anticipation—since this aspect of driver performance is a critical skill to master in the RAPT program. In the literature, hazard anticipation is described as a collection of driver behavioral attributes such as the awareness and knowledge of traffic risks, the ability to scan and understand hazardous situations that may result in crashes, the ability to anticipate latent hazards from the current field of view, and the capacity to adopt the necessary actions to safely navigate the roadway by mitigating risks [24, 25]. Researchers have learned that it is important to differentiate between hazards that are visible and those that are not visible or have not materialized but can easily be anticipated [26]. This is perhaps best understood using examples. An example of a visible hazard is a vehicle in the opposing lane crossing into the driver's lane, and an example of a hazard that is not visible but can be anticipated is a pedestrian in a crosswalk hidden by a stopped vehicle in a travel lane. For an example of a hazard that has not materialized, consider a vehicle driving through a residential area on a two-lane roadway with a hidden driveway on the right side. The driveway is obscured by vegetation, and any potential hazard coming onto the road from the driveway is also obscured. To minimize any potential conflicts, the driver would need to identify and continuously scan the driveway for any potential hazards that may emerge until safely passing through that area of the roadway [27].

There are two reasons we focus on hazard anticipation. On the one hand, there is a consensus that young, novice drivers lack the ability to acquire and assess information relevant to the recognition of risks on the road ahead [28, 29, 30, 31]. A driving simulator study by Pradhan et al. [32] reported that while 69.59% of older, experienced drivers engaged in behaviors indicative of successful latent hazard detection in the scenarios, only 25.82% of younger, inexperienced drivers and 40.14% of younger, experienced drivers depicted such behaviors. In summary, hazard anticipation has been shown repeatedly to be significantly poorer in young drivers than more experienced drivers [32], and therefore can serve as a standard for comparing the performance of VR headsets with other measures of latent hazard anticipation.

On the other hand, the inability to detect latent hazards has been linked to the increased rate of [33, 34], making it one of the more critical skills with which to assess VR headsets. In one study, it was reported that out of 1000 crashes reviewed, inexperience and failure to scan for hazards were the main factors contributing to approximately 42.7% of the crashes [35]. It was argued that this was mostly due to the fact that younger drivers are generally inexperienced rather than that they have an increased risk-taking tendency.

While the use of educational tools was proved to be inefficient to prevent driving under the influence and high speeding incidents [36], preventing crashes through additional driver training could play a significant role. Several training programs have been developed and proven to be highly efficient in enhancing young drivers' hazard anticipation skills [37, 38]. These training programs allowed the drivers to retain these skills for up to a year before an on-road evaluation, which showed the long-term effects of these hazard anticipation training programs on novice drivers [39].

The RAPT program is a popular training program that has concentrated on hazard anticipation and poor scanning behavior [29]. Over the years, various versions of RAPT, both on-field and simulator-based, have been incorporated into multiple studies to train novice drivers and analyze their behavior and performance [38, 24]. Results from these studies have indicated the effectiveness of RAPT in the enhancing the hazard anticipation skills of young drivers, with RAPT-trained novice drivers being twice as likely as untrained drivers to recognize hazardous situations or fixate on signs that indicate potential hazards [40]. The RAPT program was also utilized and modified by State Farm into a training suite called RoadAware™, which like all other versions of RAPT has shown to be very effective in improving drivers' performance and behavior. As mentioned before, RAPT has used VR components in its mediation stage, but with limited capacity [20] and has yet to be fully designed around a VR-headset-based driving simulator. This is because such a driving simulator has never been validated for training or driving simulation purposes.

To begin the validation of the VR platform for driving simulation and training purposes, it is vital to replicate results previously validated on another platform. A fixed-base driving simulator was chosen for comparison due to similarities in the manner of simulation and possibility of performance measurement. The current study compares the hazard anticipation performance of young and more experienced, middle-aged drivers on the VR-headset-based driving simulator and a fixed-base driving simulator. The scenarios used by Pradhan et al. [32] were redeveloped on a VR headset using Unity 3D to the closest identifiable approximation. These scenarios were also originally featured on the RAPT program module. We want to determine whether there is a difference in the proportion of latent hazards anticipated by young drivers on the VR headset and fixed-base simulator and correspondingly whether there is a difference between middle-aged drivers on the two simulator platforms. If the differences are small, this will be important evidence that VR headsets can be used to measure one of the most critical of behaviors, latent hazard anticipation.

1.3 Glance Duration

Glance duration refers to the temporal characteristics (how long the driver looked) as opposed to the spatial characteristics of latent hazard anticipation glances (where the driver looked) mentioned in the previous section. The temporal characteristics include both the total time the driver spends glancing at a latent hazard and the duration of each glance at a latent hazard. It is important to know how long drivers glance at a latent hazard because drivers who look for only a short period of time or who take very short glances are less likely to be fully able to perceive a threat, understand what the threat means, and take appropriate action [41].

With regard to temporal characteristics, it has been reported in previous simulator studies that middle-aged drivers spend longer in total looking at latent hazards than their younger counterparts [42, 43]. As for the duration of individual glances, it has been reported that, as measured on a driving simulator or using video clips, there are only marginally significant differences in the average glance durations of middle-aged and younger drivers [44, 45]. For this reason, we have considered both total glance duration and average glance duration as our dependent variables in this study. The current study aims to determine whether these two temporal characteristics (the total duration of the glances at a latent hazard and the average glance duration of each glance at a latent hazard) of young drivers and more experienced, middle-aged drivers are similar on two platforms: a VR-headset-based driving simulator and a fixed-base driving simulator. If the differences between the results acquired on both platforms are small, this will further add to the evidence that VR headsets can be used to measure indices of safe driving behavior.

1.4 Simulator Sickness

With respect to the second concern about VR headsets, simulator sickness, we gave participants the Simulator Sickness Questionnaire [46]. If VR headsets used to evaluate hazard anticipation create increased rates of simulator sickness, then the differences should appear in the scores of the VR headset groups when compared with the fixed-base simulator groups.

Simulator sickness is a major obstacle to the use of driving simulators for research, training, and driver assessment purposes. Due to the large amount of visual flow associated with virtual environments, visual-temporal lags occur, resulting in simulator sickness. There is limited scientific literature about what influences simulator sickness and its subsequent effect on the behavior and performance of the user in the virtual environment. Factors such as age, sex, and psychological traits that increase the likelihood of simulator sickness have been identified. Other factors, such as those related to various elements of the virtual environment (curved roads, high speeds, long durations) and those related to technical setup of the simulator (controls, delay in response), have also been recognized [47, 48].

In the past, driving simulation and human factors researchers have employed several measures to limit the problem of simulator sickness. These include various pre-experimental screening questions during the recruitment stages regarding history with motion sickness and preliminary practice drivers to identify and exclude subjects prone to simulator sickness. Despite these measures, it is seemingly impossible to rule out the chances of a participant experiencing simulator sickness during simulation studies [49].

We want to determine whether there is a difference in the simulator sickness questionnaire scores between corresponding driver groups on the two simulator platforms. If the differences are small, this will further help establish VR headsets as a feasible platform for future driving simulation studies.

2 Methods

2.1 Participant Groups

The study recruited a total of 50 participants. This included 24 young drivers aged 18-21 years; 24 middle-aged drivers aged 30-55 years; and two drop-outs during the preliminary practice drive due to simulator sickness. For the 48 participants who completed the practice drive without any symptoms of simulator sickness, half of the young and middle-aged drivers were randomly assigned either to a fixed-base driving simulator or a VR-headset-based driving simulator. This resulted in four groups of drivers, with each group consisting of 12 drivers: young simulator, middle-aged simulator, young headset, and middle-aged headset. The average age and average driving experience of the participants along with their respective standard deviation are listed groupwise in Table 2.1. The participant sample according to gender is also listed in Table 2.1.

Table 2.1 - Sample characteristics

Driver Group	Age (Years)		Driving Experience (Years)		Population by Gender	
	Average	SD	Average	SD	Male	Female
Middle-Aged Simulator	38.17	7.5369	18.1522	9.6691	7	5
Young Simulator	20.25	0.8292	3.1433	1.2005	9	3
Middle-Aged Headset	39.58	8.7983	21.0142	7.5496	8	4
Young Headset	20.08	0.9538	2.6692	1.1415	8	4

There were no statistically significant differences in the mean ages or years of driving experience of the two young simulator groups or the two middle-aged simulator groups. All participants held a valid United States drivers' license, were recruited from the University of Massachusetts Amherst local area, and were remunerated for their participation. Due to the difficulty posed by eyeglasses during eye-tracking calibration, participants with eyeglasses were excluded from the study. There were no other inclusion or exclusion criteria in this study.

2.2 Apparatus and Software

The apparatus consists of a fixed-base driving simulator, an eye tracker, a VR headset, and vehicle controls. The primary software consists of various programs to create the virtual worlds and coordinate events in these worlds. These are described in more detail below, and the differences between the two simulator platforms are listed in Table 2.2.

2.2.1 Fixed-base driving simulator and eye tracker

RTI Driving Simulator: The Realtime Technologies (RTI) fixed-base driving simulator at the UMass Amherst Arbella Insurance Human Performance Laboratory consists of a fully equipped 2013 Ford Fusion placed in front of five screens with a 330-degree field of view [50]. The five front and side surrounding screens have a display resolution of 1900 x 1200 dpi, with the sixth rear screen having a resolution of 1400 x 1050 (Figure 2.1.). The cab also features two dynamic side mirrors and a rear-view mirror, which provide rear views of the scenarios for the participants. The simulator is equipped with a five-speaker surround system for exterior noise and a two-speaker system for simulating in-vehicle noise. All aspects of the simulator are monitored and coordinated on SimCreator, which is a PC-based program that launches, controls, and collects real-time data from every simulator drive. The scenarios for the driving simulator are designed and developed using software called Internet Scene Assembler (ISA), which contains various commonly used roadway and environmental assets (roads, intersections, buildings, trees, etc.) as well as a user-friendly interface, which helps coordinate scripted events in scenarios, such as the appearance of a pedestrian at a certain distance from the driver's vehicle.

Figure 2.1 - RTI driving simulator

ASL MobileEye: The Applied Science Laboratories (ASL) MobileEye is a monocular eye tracker consisting of a pair of goggles with one camera focused on the eye, another focused on the scene ahead, and a small reflective monocle for the eye camera to view the eye without obstructing the participant's view (Figure 2.2.). Calibration is conducted using a 9-point calibration screen. Eye movements are recorded at a 30 Hz refresh rate, and the gaze cursor is overlaid on the recorded video output. The eye tracker has an accuracy of 0.5 degrees of visual angle. It is used for eye tracking on the fixed-base driving simulator.

Figure 2.2 - ASL MobileEye

2.2.2 VR-headset-based driving simulator

The VR-headset-based driving simulator consists of the Tobii Pro Integrated HTC Vive connected to a Logitech G29 Driving Force steering wheel. Unity 3D was used to initiate and execute scenarios. A typical scenario in this experiment displays a virtual avatar of a generic driver with hands on the steering wheel, seated inside a standard sedan-class automobile. The virtual cab consists of shifters, pedals, a steering wheel, and side/rear view mirrors similar to cabs in the real world. This gives the driver an immersive feel of being seated in an actual car. If the participant moves the steering wheel in the real world, the avatar also move its hands similarly in the virtual world. Below, the individual components of the VR-headset-based driving simulator and their specifications are briefly explained.

Tobii Pro Integrated HTC Vive: This VR headset is a retrofitted version of the HTC Vive Business Edition head-mounted display (HMD) which is integrated with Tobii Eye Tracking [51]. The headset provides a 110-degree field of view with a display resolution of 1080×1200 at a 90 Hz refresh rate. The eye-tracking platform uses the Binocular Dark Pupil Tracking technique [52] to track the pupil and uses a five-point calibration method to provide eye-tracking with up to 0.5 degrees of visual error at a 120 Hz refresh rate (Figure 2.3., left panel).

Logitech G29 Driving Force: The steering wheel features a powerful dual-motor force feedback to simulate the force effects required for an accurate response from the driver, along with good steering action. The 900-degree lock-to-lock rotation enables the wheel to be rotated two and a half times. It also consists of a separate floor pedal unit with integrated throttle, brake, and clutch pedals (Figure 2.3., right panel).

Figure 2.3 - Tobii Pro Integrated HTC Vive (left); Logitech G29 Driving Force (right)

Unity 3D: Unity is an all-purpose game engine that supports 2D and 3D graphics, drag and drop functionality, and scripting through C# (Figure 2.4). In this study, the Unity 3D engine was used to create graphically pleasing, realistic environments featuring several on-road elements and hazards. Assets for the various on-road and environmental elements (such as trees, signage, and vehicles) featured in the scenarios were mostly designed from scratch or imported from numerous resources on the Unity Store.

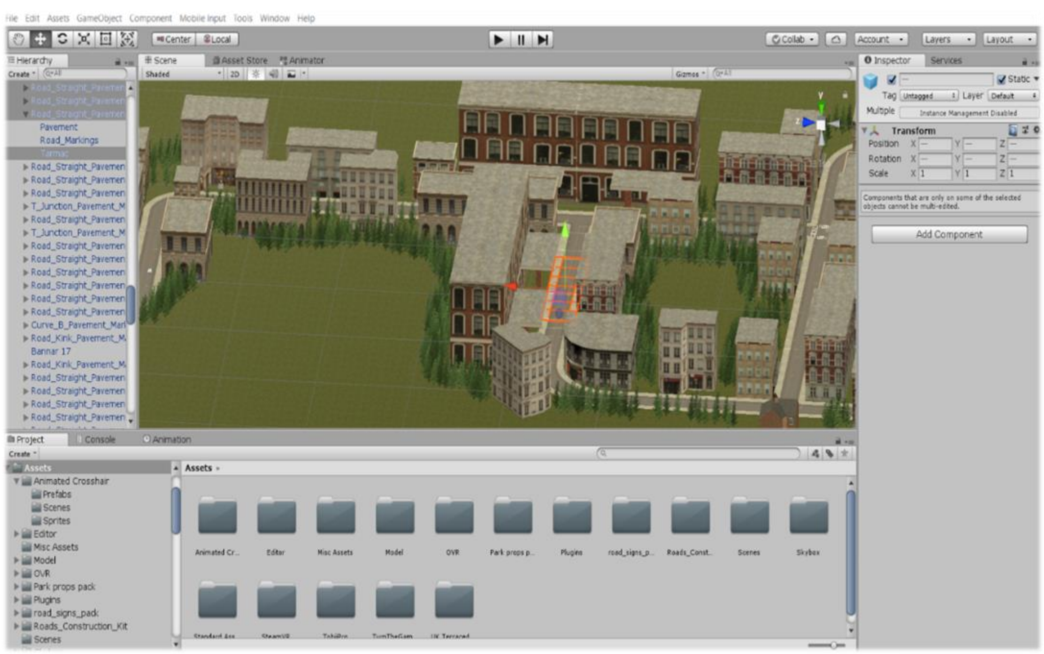

Figure 2.4 - Designing virtual worlds using Unity 3D

Table 2.2 - Differences between the two simulator/eye-tracking platforms

	Fixed-Base RTI Driving Simulator and ASL MobileEye	VR Headset-Based Driving Simulator

Fidelity	High	Low
Vehicle Measures Output	Speed, Lane Deviation, Steering wheel offset, Acceleration etc.	None, but can be programmed to collect desired output
Eye-tracking Output	Gaze point, Gaze direction, Blink Rate, Horizontal & Vertical Dispersion	Gaze point and direction are available by default. Other features are programmable.
Eye-tracking Refresh Rate	30 Hz, monocular tracking	120 Hz, binocular tracking
Field of View	330 degrees (Fixed)	110 degrees (Relative to user's head position)

2.3 Experimental Scenarios

Using Unity 3D and SimCreator, six unique scenarios were designed respectively for the VR-headset-based driving simulator and the fixed-base driving simulator, respectively, in order to examine the driver's ability to anticipate latent hazards. The design and layout of roadways as well as the latent hazard zones featured in these scenarios were identical on both platforms. The signage, traffic control, and lane markings were similar on both platforms. The six scenarios were similar to those used in the RAPT program, which were evaluated by Pradhan et al. [32]. The posted speed limit for the right turn, obscured crosswalk, and obscuring vegetation scenarios was 30 mph, while the posted speed limit for the left-turning truck, pedestrian island, and stop ahead scenarios was 45 mph.

The six scenarios are explained below.

1) Right turn: The driver approaches a stop-sign-controlled four-way intersection with a travel lane in either direction. The driver is expected to turn right at the intersection. There is a crosswalk at the intersection, and a pedestrian approaching the crosswalk is obscured by a building block on the right (Figure 2.5). The driver should scan the obscured area on the right before reaching the intersection to detect any hazards that may arise from the area or to yield to pedestrians that may attempt to cross at the crosswalk.

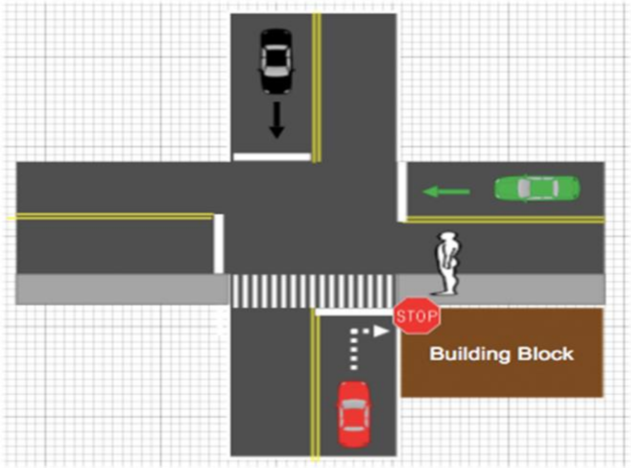

Figure 2.5 - Right turn scenario

2) Left-turning truck: The driver approaches a four-way intersection with two travel lanes in either direction and with cross traffic controlled by stop signs. In the left lane, a line of vehicles is stopped before a truck, which is attempting to make a left turn. The truck blocks the driver's view of any oncoming traffic from the opposing lanes. A car is waiting to turn left in the opposing lane but is obscured to the driver by the aforementioned truck (Figure 2.6). The driver should glance at the right occluding edge of the truck to detect any emerging hazards from obscured areas of the roadways.

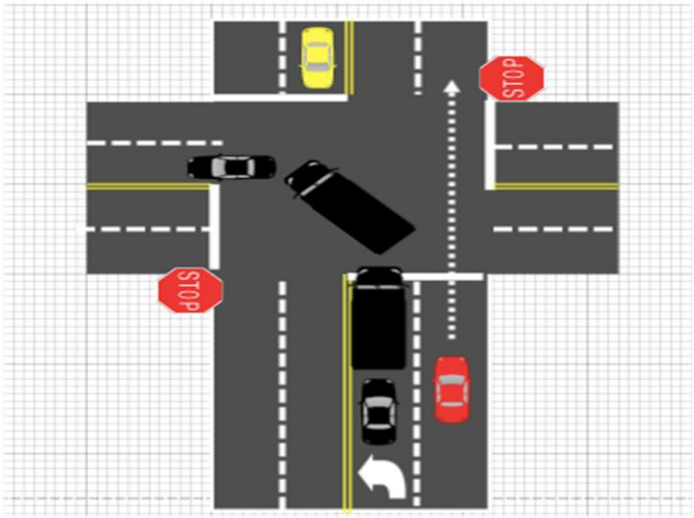

Figure 2.6 - Left-turning truck scenario

3) Obscured crosswalk: There is a truck parked on the right side of a two-lane roadway right before a crosswalk. As the driver nears the truck and tries to pass from its left side, a vehicle approaches in the opposing lane. The presence of the truck obscures any emerging hazards,

such as pedestrians from the right side of the crosswalk. Additionally, the presence of the vehicle in the opposing lane partially obscures the left side of the crosswalk, thereby making it difficult to detect sudden threats from that side of the crosswalk (Figure 2.7). The driver should scan the left side of the crosswalk now obscured by the approaching vehicle and also the left front edge of the truck on the right.

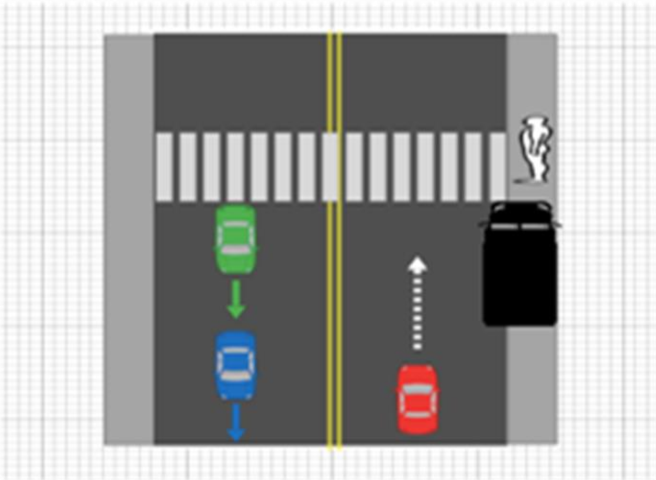

Figure 2.7 - Obscured crosswalk scenario

4) Pedestrian island: The driver is in the right lane while approaching a T-intersection. Only the stem of the T is controlled by a stop sign. In the left lane, a line of vehicles waits to turn left. The median to the left of the line accommodates a pedestrian island at the crosswalk. A pedestrian on this island is obscured by the line of vehicles (Figure 2.8). The driver should scan towards the front right edge of the first vehicle in the line of vehicles waiting to turn left to detect any obscured pedestrians who may be attempting to cross.

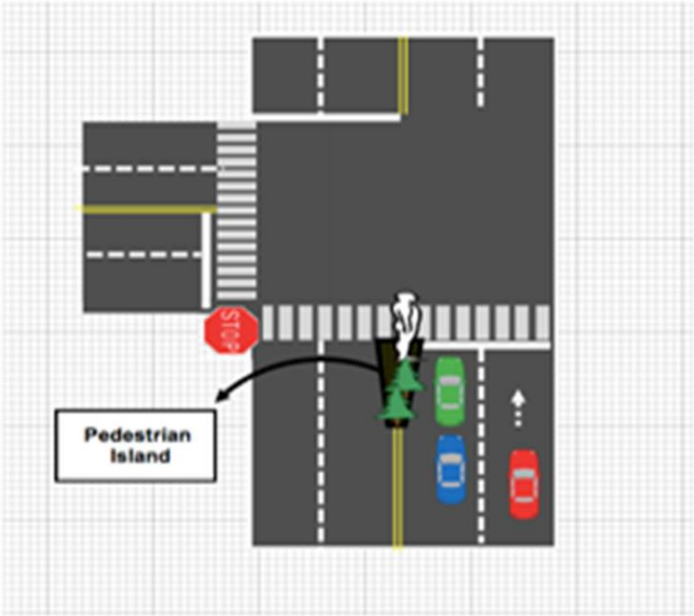

Figure 2.8 - Pedestrian island scenario

5) Obscuring vegetation: The driver is approaching a stop-sign-controlled T-intersection with one travel lane in either direction. There is a pedestrian at the crosswalk, which lies beyond the intersection to the driver's right side. Vegetation obscures the stop sign as well as the driver's view of the crosswalk (Figure 2.9). At the intersection, the driver should continuously scan towards the obscured area on his or her right side while attempting to turn right in order to detect any potential hazards emerging from the obscured area.

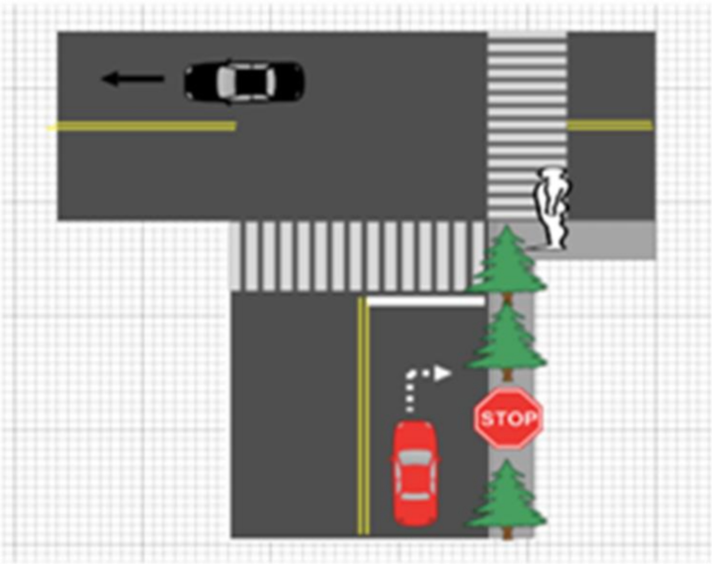

Figure 2.9 - Obscuring vegetation scenario

6) Stop ahead: The driver is traveling on a road curving to the right and approaching a stop-sign-controlled intersection. At the beginning of the curve, a Stop Ahead sign exists, and the stop sign at the end of the curve is partially obscured by vegetation (Figure 2.10). The driver should notice the Stop Ahead sign and then correctly identify the stop sign and stop at the intersection.

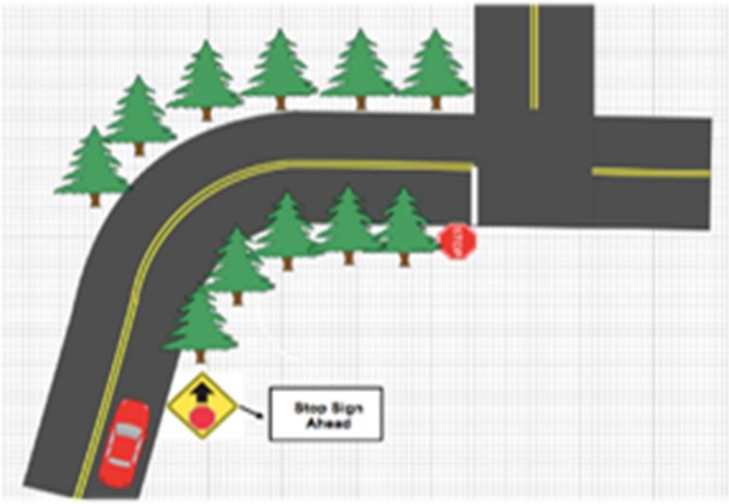

Figure 2.10 - Stop ahead scenario

2.4 Experimental Design

The experimental design was a 2 x 2 x 6 mixed with platform (fixed-base driving simulator or VR-headset-based driving simulator) and driving experience (young or middle-aged) as the two between-subject factors and scenario as the within-subject factor. The between-subject design for platform was chosen because the scenarios were conceptually identical on both platforms and, in a between-subject design, there would not be an instance where any learning effects experienced by participants after their first exposure to a specific scenario would transfer to their second exposure. Between-subject designs are valid as long as the participants are assigned randomly to the different conditions [53]. The four groups of participants navigated six scenarios overall on their assigned platform. The order of the scenarios presented to participants was counterbalanced across and within groups using a balanced Latin square method [54].

2.5 Simulator Sickness Questionnaire

The Simulator Sickness Questionnaire is the most widely used tool to measure simulator sickness [55]. Among the 16 symptoms listed on the SSQ, there were sets of symptoms that were correlated, and three subscales were identified: nausea (N), oculomotor problems (O), and disorientation (D). Each participant rated a symptom score of 0, 1, 2, or 3. For example, let's say a participant rates the seven symptoms under disorientation as, 2, 2, 3, 2, 3, 2, 1. The unweighted disorientation factor score will be 15, and the weighted disorientation score will be 15 × 7.58. Similarly, the weights for N and D are 9.54 and 13.92. The total score will be equal to N + O + D × 3.74. In our experiment we computed the total score calculated from the participants' responses on the SSQ for each of the driving experience groups (young and middle-aged) on both platforms along with the weighted nausea, oculomotor, and disorientation scores.

2.6 Driver Behavior Questionnaire

This study utilizes the North American version of the Driver Behavior Questionnaire (DBQ), which was originally developed in the United Kingdom. The DBQ is a widely used tool to measure driving behaviors linked to collision risks [56]. Three subscales were identified for the 24 items listed in the questionnaire in the form of questions: error (E), lapses (L), and violations (V). Each participant rated an item on a scale of 0 to 5 (rarely to always), based on how often they engaged in the behavior mentioned in that item. For example, "Try to pass another car that is signaling a left turn" is an error-related item, and a participant who rarely engages in such behavior would rate this item as 0. Each DBQ has eight items for each subscale. In our experiment, we computed the average score for each subscale based on each participants' responses for each of the driving experience groups (young and middle-aged) on both platforms.

2.7 Post-Study Questionnaire

The Post-Study Questionnaire (PSQ) was developed for this study to compare several user-experience-based attributes of the VR-headset-based driving simulator and the fixed-base driving simulator. The questionnaire listed the following attributes of the simulator which were to be rated on a scale of 1 to 5 (Very Bad to Very Good): Navigation, Driving Controls, Graphical Quality, Sense of Realism, Audio Quality, Wearable Equipment, and Seating Comfort. The average rating for each attribute was computed for each participant for each of the driving experience groups (young and middle-aged) on both platforms along with the overall rating by each participant, which is the average score of all the attributes' rating for each participant.

2.8 Procedure

After informed consent was obtained from the participants, a pre-study questionnaire and a DBQ were administered to record data related to demographics, driver experience, and drivers' tendency to engage in aggressive behavior while driving. Next the participants were given basic instructions such as to follow on-screen/audio instructions and maintain the posted speed limit. Eye-tracking calibration was done to ensure accurate eye-tracking data. The participants on the VR-headset-based driving simulator were given a short tutorial on different aspects of the headset and steering wheel. Both sets of participants then drove through a preliminary practice drive for the next five minutes. The purpose of this practice drive was to familiarize the participants with the virtual world and the controls of the cab. The virtual world featured in the practice drive was a closed-loop roadway consisting of several left/right turns, curves, intersections, and straight roads. While navigating through the practice drive, they were shown the rearview and sideview mirrors and were asked to brake, accelerate, and make left/right turns. Once they concluded the practice drive, participants were permitted to continue to the experimental scenarios if they felt confident about driving and maneuvering through the simulation. A set of six counterbalanced scenarios was then introduced to the participants with a gap of 30 seconds between each scenario. This session lasted for approximately 45 minutes. After concluding the driving session, an SSQ was administered to track any symptoms of simulator sickness [46]. The PSQ was also filled out by the participants.

2.9 Analysis Techniques

The dependent variables considered for this experiment were binary scored latent hazard anticipation (whether the driver detected the latent hazard), glance duration (how long the driver scanned for the latent hazard), simulator sickness severity, and user-experience-based attributes of the simulator platform. To analyze these variables, eye-tracking data was decoded from the recorded videos of each participant's drives through each of the six scenarios, and their responses on the SSQ and PSQ were analyzed. In addition to these variables, driver behavior questionnaire responses were also analyzed to weed out anomalies that may arise during a between-subject design experiment.

As mentioned earlier, in order to examine the drivers' latent hazard anticipation behavior (looked or did not look), the eye-tracking data from the recorded videos were binary scored (0 or 1). A set of launch zones and target zones was predetermined for each scenario based on previous studies [57, 58]. A *target zone* is defined as an area(s) of the roadway from where potential threats may emerge. A *launch zone* is defined as that area of the roadway where the drivers should begin scanning towards the target zone to successfully identify the presence of any potential threats. Participants who successfully glanced at the target zone while in the launch zone in a given scenario were scored 1, while those who failed to do so were scored 0. The concept of launch zones and target zones is perhaps better understood with an example. Let us consider the obscured crosswalk scenario. Figure 2.11. shows the launch zone and target zones for this scenario.

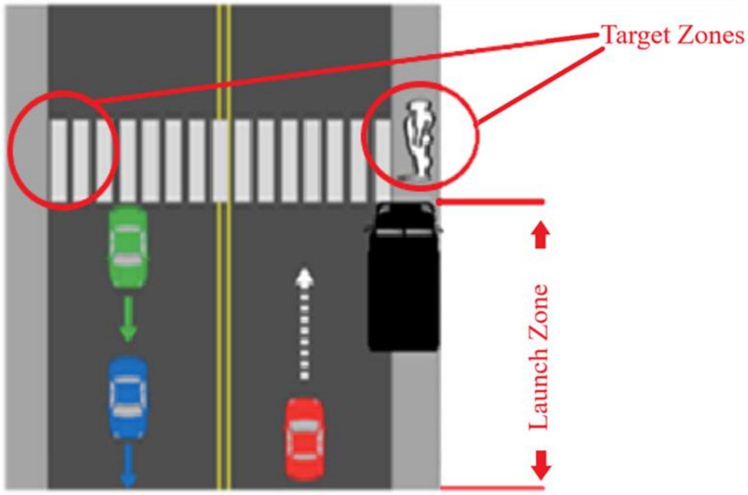

Figure 2.11 - Launch zone and target zones for the obscured crosswalk scenario

The launch zone in Figure 2.11 starts from a point that is 5 seconds before the crosswalk lying ~50 ft before the crosswalk. The target zones are the two obscured sides of the crosswalk, where potential threats can emerge. In order to be scored 1, the participant will need to scan both the target zones at least one time after entering the launch zone. Figure 2.12 and Figure 2.13 show the obscured crosswalk scenario from the driver's point of view on both platforms. The drivers in both instances have successfully identified the target zones in the scenario.

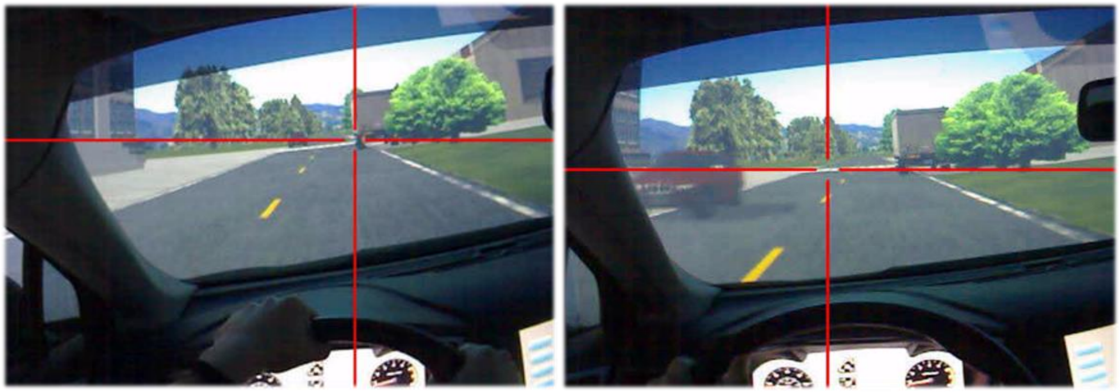

Figure 2.12 - Successful detection for the obscured crosswalk scenario on the fixed-base driving simulator

Figure 2.13 - Successful detection for the obscured crosswalk scenario on the VR-headset-based driving simulator

The term *glance* in this experiment is used to refer to one or more sequential fixations on the target zone when the participant is in the launch zone in a particular scenario. Each frame includes an indication of where the driver is looking in the frame. In a frame-by-frame tracking of the recorded videos (one frame = 33 milliseconds), every sequence of frames in which the driver is looking at the target zone from the launch zone is recorded as a glance. A participant usually makes more than one glance in the scenario where he or she successfully detected the latent hazards. The *total glance duration* is the sum of the duration of all glances made by a participant in a scenario at a latent hazard, while the *average glance duration* is the mean duration of all glances at the latent hazard.

For the scenarios where the participant successfully glanced at the target zone(s), the total and average glance duration were calculated. The process of calculating the glance duration is illustrated below with the help of figures.

At frame #3452, the participant, upon entering the launch zone, has not yet scanned the target zone (Figure 2.14).

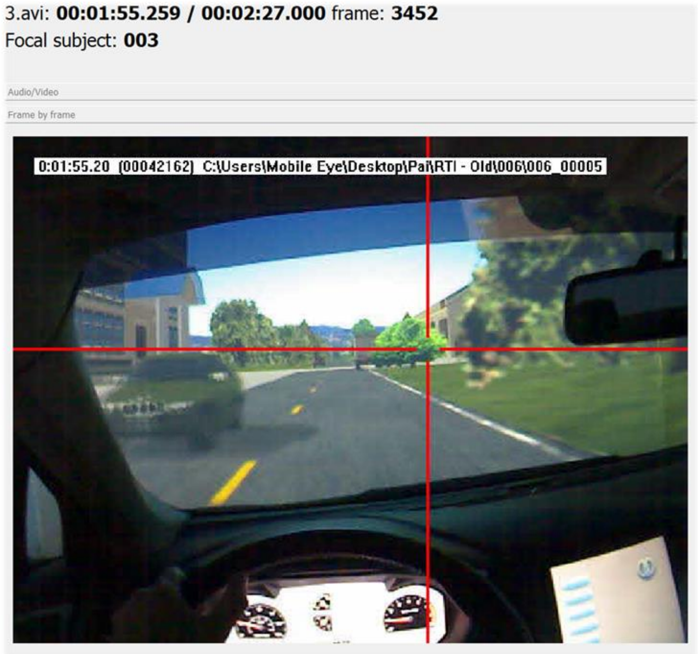

Figure 2.14 - Participant has yet to scan the target zone(s)

At frame #3453, the participant scans the left side of the crosswalk, which is one of the target zones, and continues scanning that zone until frame #3474. (Figure 2.15.)

Figure 2.15 - Participant begins scanning a target zone (left); participant stops scanning the target zone (right)

Since each frame is 33 milliseconds each, the amount of the time the participant spent glancing at the target zone, i.e., the glance duration, is 3473 – 3453 = 20 * 33 = 660 milliseconds.

At frame #3477, the participant begins scanning the left edge of the truck on the right side of the crosswalk, which is the other target zone in this scenario. He/she continues to do so until frame #3496 (Figure 2.16).

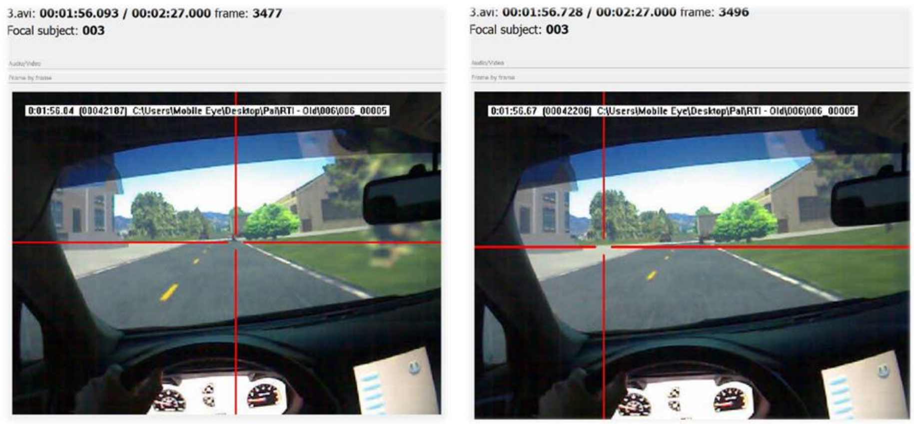

Figure 2.16 - Participant begins scanning another target zone (left); participant stops scanning the target zone (right)

The amount of the time the participant spent glancing at the target zone, i.e., the glance duration, is 3496 – 3477 = 19 * 33 = 627 milliseconds. Considering these two glances at the target zones, the total glance duration would be the sum of the glance duration, which is 1287 milliseconds, or 1.28 seconds. The average glance duration would be 643.5 milliseconds, or 0.64 seconds.

3 Results and Discussion

3.1 Latent Hazard Anticipation

In order to analyze the binary scored, binomially distributed eye-tracking data, a logistic regression model within the framework of generalized estimation equations (GEE) was used. The model included driver experience (younger and older) and the two platforms (VR-headset-based driving simulator and fixed-base driving simulator) as the between-subject factors, while scenario type was considered as a within-subject factor. The significance level was set at .05, and the participants were included as a random effect in the model. The model was used to determine whether there was a significant difference between the proportion of latent hazards detected by participants across two groups (young vs middle-aged) and two platforms (fixe-based driving simulator vs VR-headset-based driving simulator), as well as whether there was an interaction between scenario type and platform.

A backward elimination procedure was used to eliminate any non-significant higher-order interactions. The final model revealed a highly significant main effect of driving experience [Wald χ^2 = 28.72; $p < 0.001$], which is consistent with the results of Pradhan et al. [32], as well as our expected results. There was no significant effect of the platform [Wald χ^2 = 0.117; $p > 0.05$].

The second-order interaction between driver experience and platform was not significant. There was a significant effect of scenario type [Wald χ^2 = 4871.61; $p < 0.001$], but the second-order interaction between scenario type and platform was not significant.

For both platforms, the proportion of latent hazards detected was smaller for young driver groups when compared to their middle-aged driver counterparts on the same platform. On the fixed-base driving simulator, middle-aged drivers anticipated 92% of the latent hazards compared to only 64% for the young drivers. Similarly, on the VR-headset-based simulator, the middle-aged drivers anticipated 90% of the latent hazards compared to 62% for the young drivers (Figure 3.1).

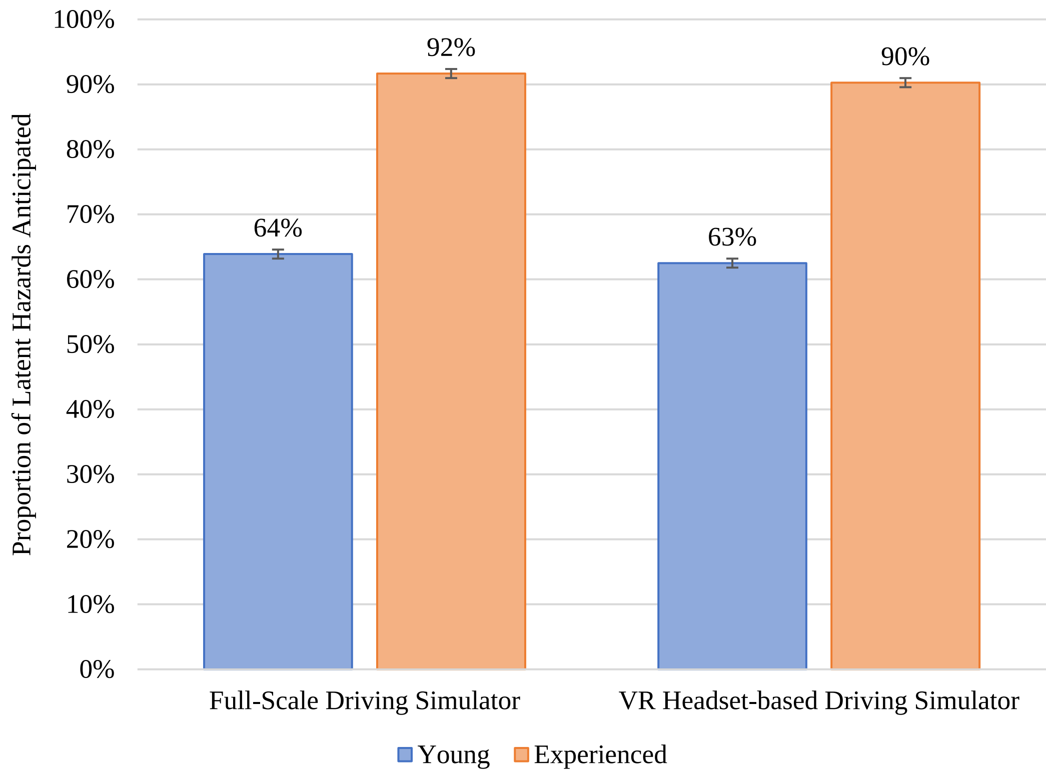

Figure 3.1 - Proportion of latent hazards anticipated by each group

3.2 Glance Duration

A 2 × 2 factorial [2 driving experience groups: young and middle-aged; 2 platforms: VR-headset-based simulator and fixed-base driving simulator] ANOVA was performed separately for the total glance duration and average glance duration for each scenario for each participant, n = 48, $\alpha = 0.05$.

3.2.1 Total glance duration

Analysis of the total glance duration indicated no main effect of platform ($F = 2.309$; p-value = 0.130; $\eta^2 = 0.010$) or interaction between experience and platform ($F = 2.733$; p-value = 0.1; η^2 = 0.012). There was a main effect of driving experience ($F = 19.9$; p-value < 0.005; $\eta^2 = 0.084$).

These results are consistent with those reported in previous studies, which state that middle-aged drivers glance in total at potential threats longer than their younger counterparts [42, 43].

3.2.2 Average glance duration

For average glance duration, there was no interaction between experience and platform ($F = 0.042$; p-value = 0.838; $\eta^2 = 0.0002$) or main effect of platform ($F = 3.42$; p-value = 0.066; $\eta^2 = 0.015$) or of driving experience ($F = 3.429$; p-value = 0.065; $\eta^2 = 0.015$). These results are consistent with those reported in previous studies, which state that the difference in the average glance duration between younger and middle-aged drivers is marginal or non-existent [44, 45].

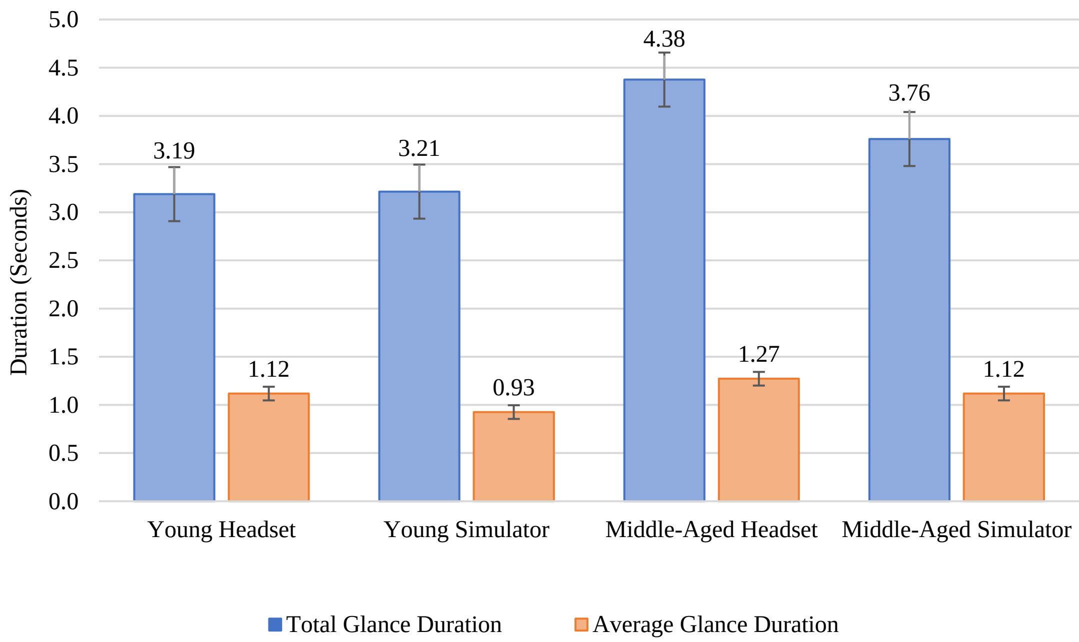

Figure 3.2 - The mean average glance duration and mean total glance duration for each driver group

3.3 <u>Simulator sickness questionnaire</u>

Data from the simulator sickness questionnaire was collected and processed. While all drivers assigned to the fixed-base driving simulator groups completed their drives, two drivers assigned to the VR-headset-based driving simulator (one young, one middle-aged) dropped out during or right after the preliminary practice drive and were immediately withdrawn from the study.

A 2 × 2 factorial [2 driving experience groups: young and middle-aged; 2 platforms: VR-headset-based driving simulator and fixed-base driving simulator] ANOVA was performed for the SSQ total scores as well as for the individual weighted scores for the three subscales (nausea, oculomotor, and disorientation) for the non-dropout participants, $n = 48$, $\alpha = 0.05$.

3.3.1 Nausea

No interaction between experience and platform (F = 1.348; p-value = 0.252; η^2 = 0.030) or main effect of platform (F = 0.84; p-value = 0.773; η^2 = 0.002) were observed, although there was a main effect of driving experience (F = 7.207; p-value = 0.010; η^2 = 0.141).

3.3.2 Oculomotor

No interaction between experience and platform (F = 1.179; p-value = 0.284; η^2 = 0.026) or main effect of platform (F = 0.354; p-value = 0.555; η^2 = 0.008) were observed, although there was a main effect of driving experience (F = 4.269; p-value = 0.045; η^2 = 0.088).

3.3.3 Disorientation

No interaction between experience and platform (F = 2.928; p-value = 0.094; η^2 = 0.062) or main effect of platform (F = 0.007; p-value = 0.932; η^2 = 0.000) were observed, although there was a main effect of driving experience (F = 7.973; p-value = 0.007; η^2 = 0.153).

3.3.4 Total Severity

No interaction between experience and platform (F = 0.322; p-value = 0.573; η^2 = 0.007) or main effect of platform (F = 0.688; p-value = 0.411; η^2 = 0.015) were observed, although there was a main effect of driving experience (F = 14.641; p-value = 0.0004; η^2 = 0.25). Similar results were found when the weighted scores for nausea, oculomotor, and disorientation were analyzed.

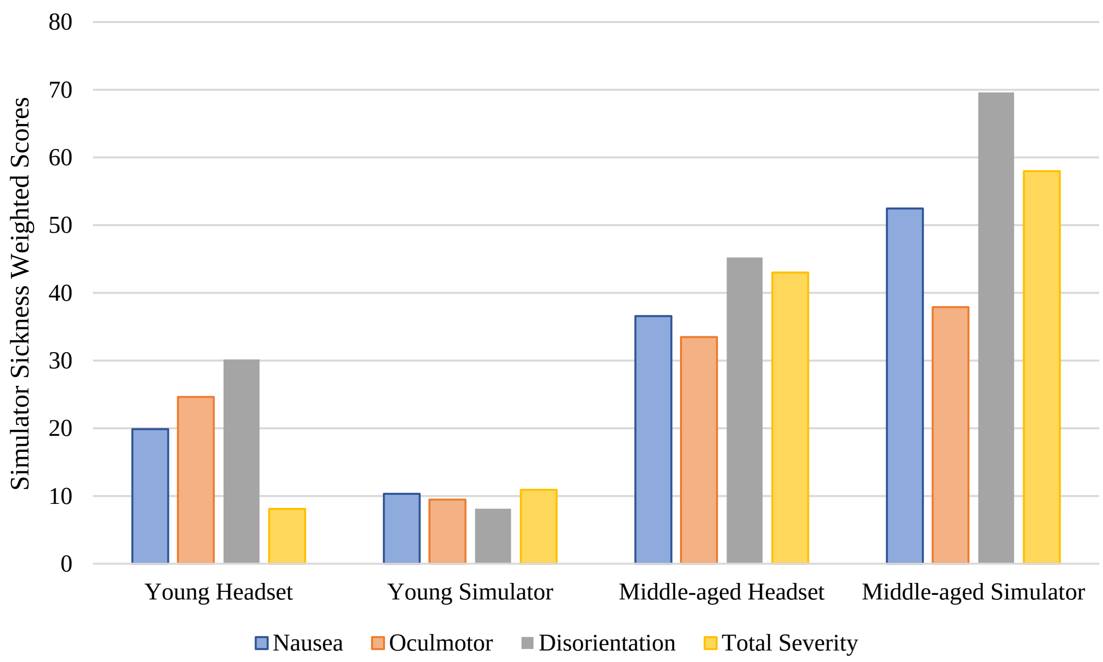

Figure 3.3 - The weighted simulator sickness scores for each driver group

3.4 Driver Behavior Questionnaire

A 2 × 2 factorial [2 driving experience groups: young and middle-aged; 2 platforms: VR-headset-based simulator and fixed-base driving simulator] ANOVA was performed for the average scores for error, lapse, and violation, $n = 48$, $\alpha = 0.05$.

3.4.1 Error

No interaction between experience and platform ($F = 0.22$; p-value = 0.641; $\eta^2 = 0.005$) or main effect of platform ($F = 0.74$; p-value = 0.394; $\eta^2 = 0.017$) or driving experience ($F = 0.055$; p-value = 0.816; $\eta^2 = 0.001$) were observed.

3.4.2 Lapse

No interaction between experience and platform ($F = 1.914$; p-value = 0.174; $\eta^2 = 0.042$) or main effect of platform ($F = 0.733$; p-value = 0.396; $\eta^2 = 0.016$) or driving experience ($F = 1.254$; p-value = 0.269; $\eta^2 = 0.028$) were observed.

3.4.3 Violation

No interaction between experience and platform ($F = 0.561$; p-value = 0.458; $\eta^2 = 0.013$) or main effect of platform ($F = 0.773$; p-value = 0.384; $\eta^2 = 0.017$) or driving experience ($F = 0.027$; p-value = 0.871; $\eta^2 = 0.001$) were observed.

3.5 Post-Study Questionnaire

A 2 × 2 factorial [2 driving experience groups: young and middle-aged; 2 platforms: VR-headset-based simulator and fixed-base driving simulator] ANOVA was performed for the average scores of each attribute to check for main effects or an interaction effect. Apart from driving controls ($F = 5.038$; p-value = 0.03; $\eta^2 = 0.103$), no other attribute had a significant main effect of driving experience. Among all the attributes analyzed, only navigation ($F = 6.856$; p-value = 0.012; $\eta^2 = 0.135$) and driving controls ($F = 36.52$; p-value < 0.005); $\eta^2 = 0.454$) had a significant main effect of platform, while no interaction effect between driving experience and platform was found for any attributes. The average scores for each of the attributes for all groups are listed below in Table 3.1.

Table 3.1 - Mean post-study questionnaire scores for each group (scaled 1 to 5)

	Young Headset	Young Simulator	Middle-aged Headset	Middle-aged Simulator
Navigation	3.75	4.25	3.33	4.08
Driving Controls	2.75	4.08	2.08	3.67
Graphical Quality	3.25	3.92	3.92	3.33
Sense of Realism	3.58	3.75	3.50	3.67
Audio Quality	4.00	4.00	3.67	3.83

Wearable Equipment	4.08	3.83	4.17	3.50
Seating Comfort	4.08	4.42	4.17	4.50
Overall Rating	3.64	4.04	3.55	3.80

4 Discussion

VR headsets are much less expensive than fixed-base driving simulators and therefore could greatly extend the power of simulation. Yet, even if valid as a way to measure something like latent hazard anticipation, they have produced documented evidence of simulator sickness [22, 23]. Thus, it is important to understand not only whether VR headsets are valid, but also whether they can be put to practical use. The current study sought to fill a gap in the literature by examining the validity of VR headsets at measuring driver performance (hazard anticipation ability) compared to a fixed-base driving simulator. While we could have chosen other metrics of performance to validate the platform, we chose to measure latent hazard anticipation ability both because it has been demonstrated to be significantly higher for middle-aged drivers than for young drivers on fixed-base driving simulators and on the open road [30, 31, 32] and because it is linked to crashes.

4.1 Latent Hazard Anticipation

Consistent with our expected results, the results of the current study showed that the proportion of latent hazards anticipated by middle-aged drivers was significantly more than that anticipated by young drivers on both the VR-headset-based driving simulator (90% for middle-aged vs 62.5% for young, a difference of 27.8 percentage points) and the fixed-base driving simulator (91.7% for middle-aged vs 64% for young, a difference of 27.7 percentage points). This result was also in line with results from previous research conducted on driving simulators and in the field that demonstrated that middle-aged drivers anticipate a significantly greater proportion of latent hazards than young drivers [32]. The result from the mixed-effect logistic regression model showed that there was no impact of platform on performance for either the young or the middle-aged drivers.

4.2 Glance Duration

The current study seeks to add more evidence in support of using VR headsets to measure driver performance (total glance duration and average glance duration of anticipatory glances) in safety-critical tasks where a fixed-base driving simulator might normally be used. In particular, the results showed that middle-aged drivers spent a longer time glancing at latent hazards than did young drivers on both the VR-headset-based and fixed-base driving simulators. With this in mind, it is also important to note that the average glance duration was the same among young and middle-aged drivers across both platforms. Had the middle-aged drivers' average glance duration at the latent hazards been longer than those of younger drivers, the middle-aged drivers would potentially have compromised their safety. Both results are in line with results from previous research conducted on driving simulators and on-road studies, which demonstrated that while middle-aged drivers gaze longer at latent hazards, i.e., have a longer total glance duration, there may only be marginal or no differences in terms of their average glance duration when compared to younger drivers. Most importantly, the results from the ANOVA models for total glance duration and average glance duration showed that there was no impact of platform on performance for either the young or the middle-aged drivers.

4.3 Simulator Sickness

Driving-simulator-based studies have always presented difficulties associated with high attrition rates due to simulator sickness or simulator adaptation syndrome for both young and old drivers [59]. Virtual reality headsets have also been associated with such difficulties, with several studies reporting a high attrition rate among users due to motion sickness [22, 23]. Hence, the current study also examined the effect of simulator sickness on both platforms by comparing data collected from a standard SSQ [46]. Consistent with this, two drivers assigned to the VR headset group dropped out of the study, but none in the fixed-base simulator group dropped out. This drop-out rate is less than 10% and, for most studies, may not pose a serious limitation. Importantly, the weighted subscale scores and total simulator sickness scores among those who completed the experiment were compared between all driver groups on both platforms. The results indicated that there was no significant difference between simulator sickness scores on either platform. There was a significant main effect of driving experience on both platforms, with middle-aged drivers having significantly higher severity scores than young drivers. This is generally consistent with previous studies, which state that older drivers are more prone to the symptoms of simulator sickness than younger drivers [49, 60]. Furthermore, the lack of significance for any second-order interaction between driving experience and platform indicated that the difference between the simulator sickness scores of middle-aged and young drivers was similar on both platforms.

4.4 Driver Behavior Questionnaire

With every between-subject design, there exists a possibility for certain confounds to arise, such as overrepresentation in one group of drivers who tend to engage in aggressive, aberrant driving behavior. In order to determine whether such confounds were present, a DBQ was administered. Results show no indication of such confounds, with no significant effect in questionnaire responses across all platforms and driving experience groups.

4.5 Post-study Questionnaire

The PSQ was administered to identify the various attributes of the VR-headset-based driving simulator that could be improved, based solely on user experience. Analysis of the participants' responses on the questionnaire indicated that, although several attributes are already on par with the fixed-base driving simulator, a few attributes, such as navigation and driving controls, could be improved on the VR-headset-based driving simulator, since the VR simulator received 15% and 38% lower rating on said attributes when compared to the fixed-base simulator. Driving controls were also perceived differently by the younger drivers and the middle-aged drivers; the younger drivers were 12% more likely to rate the controls favorably than the middle-aged drivers.

4.6 Limitations and Future Work

The study has several important limitations. First, the current study used a between-design experiment to address the hypothesis that drivers would perform similarly on a VR-headset-based driving simulator and a fixed-base driving simulator. In these kinds of experiments, it is difficult to maintain complete homogeneity across the groups despite random assignment. It would be useful to consider a within-subject design with matching or block randomization techniques to eliminate confounds. In such a case, it would be worth looking into the possibility of integrating the VR headset to the controls of the fixed-base simulator in order to improve the comparison between the two platforms. Second, this study validated the VR platform based only on the hazard anticipation skills of young and middle-aged drivers. Future studies should consider investigating other crash-avoidance skills such as hazard mitigation and attention maintenance. Third, other measures of driving performance may also be considered for validation of a platform (e.g., various vehicle measures such as the standard deviation of lane position, or other eye movement measures such as horizontal and vertical gaze dispersion, physiological variables such as percentage of eye closure and blink rate, or perhaps even workload metrics). Fourth, while the two platforms were found to differ little in terms of dropout rates and not at all in terms of severity of simulator sickness among those who completed the experiment, an evaluation of drivers aged 65 years and above needs to be considered to measure true effectiveness.

4.7 Conclusion

In summary, the current study showed that VR headsets may be used to effectively measure driver performance, specifically spatial characteristics of latent hazard anticipation behaviors and also temporal characteristics. It suggests that VR headsets can potentially be used to measure a wide range of safety-critical behaviors, not only hazard anticipation behaviors. Such additional behaviors are known to include hazard mitigation behaviors as well as attention maintenance behaviors [61]. Virtual reality headsets also appear, at least with hazard anticipation scenarios, not to generate more than minimal simulator sickness.

With the successful initial validation of the VR-headset-based driving simulator for measuring hazard anticipation behavior, future work can concentrate on recreating the RAPT program exclusively for VR-headset-based driving simulator platforms. As more measures are validated for the platform, more training programs can be designed or re-designed. This could result in VR-headset-based driving simulators being used in training novice drivers or older drivers on a widespread basis, something that is not possible with more expensive fixed-base driving simulators.

Virtual reality headsets offer promise as an alternative to conventional simulators, especially as a platform that can easily accommodate multiple users. The range of applications in which VR-headset-based driving simulators can now be employed has greatly expanded. Multiple-vehicle conflicts involving multiple drivers or road users, e.g., scenarios in which each driver is using different levels of automation, is one research theme that may be suitably addressed using VR-headset-based simulators. They could also be used during licensure to evaluate drivers' crash-avoidance skills. The opportunities are many, and the impact could potentially be equally large.

References

1. Slob, J.J. State-of-the-art Driving Simulators, A Literature Survey. DCT Report. Eindhoven University of Technology, Department Mechanical Engineering, Control Systems Technology Group, 2008. 107: 1-19.
2. Lee, J. D., Caven, B., Haake, S., & Brown, T. L. Speech-based interaction with in-vehicle computers: The effect of speech-based e-mail on drivers' attention to the roadway. *Human factors*, 2001. 43(4): 631-640.
3. Godley, S. T., Triggs, T. J., & Fildes, B. N. Driving simulator validation for speed research. *Accident analysis & prevention*, 2002. 34(5): 589-600.
4. Roenker, D. L., Cissell, G. M., Ball, K. K., Wadley, V. G., & Edwards, J. D. Speed-of-processing and driving simulator training result in improved driving performance. *Human factors*, 2003. 45(2): 218-233.
5. Metzger, P.J. Adding reality to the virtual. In *Virtual Reality Annual International Symposium (IEEE),* 1993. 7-13.
6. Winn, W. Learning in virtual environments: A theoretical framework and considerations for design. *Education Media International*, 1999. 36(4): 271-279.
7. Velev, D., & Zlateva, P. Virtual reality challenges in education and training. *International Journal of Learning and Teaching,* 2017. 3(1): 33-37.
8. Pulijala, Y., Ma, M., Pears, M., Peebles, D., & Ayoub, A. Effectiveness of immersive virtual reality in surgical training—a randomized control trial. *Journal of Oral and Maxillofacial Surgery*, 2018. 76(5): 1065-1072.
9. Oagaz, H., Sable, A., Choi, M. H., Xu, W., & Lin, F. VRInsole: An unobtrusive and immersive mobility training system for stroke rehabilitation. In *Wearable and Implantable Body Sensor Networks (BSN), 2018 IEEE 15th International Conference,* 2018. pp. 5-8.
10. Lei, X., Zhang, A., Wang, B., & Rau, P. L. P. Can Virtual Reality Help Children Learn Mathematics Better? The Application of VR Headset in Children's Discipline Education. In *International Conference on Cross-Cultural Design,* 2018. pp. 60-69.
11. DeLuca, M., & Deluca, J. (2003). *U.S. Patent No. 6,559,813*. Washington, DC: U.S. Patent and Trademark Office.
12. Marks, S., Estevez, J. E., & Connor, A. M. (2014, November). Towards the Holodeck: fully immersive virtual reality visualisation of scientific and engineering data. In *Proceedings of the 29th International Conference on Image and Vision Computing New Zealand* (pp. 42-47). ACM.
13. Burns, M. Immersive Learning for Teacher Professional Development. Retrieved from https://elearnmag.acm.org/featured.cfm?aid=2181208. Accessed July 21, 2018
14. Darken, R. P., & Sibert, J. L. Wayfinding strategies and behaviors in large virtual worlds. In *Proceedings of the SIGCHI conference on Human factors in computing systems,* 1996. 142-149.
15. McComas, J., MacKay, M., & Pivik, J. (2002). Effectiveness of virtual reality for teaching pedestrian safety. *CyberPsychology & Behavior*, *5*(3), 185-190.
16. Hilfert, T., & König, M. (2016). Low-cost virtual reality environment for engineering and construction. *Visualization in Engineering*, *4*(1), 2.
17. Desai, P. R., Desai, P. N., Ajmera, K. D., & Mehta, K. (2014). A review paper on oculus rift-a virtual reality headset. arXiv preprint arXiv:1408.1173.

18. Nichols, S. (1999). Physical ergonomics of virtual environment use. *Applied Ergonomics*, *30*(1), 79-90.
19. Tan, C. T., Leong, T. W., Shen, S., Dubravs, C., & Si, C. (2015, October). Exploring gameplay experiences on the oculus rift. In Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play (pp. 253-263). ACM.
20. Agrawal, R., Knodler, M., Fisher, D. L., & Samuel, S. Virtual Reality Headset Training: Can It Be Used to Improve Young Drivers' Latent Hazard Anticipation and Mitigation Skills. *Transportation Research Record: Journal of the Transportation Research Board,* 2018. 0361198118758311.
21. Johnston, A. P., Rae, J., Ariotti, N., Bailey, B., Lilja, A., Webb, R., ... & McGhee, J. Journey to the centre of the cell: Virtual reality immersion into scientific data. *Traffic*, (2018). 19(2): 105-110.
22. Munafo, J., Diedrick, M., & Stoffregen, T. A. The virtual reality head-mounted display Oculus Rift induces motion sickness and is sexist in its effects. *Experimental brain research*, 2017. 235(3): 889-901.
23. Jensen, L., & Konradsen, F. A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies,* 2018. 23(4): 1515-1529.
24. Vlakveld, W. Hazard anticipation of young novice drivers: Assessing and enhancing the capabilities of young novice drivers to anticipate latent hazards in road and traffic situations. Doctoral dissertation, University of Groningen, 2011.
25. McDonald, C. C., Goodwin, A. H., Pradhan, A. K., Romoser, M. R., & Williams, A. F. A review of hazard anticipation training programs for young drivers. *Journal of Adolescent Health*, 2015. 57(1): 15-23.
26. Borowsky, A., Shinar, D., & Oron-Gilad, T. (2010). Age, skill, and hazard perception in driving. *Accident Analysis & Prevention*, 42(4), 1240-1249.
27. Mehranian, H. (2013). Novice drivers: Development and evaluation of training program for hazard anticipation, hazard mitigation and attention maintenance skills in complex driving scenarios. University of Massachusetts Amherst.
28. Mayhew, D. R. (2007). Driver education and graduated licensing in North America: Past, present, and future. *Journal of safety research*, 38(2), 229-235.
29. Fisher, D. L., Laurie, N. E., Glaser, R., Connerney, K., Pollatsek, A., Duffy, S. A., & Brock, J. The use of an advanced driving simulator to evaluate the effects of training and experience on drivers' behavior in risky traffic scenarios. *Human Factors*, 2002. 44(2): 287-302.
30. Lee, S. E., Klauer, S. G., Olsen, E. C. B., Simons-Morton, B. G., Dingus, T. A., Ramsey, D. J., Ouimet, M. C. Detection of road hazards by novice teen and experienced adult drivers. *Transportation Research Record*: *Journal of the Transportation Research Board*, 2008. 2078: 26-32.
31. Romoser, Matthew RE, Alexander Pollatsek, Donald L. Fisher, and Carrick C. Williams. Comparing the glance patterns of older versus younger experienced drivers: Scanning for hazards while approaching and entering the intersection. *Transportation research part F: traffic psychology and behavior*, 2013. 16: 104-116.
32. Pradhan, A. K., Hammel, K. R., DeRamus, R., Pollatsek, A., Noyce, D. A., & Fisher, D. L. Using eye movements to evaluate effects of driver age on risk perception in a driving simulator. *Human factors*, 2005. 47(4): 840-852.

33. Horswill, M. S., & McKenna, F. P. Drivers' hazard perception ability: Situation awareness on the road. In *A Cognitive Approach to Situation Awareness* (S. Banbury & S. Tremblay, eds), Aldershot, UK, Ashgate, 2004, pp.155-175.
34. Thomas, F. D., Rilea, S. L., Blomberg, R. D., Peck. R. C., & Korbelak, K. T. *Evaluation of the safety benefits of the risk awareness and perception training program for novice teen drivers*, Report No. DOT HS 812 235. Washington, DC: National Highway Traffic Safety Administration, 2016.
35. McKnight, A. J., & McKnight, A. S. Young novice drivers: careless or clueless? *Accident Analysis & Prevention*, 2003. 35(6): 921-925.
36. Mayhew, D. R. and Simpson, H. M. *The role of driving experience: Implications for training and licensing of new drivers*, Insurance Bureau of Canada, Toronto, Canada, 1995.
37. Crundall, D., Andrews, B., Van Loon, E., & Chapman, P. (2010). Commentary training improves responsiveness to hazards in a driving simulator. *Accident Analysis & Prevention*, 42(6), 2117-2124.
38. Pradhan, A., Masserang, K. M., Divekar, G., Reagan, I., Thomas, F. D., Blomberg, R., ... & Fisher, D. (2009). Attention maintenance in novice drivers: Assessment and training. In Proceedings of the *International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design* (Vol. 2009, p. 349). NIH Public Access.
39. Taylor, T. G., Masserang, K. M., Pradhan, A. K., Divekar, G., Samuel, S., Muttart, J. W., ... & Fisher, D. L. (2011). Long term effects of hazard anticipation training on novice drivers measured on the open road. In Proceedings of the... International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design (Vol. 2011, p. 187). NIH Public Access.
40. Pollatsek, A., Narayanaan, V., Pradhan, A., & Fisher, D. L. (2006). Using eye movements to evaluate a PC-based risk awareness and perception training program on a driving simulator. *Human Factors*, 48(3), 447-464.
41. Endsley, M. R. (1995). Measurement of situation awareness in dynamic systems. *Human factors*, 37(1), 65-84.
42. Urwyler, P., Gruber, N., Müri, R. M., Jäger, M., Bieri, R., Nyffeler, T., ... & Nef, T. (2015). Age-dependent visual exploration during simulated day-and night driving on a motorway: a cross-sectional study. *BMC geriatrics*, *15*(1), 18.
43. Crundall, D., Chapman, P., Trawley, S., Collins, L., Van Loon, E., Andrews, B., & Underwood, G. (2012). Some hazards are more attractive than others: Drivers of varying experience respond differently to different types of hazard. *Accident Analysis & Prevention*, 45, 600-609.
44. Chan, E., Pradhan, A. K., Pollatsek, A., Knodler, M. A., & Fisher, D. L. (2010). Are driving simulators effective tools for evaluating novice drivers' hazard anticipation, speed management, and attention maintenance skills? *Transportation research part F: traffic psychology and behaviour*, 13(5), 343-353.
45. Chapman, P. R., & Underwood, G. (1998). Visual search of driving situations: Danger and experience. *Perception*, 27(8), 951-964.
46. Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology,* 1993. 3(3): 203-220.

47. Classen, S., Bewernitz, M., & Shechtman, O. (2011). Driving simulator sickness: an evidence-based review of the literature. *American journal of occupational therapy*, *65*(2), 179-188.
48. Milleville-Pennel, I., & Charron, C. (2015). Do mental workload and presence experienced when driving a real car predispose drivers to simulator sickness? An exploratory study. *Accident Analysis & Prevention*, *74*, 192-202.
49. Brooks, J. O., Goodenough, R. R., Crisler, M. C., Klein, N. D., Alley, R. L., Koon, B. L., ... & Wills, R. F. (2010). Simulator sickness during driving simulation studies. *Accident Analysis & Prevention*, *42*(3), 788-796.
50. Realtime Technologies Catalog. Realtime Technologies Product Catalog. Retrieved from https://bit.ly/2RsFvfr. Accessed October 8, 2018.
51. Tobii VR Integration. Eye tracking Specifications. Retrieved from https://www.tobiipro.com/product-listing/vr-integration/#Specifications. Accessed October 8, 2018.
52. Morimoto, C. H., & Mimica, M. R. Eye gaze tracking techniques for interactive applications. *Computer vision and image understanding*, 2005. 98(1): 4-24.
53. Campbell, D., & Stanley, J. Experimental and quasi-experimental designs for research, 1963. Boston: Houghton Mifflin.
54. Williams, E. J. Experimental designs balanced for the estimation of residual effects of treatments. *Australian Journal of Chemistry*, 1949. 2(2): 149-168.
55. Stoner, H., Fisher, D., & Mollenhaurer, M. J. Simulator and Scenario Factors Influencing Simulator Sickness. In *Handbook of Driving Simulation for Engineering, Medicine and Psychology*, 2011, pp. 14-1:14-23. Boca Rator: CRC.
56. 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.
57. Muttart, J. W. Identifying hazard mitigation behaviors that lead to differences in the crash risk between experienced and novice drivers (Doctoral dissertation, University of Massachusetts Amherst), 2013.
58. Samuel, S., & Fisher, D. L. Evaluation of the minimum forward roadway glance duration. *Transportation Research Record: Journal of the Transportation Research Board*, 2015. 2518: 9-17.
59. Helland, A., Lydersen, S., Lervåg, L. E., Jenssen, G. D., Mørland, J., & Slørdal, L. Driving simulator sickness: Impact on driving performance, influence of blood alcohol concentration, and effect of repeated simulator exposures. *Accident Analysis & Prevention*, 2016. 94: 180-187.
60. Keshavarz, B., Ramkhalawansingh, R., Haycock, B., Shahab, S., & Campos, J. L. Comparing simulator sickness in younger and older adults during simulated driving under different multisensory conditions. *Transportation research part F: traffic psychology and behaviour*, 2018. 54, 47-62.
61. Fisher, D. L., Young, J., Zhang, L., Knodler, M., & Samuel, S. (2017). Accelerating Teen Driver Learning: Anywhere, Anytime Training. *Behavior Research Methods*, 37, 379-384.