Effects of using headset-delivered virtual reality in road safety research: A systematic review of empirical studies

Daniel VANKOV*, David JANKOVSZKY

1. Queensland University of Technology, Centre for Accident Research and Road Safety-Queensland, Kelvin Grove 4059, Queensland, Australia
2. Asociatia Pentru Tinerii si Studentii din Partium, Oradea, Romania

*Corresponding author, d.vankov@qut.edu.au
Received: 26 March 2021 Accepted: 8 May 2021

Citation: Daniel VANKOV, David JANKOVSZKY. Effects of using headset-delivered virtual reality in road safety research: A systematic review of empirical studies. Virtual Reality & Intelligent Hardware, 2021, 3(5): 351—368
DOI: 10.1016/j.vrih.2021.05.005

Abstract To reduce serious crashes, contemporary research leverages opportunities provided by technology. A potentially higher added value to reduce road trauma may be hidden in utilising emerging technologies, such as headset-delivered virtual reality (VR). However, there is no study to analyse the application of such VR in road safety research systematically. Using the PRISMA protocol, our study identified 39 papers presented at conferences or published in scholarly journals. In those sources, we found evidence of VR's applicability in studies involving different road users (drivers, pedestrians, cyclists and passengers). A number of articles were concerned with providing evidence around the potential adverse effects of VR, such as simulator sickness. Other work compared VR with conventional simulators. VR was also contributing to the emerging field of autonomous vehicles. However, few studies leveraged the opportunities that VR presents to positively influence the involved road users' behaviour. Based on our findings, we identified pathways for future research.

Keywords Empirical evidence; Headset; Road safety; Systematic review; Virtual reality

1 Introduction

In 2027, the virtual reality (VR) market is estimated to reach 62.10 billion USD, up from 10.32 billion USD in 2019[1]. Such growth represents a 21.6% compound annual growth rate (CAGR)[1].

VR is becoming a popular tool for researchers and practitioners[2]. For example, it is successfully applied to gamify education[3]. As a result, until 2024, the use of VR in the education sector has a forecasted CAGR of 59%[4], almost triple the overall CAGR.

The popularity of VR may be sought in its ability to stimulate simultaneously multiple senses, such as vision, hearing or even touch and sense of heat[3]. At the same time, it offers a safe training environment with minimised treatment risks[5]. VR can also review and analyse users' behaviour or deliver real-time performance feedback and cues on how to modify behaviour[5]. Researchers suggest that such information...
VR capabilities can be valuable for road safety research, which often tries to prevent life-threatening situations by influencing road users' behaviour. For example, driving while intoxicated negatively affects driving performance\cite{7}, increasing the risk of crashes\cite{8}. It is illegal to drive above certain intoxication levels, which raises ethical and legal considerations when the behaviour is researched\cite{9}, even when people drive intoxicated on a closed circuit\cite{10}. VR may present opportunities for a safer way to undertake such road safety research than other available technologies, bringing the experience closer to real-life experience with an unmatched level of immersion (Section 2).

In the current paper, we define VR as technology that delivers immersive experiences through a headset. Experience is considered immersive when a user can enter a computer-generated world of imagery and sounds\cite{11}. A full immersion would allow the user to use controllers to move around and manipulate objects in the VR environment\cite{11}.

Outside the research environment, VR is making its way in industry road safety interventions\cite{12}. Such interventions attract media interest, while prestigious funding enables their implementation across borders\cite{12}. At the same time, those interventions seem to miss out on generating new knowledge around the VR success in triggering behavioural changes or the lack of it\cite{12}.

One such VR intervention with the support of the European Union Erasmus+ Programme is foreseen in the framework of the future "Drugs and Alcohol use Reduction: Engage with young people" (DARE) project. To avoid pitfalls of previous projects, the DARE team seeks a better understanding of the available empirical evidence around using VR in road safety. To the best of the authors' knowledge, no systematic investigation of such evidence is available. Hence, their motivation to shed light on the subject through the current article.

The current study reviews the empirical evidence around using VR applications in road safety research. Such a review uncovers state of the art, which can help future research and industry interventions' design better meet road users' needs. Thus, we argue that this review is timely and necessary to explore how effective was VR when implemented as a road safety tool.

To shed light on the VR implementations' effectiveness, weexploited various intervention characteristics, such as samples and targeted behaviours. We aimed to reveal to interested practitioners and researchers relevant issues to consider when utilising VR in their future work. To achieve this aim, we sought to answer the research question: What VR application effects are reported in road safety research?

The answer to the research question is structured in three sections: Methodology, Results and Discussion. In the Methodology section, we explain our preparation for the undertaken work. Our motivation for providing this information is to make it easier for other researchers to replicate our study. In the Results section, we describe what we have found. Thus, we provide a guide for interested parties into the empirical evidence around the effects of VR in road safety research. In the Discussion section, we explore the implications stemming from our findings. Thus, in turn, we are able to suggest directions in which previous work can be extended in the future.

However, before we started answering the research question, we acknowledged that our impressions might lead to a traditional literature review bias\cite{13}. To increase the readers' confidence in our work and improve our findings' overall reliability, we decided to systematically review the literature to address any potential bias\cite{14}. For this purpose, we followed a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) study design\cite{14}.

The PRISMA design allowed us to focus our attention on specific aspects of the explored body of research. In our case, we were interested to learn whether VR triggered any behavioural change in the

352
studies’ participants. At the same time, we acknowledged that our adopted systematic design imposed other limitations to our work. For example, a traditional literature review can explore in more details the reviewed studies, discussing their method design. Such breadth of the analysis is hardly achievable in a systematic review due to the many studies that need to be analysed. The trade-offs between the strengths and the weakness of traditional and systematic reviews may have led some study features to fall outside the scope of our work. Nevertheless, interested readers can use the current work as a guide by locating studies with relevant findings and exploring additional study characteristics directly in the respective article. Overall, we focused on delivering behavioural change in a safe environment, and, in turn, we introduce related work.

2 Related work

One of the biggest challenges in road safety research is to investigate and influence risky behaviour in safety. Technologies, such as driving simulators, smartphones or VR, offer such safer environments. However, each technology comes with its strengths and limitations.

2.1 Driving simulators

Researchers focus on developing and improving driving simulators to offer a safe environment to conduct research, education, and training\(^\text{[16]}\). Such simulators find numerous applications in developing road and vehicle systems, understanding how road users interact with them and educating them to adopt safer behaviours\(^\text{[16]}\). In addition to removing associated risks, the convenience of data collection and the driving scenarios’ controllability and reproducibility are amongst the many advantages offered by driving simulators\(^\text{[17]}\). Nevertheless, problems, such as limited fidelity, questionable simulation validity or simulator sickness, are known issues for the research community\(^\text{[17]}\). For example, the fidelity of a simulation may affect the research outcomes and, in extreme cases, instil unrealistic self-perceptions of competence or abilities in the users\(^\text{[17]}\). If driving performance is improved in the simulator, this outcome does not guarantee that the improved skills will be transferred in the real world, i.e., the validity of the simulation can be questioned\(^\text{[17]}\). Lastly, regardless of the simulation validity level, simulators are shown to make some people sick, undermining any delivered training\(^\text{[17]}\).

Some of the identified issues can be tackled through technology, reproducing an environment close to a real-world experience\(^\text{[18]}\). Nevertheless, expensive equipment is generally hard to obtain. Thus, different parameters characterise the multitude of utilised simulators, ranging from hobby-grade to physical vehicles. Regardless of the simulator grade, the environment is typically 2D or a non-immersive 3D.

2.2 Smartphones

The advancement of smartphones offered an upgrade of the simulator research capabilities at a relatively low cost. For example, combining an existing simulator and a smartphone allowed re-engaging the drivers in the driving task when speed changes were necessary due to speed limits\(^\text{[19]}\), alerting for traffic hazards (headway, speed, and acceleration) through sound, visual, and voice warnings\(^\text{[20]}\), or calculating a mood-fatigue profile of the driver in real-time\(^\text{[21]}\). Nevertheless, smartphones are known to introduce risks amongst road users\(^\text{[22]}\).

2.3 Virtual reality

Like smartphones, VR technology is becoming increasingly available\(^\text{[23]}\) and may easily find its way into
road safety prevention efforts. VR seems to address issues related to traditional simulators and smartphones simultaneously. For example, it may offer an immersive 3D experience at a low cost. At the same time, it can deliver similar benefits to a smartphone, such as providing driving cues, while there is no evidence of distracting the user. Those advantages may help the technology's wider adoption, making it possible to replace older and less efficient simulators in community interventions, such as in Vankov and Kaufmann[24].

The potential VR benefits are not restricted to the road safety field. VR technology has already been applied in education[3] or health research. For example, Rizzo and Koenig[5] provide evidence of its potential to treat clinical impairments. At the same time, there is more limited evidence of VR effectiveness in clinically healthy populations, such as the typical road users. In some applications aimed at clinically healthy populations, VR has shown its potential to trigger a change in health behaviour[25]. Several empathy-targeted studies support VR use[26-28]. For example, VR helped significantly increase the willingness to intervene in bullying events and the feelings of belonging among school students (n=118)[27]. Other VR applications motivated users to self-manage their diabetes better[29], to reduce their use of paper[30], or not to text when crossing a street[31].

In conclusion, VR attracts the interest of researchers from various fields. This interest may be due to the novelty of the technology or to the capabilities that it offers. However, to our best knowledge, a study to systematically review the utilisation of VR in road safety research is still missing. To bridge the identified gap, this study employed a PRISMA protocol[15] to examine the latest evidence, going beyond the separate studies' findings to provide insights for future research.

3 Methodology

Following a PRISMA design, this study reviewed the available literature regarding VR deployment and its effects on road users' behaviour and safety. Our review followed linear procedural phases[15]. The review protocol steps were as follows: Development of the research question (Section 1), Identification of search databases, Definition of scope, inclusion, and exclusion criteria, Definition of a search term, A systematic search for information, Screening and selection of studies, Review of selected articles and Summarising of findings.

3.1 Search databases

Relevant papers were identified through searches in Scopus (www.scopus.com), Web of Science (www.webofknowledge.com) and ProQuest (www.proquest.com). All three databases are widely used for PRISMA-guided systematic reviews in the context of road safety[32-34]. The searches were last updated on March 5, 2021.

3.2 Literature search criteria

The focus of our search was literature on immersive headset-delivered VR as defined in Section 1. Projection-based VR, such as Cave Automatic Virtual Environment or CAVE, was not considered within the scope of the current investigation. In addition, only papers written in English were considered for inclusion in the review. The investigation focused on the actual and potential application and utility of VR in road safety research and practice. Papers were excluded when they focused on augmented reality, technical ITS solutions, theoretical discussions, and concepts' development, or when they were not relevant for road safety.
3.3 Search term

The search term deployed in Scopus, Web of Science and ProQuest was the same. We searched all fields of the databases' records for (road OR driver OR traffic) AND safety AND "virtual reality" AND (headset OR "HTC Vive" OR "Oculus Rift" OR "Oculus Quest" OR valve OR "PlayStation VR" OR "Google Cardboard" OR "Samsung Gear" OR "HP Reverb" OR "Pansonite VR" OR "Veer Falcon" OR retrak). The most common headsets' brands were specified because otherwise, the returned search result contained many regular driving simulator studies, referring to themselves as VR ones. Such reference might have been technically correct, but it does not reflect our adopted VR definition.

We acknowledge that our list of VR headsets' brands is not exhaustive as new brands constantly appear in the market. We intended to include the most common brands because they might be mentioned in a text without the accompanying word "headset". For less popular brands, we believe that for clarity, the brand name will likely be accompanied by the word "headset". Thus, we have included "headset" in the search term separately. With this solution, we believe our search term to be as inclusive of all brands as possible.

4 Results

A total of 838 records were identified through searching Scopus, Web of Science and ProQuest (Figure 1). Their titles were screened for relevance. This step excluded 762 titles, leaving 76 documents for abstract review. After the abstract screening, 21 were excluded from further analysis. The full texts of 55 papers remained to be assessed for eligibility. Six duplicates were excluded, which left 49 full texts to be processed. The full texts were downloaded for review. Ten of the studies were found not to comply with our inclusion criteria. Thus, 39 papers were retained to be included in the qualitative synthesis.

We organised our results in two main subsections, a general overview and studies' characteristics. The general overview provides descriptive information such as document type, year of publication and type of headset. The studies characteristics subsection explores our final selection of papers in more detail. It includes elements of research interest, such as the type and number of participants, the measures taken, and summaries of the studies' findings. Those elements were identified in line with the motivation behind conducting the current systematic review (Section 1). By focusing on a carefully chosen number of variables, we intended to provide researchers and practitioners with a clearer and simple comparative map to navigate and gauge the design and implementation of their future VR interventions.

4.1 General overview

Road safety VR publications started appearing in 2015 (Figure 2). Despite the small number of publications in 2016, the trend was upward. The most number of works (n=10) was published in 2020. However, for the first two months of 2021, there are already three publications. This comparatively high number hinted that an increase in the output might be expected.

Since we looked only at peer-reviewed work, two types of documents were included in our final analysis: journal articles and conference papers. The number of journal articles was 21, while the number of papers was 18. This finding indicates that authors seem to publish their works in scholarly journals and conference proceedings almost equally.

Similar parity was observed with regards to the VR headsets utilised by the studies. HTC Vive was the most popular, used in 19 studies. Oculus Rift was identified in 15 studies and came in second. One study used both. From the remaining four studies, one did not specify the headset, one used FOVE and two used Google technology.
4.2 Studies' characteristics

The in-depth analysis of the studies revealed the variability of the undertaken work (Table 1). For example, Ropelato, Ropelato et al.\textsuperscript{[35]} focused on simulator sickness only, while Gaibler et al.\textsuperscript{[36]} used a keyboard to control the software. Regarding the involved road users, most of the studies focused on drivers ($n=23$). Pedestrians came in second ($n=10$), followed by cyclists ($n=4$) and passengers ($n=2$).

Two comparatively large groups of studies were identified with regards to their focus. The first one (8 studies) investigated interactions with autonomous vehicles (AV)\textsuperscript{[37–42]}, including research on controversial topics, such as decision making and moral dilemmas\textsuperscript{[43,44]}. The second group (8 studies) was directed towards identifying whether the technology is user-friendly\textsuperscript{[45–52]}. This large amount of work indicated that VR applications can still be considered an emerging technology, i.e., they are at a very early stage of their development. Further, we focused on three main characteristics of our study selection: research samples, measures taken and reported results (Table 1).
### Table 1  Virtual reality in the road safety literature (arranged by first author’s last name)

<table>
<thead>
<tr>
<th>Authors</th>
<th>Type of participants</th>
<th>Number of participants</th>
<th>Measures taken</th>
<th>Summary of findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lang et al.[32]</td>
<td>Drivers: 1 to 20 years of driving experience</td>
<td>50</td>
<td>Emergency response time, a safe-driving score for crossing pedestrian or signalling on turn events.</td>
<td>The VR intervention group showed both better response time improvement and higher safety scores.</td>
</tr>
<tr>
<td>Ropelato et al.[33]</td>
<td>Drivers: 5 females and 12 males, mean age 29.5</td>
<td>17</td>
<td>Simulator sickness</td>
<td>A slight increase in discomfort was observed with 4 participants aborting the VR during the 15 minutes of simulation.</td>
</tr>
<tr>
<td>Fuest et al.[37]</td>
<td>Pedestrians: 47 males and 24 females</td>
<td>71</td>
<td>Ability to recognise the vehicle’s intention</td>
<td>The VR setup could not provide enough evidence as to absolute values to form any conclusions.</td>
</tr>
<tr>
<td>Deb et al.[38]</td>
<td>Pedestrians: 12 males (aged 20–34) and 12 females (aged 19–36)</td>
<td>24</td>
<td>While crossing: the gap from the approaching AV, frequency of stopping and waiting time</td>
<td>By observing vehicles' legacy behaviour, the participants responded similarly to a human-driven car and AV.</td>
</tr>
<tr>
<td>Faulhaber et al.[39]</td>
<td>Drivers: 62 females and 127 male, age range 18 to 67</td>
<td>189</td>
<td>Moral decisions</td>
<td>While driving, the participants aimed at saving as many virtual traffic participants as possible, acting according to a principle for the quantitative greater good.</td>
</tr>
<tr>
<td>Helgath et al.[40]</td>
<td>Passengers: 6 males and 11 females, age range 20 to 29</td>
<td>17</td>
<td>User experience and user acceptance when handling AV</td>
<td>In AV mode, observed significantly greater levels of safety, performance expectations and social influence. No evidence of lower user experience and user acceptance when the level of autonomy is increased.</td>
</tr>
<tr>
<td>Ferrier-Barbut et al.[41]</td>
<td>Passengers: 15 males and 4 females</td>
<td>19</td>
<td>Feelings when an AV passenger</td>
<td>A comfort zone was observed. When a pedestrian crosses this zone, the AV passenger feels discomfort.</td>
</tr>
<tr>
<td>Rangelova et al.[42]</td>
<td>Drivers: Age range 30 to 53</td>
<td>9</td>
<td>Simulator sickness</td>
<td>Regardless of whether in combination with a motion platform or not, simulating AV in VR did not result in substantial simulator sickness.</td>
</tr>
<tr>
<td>Leon René et al.[43]</td>
<td>Drivers: 43 females and 42 males, mean age 23</td>
<td>85</td>
<td>Decision to save an avatar</td>
<td>Gender and age were the consideration factors for a decision of preference to be saved, e.g., 90% of children saved and 60% of women saved.</td>
</tr>
<tr>
<td>Süttdfeld et al.[44]</td>
<td>Drivers: 76 males and 29 females, age range 18 and 60</td>
<td>105</td>
<td>Ethical decisions</td>
<td>When crashes are unavoidable, human moral decisions can be approximated by value-of-life models.</td>
</tr>
<tr>
<td>Sakhare et al.[45]</td>
<td>Cyclists: 20 younger (m. age 26) and 20 older (m. age 64)</td>
<td>40</td>
<td>Physical exertion, adverse effects, and presence</td>
<td>When cycling in the VR environment, most participants enjoyed the spatial navigation, showing high arousal and low stress levels.</td>
</tr>
<tr>
<td>Wang et al.[46]</td>
<td>Cyclists: 2 females and 2 males</td>
<td>4</td>
<td>Usability</td>
<td>The users could not understand the cycling scenarios well.</td>
</tr>
<tr>
<td>Deb et al.[47]</td>
<td>Pedestrians: 10 males (aged 21–39) and 11 females (aged 22–50)</td>
<td>21</td>
<td>Simulator sickness, system usability, immersion, walking speed and crossing</td>
<td>The VR was confirmed as capable of capturing pedestrian behaviour on a crosswalk.</td>
</tr>
<tr>
<td>Walch et al.[48]</td>
<td>Drivers: 5 females and 14 males, 1 unreported</td>
<td>20</td>
<td>Simulator sickness, immersion, presence</td>
<td>The VR did not appear to be better than flat screens.</td>
</tr>
<tr>
<td>Cao et al.[49]</td>
<td>Drivers: 12 females and 35 males, age range 18 to 38</td>
<td>47</td>
<td>Time gameplay, simulator sickness</td>
<td>The VR setup did not deliver significantly better game engagement and game performance than a regular desktop LCD. However, it was associated with an increase in simulator sickness.</td>
</tr>
<tr>
<td>Authors</td>
<td>Type of participants</td>
<td>Number of participants</td>
<td>Measures taken</td>
<td>Summary of findings</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---------------------------------------------</td>
<td>------------------------</td>
<td>-------------------------------------------</td>
<td>--------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Mortimer et al. ([31])</td>
<td>Drivers: 64 years old and above</td>
<td>56</td>
<td>Ease of use</td>
<td>Evidence for a steady increase in the VR ease of use.</td>
</tr>
<tr>
<td>Xuan and Albert ([32])</td>
<td>Pedestrians: 11 to 40</td>
<td>15</td>
<td>Immersion, ease-of-use</td>
<td>In the VR, the participants showed greater immersion and familiarity with the correct methods for crossing roads.</td>
</tr>
<tr>
<td>Gašiorek et al. ([33])</td>
<td>Drivers: 15 males, aged 20 to 29, and 15 males, aged 50 to 65</td>
<td>30</td>
<td>Simulator sickness, immersion, usability</td>
<td>VR delivered greater realism than projection screens, with 43.3% of participants voting for Oculus Rift, 26.7% for HTC Vive, and 30% for projection screens.</td>
</tr>
<tr>
<td>Agrawal et al. ([34])</td>
<td>Drivers: aged 18 to 25 years</td>
<td>24</td>
<td>Eye movements, hazards anticipation</td>
<td>The young drivers improved their ability to detect threats. Compared to the control groups, V-RAPT users anticipated significantly more latent hazards.</td>
</tr>
<tr>
<td>Agrawal et al. ([35])</td>
<td>Drivers: aged 18 to 25 years</td>
<td>36</td>
<td>Hazard mitigation, hazard anticipation</td>
<td>The V-RAPT-trained drivers were significantly better at anticipating and mitigating potential hazards than the regular RAPT or placebo trained drivers.</td>
</tr>
<tr>
<td>Bogacz et al. ([36])</td>
<td>Cyclists: 18 males and 28 females (mean age 30.7 years)</td>
<td>46</td>
<td>Speed, acceleration, braking and head movements</td>
<td>The VR was shown to trigger behavioural patterns as in real-world driving.</td>
</tr>
<tr>
<td>Bozkir et al. ([37])</td>
<td>Drivers: 4 females and 12 males, age range from 25 to 50</td>
<td>16</td>
<td>Changes of pupil diameter, input on breaks, input on the accelerator, crossing distances between vehicles and pedestrians</td>
<td>The VR experience was characterised by low resolution and narrow field of view. Nevertheless, it was shown that visual cues in critical situations could stimulate the drivers’ attention.</td>
</tr>
<tr>
<td>Brown et al. ([38])</td>
<td>Drivers: 57% male and 43% female, age range 18 and 22</td>
<td>23</td>
<td>Following distances, comfort with AV, number of crashes</td>
<td>Following distances affected comfort levels. Introducing AVs is not likely to entirely eliminate crashes due to other human drivers.</td>
</tr>
<tr>
<td>Chen et al. ([39])</td>
<td>Drivers: 15 females and 11 males; 14, age range 18 to 35, and 12 age range 65 to 75</td>
<td>26</td>
<td>Target detection</td>
<td>Younger drivers were twice more successful in target detection in less time than older ones. They also rotated their trunks on average in two-times greater radius than the older drivers.</td>
</tr>
<tr>
<td>Cherix et al. ([40])</td>
<td>Pedestrians: 11 males and 4 females, age range 9 to 18</td>
<td>15</td>
<td>Waiting time before crossing</td>
<td>After a series of sessions, a learning effect was observed. The VR was associated with good acceptability.</td>
</tr>
<tr>
<td>Cutello et al. ([41])</td>
<td>Drivers: age range 18 to 25, mean age 20.97</td>
<td>146</td>
<td>Risky driving behaviours engagement</td>
<td>The VR was perceived as immersive, leading to better results in decreasing self-reported risky driving behaviours after watching positively framed films.</td>
</tr>
<tr>
<td>Ojados Gonzalez et al. ([42])</td>
<td>Drivers: 93 males and 34 females, aged range 16 to 56</td>
<td>127</td>
<td>Self-reported perception of the risk and safety</td>
<td>Participants reported positive experience using the tractor driving simulator to help them drive safer, but they would need more training.</td>
</tr>
<tr>
<td>González-Ortega et al. ([43])</td>
<td>Drivers: Age range 20 to 55</td>
<td>12</td>
<td>Gaze estimation</td>
<td>Oculus Rift performed better than Kinect due to more precise head orientation and larger distance between vision regions.</td>
</tr>
</tbody>
</table>

**Summary of findings:**
- Evidence for a steady increase in the VR ease of use.
- In the VR, the participants showed greater immersion and familiarity with the correct methods for crossing roads.
- VR delivered greater realism than projection screens, with 43.3% of participants voting for Oculus Rift, 26.7% for HTC Vive, and 30% for projection screens.
- The young drivers improved their ability to detect threats. Compared to the control groups, V-RAPT users anticipated significantly more latent hazards.
- The V-RAPT-trained drivers were significantly better at anticipating and mitigating potential hazards than the regular RAPT or placebo trained drivers.
- The VR was shown to trigger behavioural patterns as in real-world driving.
- The VR experience was characterised by low resolution and narrow field of view. Nevertheless, it was shown that visual cues in critical situations could stimulate the drivers’ attention.
- Following distances affected comfort levels. Introducing AVs is not likely to entirely eliminate crashes due to other human drivers.
- Younger drivers were twice more successful in target detection in less time than older ones. They also rotated their trunks on average in two-times greater radius than the older drivers.
- After a series of sessions, a learning effect was observed. The VR was associated with good acceptability.
- The VR was perceived as immersive, leading to better results in decreasing self-reported risky driving behaviours after watching positively framed films.
- Participants reported positive experience using the tractor driving simulator to help them drive safer, but they would need more training.
- Oculus Rift performed better than Kinect due to more precise head orientation and larger distance between vision regions.
Some studies implied the importance of experience rather than any other demographic characteristic as in Agrawal et al.,

46, 47

When those two studies are excluded, the mean sample size is 55.

4.2.1 Research samples

As shown in Table 1, the reported numbers of involved participants vary significantly between the studies, ranging from 45 to 1900. When those two studies are excluded (as outliers), the mean sample size is 55.

Potentially because the VR technology is still in its early days of application, a common impression from the studies is that their samples seem convenient, i.e., to test the technology, as in Wang et al.,

46, Walch et al.,

48, or Orfila et al.,

88, rather than purposefully recruited to understand its impact, as in Agrawal et al.,

84. Some studies implied the importance of experience rather than any other demographic characteristic.

Nevertheless, studies involved participants of all ages, e.g., children under 10,

67, children between 9 and 18,

89, young people 18 to 25,

54, adults from 25 to 50,

86, older adults from 50 to 65,

52, and elderly above 65.
65[60]. Most studies were also non-discriminative regarding gender, involving both males and females[40,63,65].

4.2.2 Measures taken

The studies' variability regarding their samples continued in what was measured within them and, thus, in their focus. Some studies focused on self-reports, collecting data on risk and safety perception[61], simulator sickness[51] or VR simulation user-friendliness[51]. Other studies focused on the participants' driving, looking into their hazards anticipation[51], target detection[54] and fuel consumption[60]. The third type of studies collected objective data, such as gaze estimation[62]. A large number of studies (n=8) focused on addressing one of the known simulator issues[17], such as simulator sickness[42,60,70].

4.2.3 Reported results

Most studies in our sample used VR as a one-off measurement tool. For example, similar to Agrawal et al. [51], Chen et al. [58] also focused on detection. However, in their study, the VR was used to compare younger and older participants rather than produce any effect on them. Several studies measured pedestrians' travel behaviours, such as walking time and gap to a vehicle[37,38,59], while others focused on understanding the VR experience of cyclists[45,46,63].

Although increasing those users' safety should be one of the main focuses of road safety research, not many of our sample studies focused on that. Two papers[43,54] reported significant improvements in anticipating and mitigating potential hazards. One study[36] showed that visual cues could stimulate the drivers' attention in critical situations. Another investigation[12] reported a change of attitude towards safer driving without presenting statistical information to substantiate the claim. Similarly, Cutello et al. reported decreasing self-reported risky driving behaviours after watching positively framed films[60]. Overall, we found little evidence about VR delivering behavioural change.

A challenge for the meaningful interpretation of the reported results was the missing discussion around the strengths and limitations of the implemented VR interventions in a significant proportion of the studies. In total, 13 studies discussed the limitations of their research. The reported limitations were technical deficiencies in the VR simulations[37,39,54,57,63,73], problems with data reliability and comparability[23,37,57,60], participation discontinuation due to simulator sickness[45,60], and small samples and lack of age or gender balance[42,54,60,67,73]. Only two studies reported on their strengths. Those were assessing behaviour through immersive VR[53,73], a large sample size[73] and random assignment[63].

5 Discussion

This article presented a systematic literature review of empirical applications of emerging VR systems in road safety research. We intended to provide a comprehensive, in-depth understanding of the state of art in the field. Our investigation revealed a growing interest in VR implementation since 2015. It is interesting to underline that in their work, researchers do not opt for the cheapest options, such as Google Cardboard, but rather prefer higher-end devices, such as HTC Vive and Oculus Rift. Regardless of the utilised headset, there are multiple avenues for researchers to expand on the state of the art. Thus, following this Discussion section, we suggest some future research directions.

5.1 Research samples

Our data revealed a mean sample size of 55 when outliers are excluded. This number is considerably higher than similar metrics reported in other systematic reviews of road safety literature. For example, in mobile phone distraction studies, the respective number is 30 or 40, depending on the analysed period[22].
The difference may be because there are both simulator and naturalistic studies in mobile phone research\textsuperscript{[22]}. Naturalistic studies are generally more challenging to implement and, thus, may involve fewer participants.

Simulator studies are likely to continue to be more common due to ethical and safety considerations\textsuperscript{[16]}. Nevertheless, their known fidelity, validity and sickness problems\textsuperscript{[17]} seem to persist in VR studies. For example, several studies reported imperfections of the offered VR experience\textsuperscript{[37,39,54,57,61,73]}. Such imperfections are likely to misrepresent users' abilities, regardless of whether positively or negatively, which questions the fidelity of the simulation. Other studies reported data-related issues\textsuperscript{[23,37,57,60]}. If the collected data is not of sufficient quality, then the validity of the simulation may be considered compromised. And last, simulator sickness was reported as a limitation in some studies, which undermined the delivered intervention\textsuperscript{[45,66]}

Another reported problem was related to the sample composition. For example, despite that many studies involved both males and females\textsuperscript{[40,41,61]}, it seemed that gender diversity was not a concern for all researchers. Gender diversity, i.e., involvement of both males and females, was reported in 22 studies. At the same time, only nine of them had balanced samples. Furthermore, several studies reported sample balance as a limitation\textsuperscript{[42,54,60,67,71]}. With such recognition, we can expect sample-related gaps to be better addressed in future research.

Regardless of the encountered problems, the reviewed articles revealed VR road safety research to be age- and gender-inclusive. Such broad applicability of the VR technology suggest a potential for diverse impact in future research and potentially in implementations outside the research domain, such as community interventions\textsuperscript{[29]}

\subsection{5.2 Measures taken}

The variability of taken measures made meaningless any attempt to categorise the review sample according to this characteristic. We observed risk and safety perception\textsuperscript{[61]}, simulator sickness\textsuperscript{[15]}, simulation user-friendliness\textsuperscript{[51]}, hazards anticipation\textsuperscript{[53]}, target detection\textsuperscript{[59]}, fuel consumption\textsuperscript{[68]}, and gaze estimation\textsuperscript{[62]}

Regardless of the measures, what variables should be measured in future research can be defined based on theories, such as the Theory of Planned Behaviour\textsuperscript{[59]} or the Social Cognitive Theory\textsuperscript{[75]}. Such an approach seemed to be underutilised in the reviewed studies, although it is capable of helping the researchers to understand better the impact of their interventions\textsuperscript{[12]}. This approach may also assist developers in designing better VR scenarios. Such theory-grounded scenarios may, in turn, help researchers investigate how VR modifies the road users' behaviour and how it motivates their decisions, such as who to sacrifice when a crash is unavoidable\textsuperscript{[43,44]}. Those deeper investigations can produce reliable results in a more efficient and safer simulated environment.

\subsection{5.3 Reported results}

As already mentioned earlier, we found little evidence about VR delivering behavioural change. Some researchers tried to understand VR from a user-experience perspective\textsuperscript{[40,47,50]}. Others tried to prove that VR delivers realism\textsuperscript{[42,66]}. A third group compared VR with older simulator types\textsuperscript{[63,65,70]}. Rather than employing it as a behavioural change tool. Those findings suggested that research on VR as a behavioural change tool might be at its very early stages.

This early stage made it hard to draw overall conclusions or to support findings across VR studies.
Nevertheless, it opened up numerous possibilities for research beyond the current stage, with VR simulations being applied to reduce risky behaviours, such as driving intoxicated, crossing distracted or speeding. VR has already been applied in other educational contexts\cite{3}, which supports expanding its use in road safety educational programs.

\section*{5.4 Limitations}

Our findings should be considered in light of some limitations. First, our investigation was limited to peer-reviewed articles and conference papers. Thus, grey literature was not considered. Second, non-English documents were excluded. This exclusion means that studies that are published in other languages were not included in our analysis. Third, we searched three databases that we believe to be comprehensive and widely used. However, there are other research databases, and by not using them, we potentially could have missed relevant studies.

In our work, we focused on whether VR triggered a behavioural change. Such a focus limited our discussion by not providing details on other study characteristics, such as method design. Nevertheless, our work can serve as a guide to interested readers. By easily locating relevant findings through our work, interested researchers can explore additional characteristics of interest directly in the respective article.

A final limitation of our systematic review is that we considered only studies that deliver VR through a headset. By limiting the scope, we did not look at other available tools, such as CAVE, which encompass a broader definition of VR.

\section*{6 Future research}

It is still to be seen whether VR will have a different impact when controlling for demographics (e.g., age, gender and experience). Thus, there is research potential in the area of behavioural change for participants with different demographic characteristics. For example, road safety behavioural models can be calibrated by measuring behavioural adaptation levels of males and females. Those models can help measure the VR technology effect on any of the encountered road user types (drivers, passengers, pedestrians and cyclists). Thus far, our review suggested that the technology might be capable of obtaining useful measurements in the safe environment of VR.

Creating safer VR environments for future research\cite{5} can be particularly valuable when investigating high-risk behaviours, such as intoxicated driving\cite{10}. Like some of the investigated VR applications, road users' data seemed easy to be collected while the participants experienced VR scenarios\cite{67,69,71}.

However, such data would deliver the highest benefit if compared at several time points of the experience\cite{5}. To enable such comparisons, VR developers can design similar challenges throughout the scenarios. For example, a participant may need to break on a crossroad in sunny weather and during an emergency on a frozen road. Such tasks can be further complicated. For example, after a moral dilemma is presented about who to sacrifice in a situation where a crash cannot be avoided\cite{43,44}, the same dilemma can be presented in intoxicated mode to assess the additional damage, caused by reduced abilities to react\cite{5}.

A variety of quantitative measurements, such as speed, brake reaction time or following distance, can help in-depth evaluate the VR technology, ultimately leading to improvements in the road users' performance. Such improvements may be further facilitated by the timely provision of meaningful and informative feedback\cite{5}. In turn, this feedback can potentially influence the participants' attitudes and norms, triggering the desired behaviour change\cite{76}.
7 Conclusion

In cases of awareness-raising and road safety education programmes, researchers can provide essential insights into understanding the capabilities of emerging technologies that can promote safer road behaviour[^6]. VR can be considered such technology which is becoming increasingly available[^23]. Thus, understanding its current potential to improve road safety was due for an in-depth investigation.

This study described the VR state of the art in road safety. To our knowledge, this is the first systematic attempt to synthesising the relevant literature. We reviewed a total of 39 documents under the PRISMA guidelines. The literature suggested the users' sense of realism and immersion as the most critical VR value. The results revealed the empirical evidence around the current VR applications, which are simultaneously controversial and full of hope.

Overall, the study results unveiled valuable insights related to the type and number of studies' participants, measures taken, and findings. At the same time, we acknowledge that it is not easy to draw parallels beyond those parameters due to the numerous VR variations found in research. Nevertheless, by summarizing the current state of scientific progress, we provided suggestions not only for future research but potentially for non-academic road safety practitioners.

Declaration of competing interest

We declare that we have no conflict of interest.

References

Can virtual reality exposure therapy gains be generalized to...


37 Fuest T, Schmidt E, Bengler K. Comparison of methods to evaluate the influence of an automated vehicle’s driving behavior on pedestrians: wizard of oz, virtual reality, and video. Information, 2020, 11(6): 291 DOI:10.3390/info11060291


DOI:10.1109/ivs.2018.8500648

DOI:10.1007/978-3-319-91581-4_12

DOI: 10.1371/journal.pone.0223108

DOI:10.3389/fnbeh.2017.00122

DOI:10.3389/fnagi.2019.00218

DOI:10.1145/2957265.2961846

DOI:10.1016/j.apergo.2017.03.007

DOI:10.1145/3027063.3053202

49 Cao S, Nandakumar K, Babu R, Thompson B. Game play in virtual reality driving simulation involving head-mounted display and comparison to desktop display. Virtual Reality, 2020, 24(3): 503–513
DOI:10.1007/s10055-019-00412-x

DOI:10.1109/sysosc.2019.8753836

DOI:10.1145/3328243.3328261

DOI:10.1007/978-3-030-27687-4_14

54 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, 2672(33): 20–30 DOI:10.1177/0361198118758311


64 Madigan R, Romano R. Does the use of a head mounted display increase the success of risk awareness and perception training (RAPT) for drivers? Applied Ergonomics, 2020, 85, 103076 DOI:10.1016/j.apergo.2020.103076


DOI:10.1136/injuryprev-2014-041357

DOI:10.1109/ivs.2015.7225781

Pai Mangalore G, Ebadi Y, Samuel S, Knodler M A, Fisher D L. The promise of virtual reality headsets: can they be used to measure accurately drivers' hazard anticipation performance? Transportation Research Record: Journal of the Transportation Research Board, 2019, 2673(10): 455–464
DOI:10.1177/0361198119847612

DOI:10.1016/j.aap.2021.106004

DOI:10.1177/0018720820987446


DOI:10.1136/injuryprev-2019-043268

DOI:10.1016/0749-5978(91)90020-t
