



Road user hazard perception tests: A systematic review of current methodologies



Caroline Moran^a, Joanne M. Bennett^{a,*}, Prasannah Prabhakaran^b

^a School of Behavioural and Health Sciences, Australian Catholic University, Strathfield, NSW, Australia

^b Research Centre for Integrated Transport Innovation, School of Civil and Environmental Engineering, University of New South Wales, Sydney, NSW, Australia

ARTICLE INFO

Keywords:

Hazard perception test
Methodology
Road user
Graduated Driver Licensing (GDL) young drivers
Novice drivers
Older drivers

ABSTRACT

Background: Poor hazard perception, or the ability to anticipate potentially dangerous road and traffic situations, has been linked to an increased crash risk. Novice and younger road users are typically poorer at hazard perception than experienced and older road users. Road traffic authorities have recognised the importance of hazard perception skills, with the inclusion of a hazard perception test in most Graduated Driver Licensing (GDL) systems.

Objectives: This review synthesises studies of hazard perception tests in order to determine best practice methodologies that discriminate between novice/younger and experienced/older road users.

Data sources: Published studies available on PsychInfo, Scopus and Medline as at April 2018 were included in the review. Studies included a hazard perception test methodology and compared non-clinical populations of road users (car drivers, motorcyclists, bicyclists and pedestrians), based on age and experience, or compared methodologies.

Results: 49 studies met the inclusion criteria. There was a high degree of heterogeneity in the studies. However all methodologies – video, static image, simulator and real-world test-drive were able to discriminate road user groups categorised by age and/or experience, on at least one measure of hazard perception.

Conclusions: Whilst there was a high level of heterogeneity of studies, video methodology utilising temporal responses (e.g. press a button when detecting the potential hazard) are a consistent measure of hazard perception across road user groups, whereas spatial measures (e.g. locate potential hazard in the scenario) were inconsistent. Staged footage was found to discriminate as well as unstaged footage, with static images also adding valuable information on hazard perception. There were considerable inconsistencies in the categorising of participants based on age and experience, limited application of theoretical frameworks, and a considerable lack of detail regarding post hoc amendments of hazardous scenarios. This research can guide further developments in hazard perception testing that may improve driver licensing and outcomes for road users.

1. Introduction

Road traffic crashes result in an annual global fatality rate of 1.35 million people, of which 54% are vulnerable road users such as motorcyclists, bicyclists and pedestrians (World Health Organisation, 2018). Younger road users are over-represented in crash statistics, with traffic crashes the leading cause of death for 15–24 year olds (International Traffic Forum, 2018). For younger drivers, the risk of injury or death rapidly declines in the months after licensing, suggesting that both age and experience are contributing factors to crash risk (Mayhew et al., 2003). Younger drivers experience cognitive changes associated with brain maturation as they develop from late adolescence to adulthood, at the same time as they transition from a

novice to more experienced driver through practice on the roads (Foy et al., 2016). There is difficulty in disentangling the impact of brain maturation and experience on the roads due to their confounding effect, however time facilitates both in younger novice drivers.

One of the driving skills that is consistently linked to crash risk, is the ability to anticipate potentially dangerous road and traffic situations, a skill known as hazard perception (Horswill and McKenna, 2004). Hazard perception improves with both age and experience, and while brain maturation cannot be accelerated, experience can be advanced through hazard perception training (Horswill, 2016a). One way of encouraging training in hazard perception is to implement a test that sets a standard for hazard perception that is above the normal skill level of young and/or novice drivers (Horswill, 2016b). This has been

* Corresponding author.

E-mail address: joanne.bennett@acu.edu.au (J.M. Bennett).

<https://doi.org/10.1016/j.aap.2019.05.021>

Received 6 February 2019; Received in revised form 7 May 2019; Accepted 23 May 2019

Available online 07 June 2019

0001-4575/ © 2019 Elsevier Ltd. All rights reserved.

recognised by road traffic authorities, such that hazard perception testing has become part of the Graduated Driver Licensing (GDL) systems in countries such as Australia, the UK and the Netherlands (Australian Government, 2018; Bates et al., 2014; Drive Right Netherlands, 2019). Studies have shown that those who fail the hazard perception test twice are more likely to be involved in traffic crashes; and for those that pass on the first attempt there is a 17% reduction in at-fault injury crashes when tested during the provisional licensing period (Boufous et al., 2011; Horswill et al., 2015; Senserrick and Williams, 2015; Thomas et al., 2016). This suggests a degree of predictive validity of the hazard perception test. Given the importance of hazard perception to driver safety, a range of hazard perception test methodologies have been developed to assess a driver's competence in this domain. This has driven a large body of research into which methodologies best capture hazard perception ability in drivers, as well as other road users groups.

1.1. Development of hazard perception testing

An early hazard perception test used a driving simulator with films of traffic incidences, and found novice drivers were slower at detecting hazards than more experienced drivers (Quimby and Watts, 1981). The importance of this early work was the within-subject design that included a simulator and real-world test-drive, with findings suggesting slower response times on the simulator were correlated with errors made on the test-drive. However, the use of on-road drives are not always feasible due to the expense and time involved, as well as being unpredictable in terms of encountering hazardous events. McKenna and Crick (1994) therefore sought an alternative by creating a computer-based laboratory test, and were able to discriminate between novice and experienced drivers based on response times measured by a button press. This finding, coupled with advances in computer technology, created the impetus for developing a range of computer- and simulator-based hazard perception test methodologies for which validity was based on the ability to discriminate between novice and experienced drivers. Numerous theories and models underpinned these developments, in an effort to link the methodology to cognitive and behavioural processes involved in hazard perception.

1.2. Theoretical framework

One theory that has been applied to hazard perception is signal detection theory (SDT; Green and Swets, 2000). SDT suggests that the ability to discriminate between hazardous and non-hazardous events is an important aspect of hazard perception (Ventsislavova et al., 2016). However, it has been suggested that traffic scenarios may not be definitively hazardous or non-hazardous, but *potentially* hazardous. Given this, fuzzy signal detection theory was applied, which suggests hazard responses are continuous rather than dichotomous variables based on hits, misses, correct rejections and false alarms, thereby capturing a wider range of responses (Parasuraman et al., 2000; Wallis and Horswill, 2007).

Endsley's model of situation awareness (1995) in which perception, comprehension, and projection of future status are believed to be antecedents to decision making, has also been applied to hazard perception testing (Endsley, 1995; Jackson et al., 2009). Likewise the model of responding to risk, which is a driver-specific model, proposes that variances in crash liability are due to individual differences in hazard detection ability, critical appraisal of the situation, action plan development and implementation (Grayson et al., 2003; Bellet and Banet, 2012). The multifactorial model of driver safety has also underpinned hazard perception test development, and suggests that the capacity to drive safely, self-monitoring ability, and beliefs about driving capacity are important factors in driver behaviour (Anstey et al., 2005; Horswill et al., 2008).

As at the search date, there was no consensus on which of these

frameworks is the best fit for explaining hazard perception. This being noted, the use of these frameworks has influenced hazard perception test development, by linking methodology to an understanding of cognitive and perceptual processes.

1.3. Hazard perception test methodologies

There have been a number of advances in hazard perception test methodologies since the work of McKenna and Crick (1994). Videos that contain live, staged or unstaged traffic footage continue to be widely used due to their greater ecological validity. However, genuine hazardous incidences occur infrequently when driving. To capture these events requires hours of footage which is both costly and time-consuming. There can also be issues with the ethics of deliberately aiming events that could lead to a crash. With improvements in computer graphics, animated videos were developed that allowed greater manipulation of driving scenarios, and also static image tests, using live photos or animated images (Vlakveld, 2014; Huestege et al., 2010).

Further advances in technology have seen the creation of simulated environments that better reproduce the road user's experience, such as driving simulators. Use of these simulators are considered more realistic as the participant must respond to events in a similar way as they would whilst normally driving (Crundall et al., 2012). While a test-drive is the most realistic way of assessing hazard perception, customising vehicles with in-built data recorders is costly, and there could be a confounding effect of having researchers in the car to record driver behaviour (Lee et al., 2008).

The four methodologies – video, static image, simulator and test-drive – have evolved over time, and are now used singularly or in combination to assess hazard perception for research purposes or licensing. They all present scenarios from the point of view of the road user, and methodologies have been applied to car drivers, motorcyclists, bicyclists and pedestrians across a wide age span. The challenge for researchers has been to define and then measure the important skills needed for hazard perception.

1.4. Measuring hazard perception skills

Early studies into hazard perception focused on temporal measures, often referred to as response time, which captured the time from hazard onset until detection by pressing a button (McKenna and Crick, 1994). The term 'response time' as opposed to 'reaction time' is an important distinction as simple reaction times have been shown to decline with age, however response times to hazards have been shown to improve with age and experience (Grayson and Sexton, 2002). Critically, temporal responses recorded by button press indicate hazard detection but fail to provide a measure of accuracy.

Spatial measures, such as verbal identification or touch screen, assess accuracy based on the hazard's location, and are particularly relevant for ambiguous traffic scenes or multiple hazard scenarios. Mouse click and touch screen methods allow for both temporal and spatial responses to be captured (Isler et al., 2009; Wetton et al., 2010). While individual differences in simple reaction times could potentially have a confounding effect, studies have found that when including simple reaction time as a covariate it fails to account for differences between novice and experienced drivers (Smith et al., 2009; Wetton et al., 2011). The introduction of eye tracking technology as an adjunct to all four methodologies provided both a temporal and spatial measure based on first fixation times, location of eye fixation as well as visual search strategies (Chapman and Underwood, 1998; Crundall et al., 2012).

Questionnaires have also been used to measure hazard perception and probe decision-making without providing a traditional response time measure. Many studies are based on the model of situation awareness and provide a total score based on correct answers to What? (identification), Where? (accuracy) and What happens next? (prediction) questions, referred to as the WWW format (Endsley, 1995;

Jackson et al., 2009). This has been further enhanced with questions regarding “What would you do next?” to explore risk acceptance and decision-making regarding evasive action (Gugliotta et al., 2017). These studies are largely based on the Situational Awareness Global Assessment Technique (SAGAT), where the clip is occluded at the point where the hazard appears and participants are probed for their understanding about the hazard and prediction of consequences (Endsley, 1988).

The type of hazard scenario also influences hazard perception ability as there can be single or multiple hazards, with varying levels of ambiguity or complexity, and they can be abruptly presented or contain anticipatory cues. Hazards can be part of the environment such as a parked car that may obscure ones view, or involve the behaviour of others such as a pedestrian looking to cross the road. Hazards can be classified as requiring evasive action or no action required, and there can be control scenarios that depict normal driving.

The measurement of hazard perception skills has largely been in cross-sectional studies, primarily using between-group designs comparing road user groups based on age and/or experience. Within-group designs have also been used to compare methodologies or types of hazardous scenarios. As evident, there are a variety of methodologies that include a range of stimulus and response types. However at the search date there was no consensus on which is best practice for assessing hazard perception skills.

1.5. Rationale

Few studies have shown an association between hazard perception, or the ability to anticipate potentially dangerous road and traffic situations, and crash risk. Given its importance to safe driving, hazard perception has been widely researched using a range of test methodologies, stimuli and response types. Studies have been extended from drivers to other road user groups of varying ages and levels of experience. Despite its recognised importance, to date there has been no synthesis of the available literature on hazard perception test methodologies in order to identify best practices. This review intends to provide direction for future developments in hazard perception testing that could improve outcomes for younger and novice drivers and other vulnerable road users.

1.6. Objectives

This systematic review aims to synthesise the available literature on hazard perception test methodologies. Specifically, the primary aim is to examine the methodologies for testing hazard perception, and determine best practice based on their ability to discriminate between road users that have been grouped by age and/or experience. A secondary aim is to evaluate the studies that directly compare methodologies to determine if one test is superior to another at hazard perception ability, and under what conditions this applies.

2. Method

2.1. Literature search

A systematic review of hazard perception methodologies was conducted in accordance with the PRISMA protocol (Moher et al., 2009). A comprehensive search of peer-reviewed literature within Medline, Scopus and PsycInfo databases was completed in April 2018 with no date restrictions. In order to target the search to specific hazard perception methodologies, phrase searching specific to road users was utilised with the full list of search terms in Fig. 1. A manual search of the reference lists of relevant articles was also conducted.

2.2. Inclusion and exclusion criteria

The lead researcher initially screened abstracts followed by the full text articles for eligibility, with any uncertainty resolved by screening by a second researcher until consensus was reached. Eligibility was based on the following criteria: included a methodology for assessing hazard perception skills in road users (car drivers, motorcyclists, bicyclists and pedestrians) with a quantitative measure of hazard perception as the dependent variable. Studies that also included either a between-group comparison of road user groups based on age and/or experience, or within-group comparison of hazard perception test methodologies were deemed eligible. Excluded articles were in the form of case studies, case reports, letters, editorials, guidelines, books, reviews, abstracts only, dissertations, conference and congress papers, or were not in English. Furthermore, articles were excluded if the focus was workplace, medical or environmental hazards, and if participants were from clinical populations with physical or cognitive impairment, including without normal or corrected-to-normal vision.

2.3. Data extraction

Details of selected studies were coded into a worksheet, which was piloted prior to data collection amongst two authors and refined accordingly. Data on methodology, year of publication, road user group (s), study design, participant characteristics and sample sizes, key findings and theoretical framework were extracted from each study. For each road user group, data extraction included the key measure of hazard perception ability (dependent variable), hazard response and stimulus type, quantity of a priori hazardous events, post hoc amendments to the number or type of hazardous events, and differences in hazard perception between participant groups or methodologies.

3. Results

3.1. Study search

The study search is summarised in Fig. 1. After combining and then de-duplicating the results from the three databases, an initial screening of titles was conducted, followed by abstract screening resulting in 211 titles for review. Full text screening resulted in 43 articles that met the inclusion criteria, with a further six articles added after a manual search and expert recommendation resulting in 49 studies for review.

3.2. Study characteristics

Table 1 displays the characteristics of studies selected for review, with a superscript numbering system from 1–49 used throughout to assist in the interpretation of results. Given some studies contained multiple hazard perception measures, some numbers appear more than once.

The majority of studies, 84% ($n = 41$), have been published since 2008, with the earliest study being published in 1981. Four key methodologies were used for hazard perception testing either singularly or in conjunction with other methods: videos were employed in 37 studies (76%)^{2,3,6–9,11–13,15,16,18,19,21–26,29–37,39,41–48}, five used static images (10%)^{17,19,28,32,49}, ten studies used simulators (21%)^{1,4,5,14,15,26,27,36,38,40}, and two involved a real-world test-drive (4%)^{1,9}. Single methodologies were used in 44 studies (90%) with five studies comparing two methodologies—test-drive and simulator¹, video and driving simulator^{26,36} and video and static image^{19,32}. The addition of eye tracking as an adjunct to video, static image and simulator methodologies was featured in 14 studies (29%)^{3,5,6,16,17,20,25,27,31,36,38,40,45,48}.

Participants ($N = 4,322$) across the 49 studies were largely recruited through convenience sampling, with 40 studies (82%)^{1–13,15–19,21–23,25–28,30–33,35–37,40–44,46,47,49} in car driving

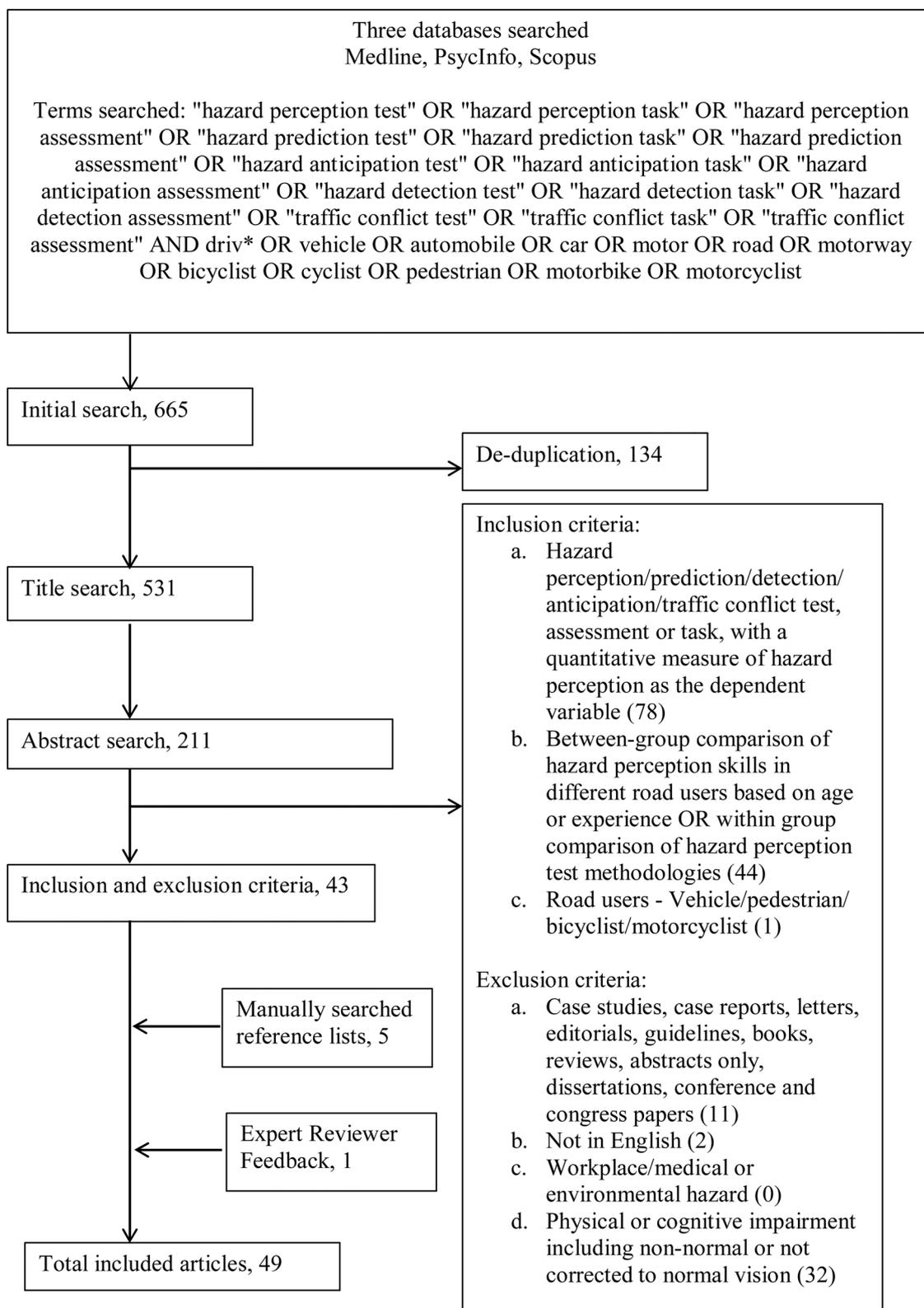


Fig. 1. Study selection procedure.

populations, followed by motorcyclists ($n = 6$; 12%)^{4,14,20,21,24,29}, pedestrians ($n = 3$; 6%)^{34,38,39} and bicyclists ($n = 2$; 4%)^{45,48}. The majority of studies ($n = 27$; 55%) compared road user groups based on experience (novice and experienced), with 11 studies (22%) categorising participants based on age (younger and older). The remaining 11 studies (23%) compared road user groups based on both age and

experience (young/novice and old/experienced). Two studies compared road user populations—car drivers and motorcyclists^{4,21}.

All 49 studies were cross-sectional, and those with a between-group design based on driving experience, had broad and varied categorisation of participants based on mean years since licensing. The range of driving experience for novice drivers ($M = 0.1$ –4.5 years since

Table 1
Characteristics of studies that include hazard perception testing methodologies and road users.

Study	Methodology	Road user group	Design	Participant groups	Conclusion	Theoretical framework
Quimby ¹ 1981 (UK)	Test-drive and simulator	Car drivers	Cross-sectional, between-group (6): age < 25, 25–34, 35–44, 45–54, 55–64, > 64 yrs; within-group (2): test drive and simulator	Age < 25 (n = 10), 25–34 (n = 11), 35–44 (n = 13), 45–54 (n = 14), 55–64 (n = 7), > 64 yrs (n = 5)	Younger drivers had slower response times to hazards in a simulator compared to mid-age drivers, who were faster than older drivers. Hazard perception skills correlated with driver error on a test drive and crash risk	N/A
McKenna ² 1994 (UK)	Video	Car drivers	Cross-sectional, between-group (2): novice, expert drivers	Novice (n = 17, M _{age} = 21.4, M _{exp} ≤ 3), expert driving instructors (n = 17, M _{age} = 34.0, M _{exp} = not specified)	Expert drivers were significantly faster at hazard perception than novice drivers	N/A
Chapman ³ 1998 (UK)	Video with eye tracker	Car drivers	Cross-sectional, between-group (2): young novice, old experienced drivers; within-group (2): road type, danger window	Young novice (n = 51, M _{age} = 18, M _{exp} < 0.25), old experienced (n = 26, M _{age} = 27, M _{exp} = 5–10)	No significant differences in hazard response times between driver groups. Novice drivers had marginally longer fixation durations than old, experienced drivers but this was not significant	N/A
Horswill and Helman, 2003 Horswill ⁴ 2003 (UK)	Simulator	Motorcyclists and car drivers	Cross-sectional, between-group (3): non-motorcycling car drivers, motorcyclists, motorcyclists as car drivers	Non-motorcycling car drivers (n = 48, M _{age} = 38.2, M _{exp} = 19.89), motorcyclists (n = 47, M _{age} = 37.5, M _{exp} = 18.79), motorcyclists as car drivers (n = 47, M _{age} = 41.1, M _{exp} = 23.00)	Motorcyclists as car drivers were significantly faster at hazard perception than both motorcyclists and non-motorcycling car drivers	N/A
Pradhan ⁵ 2005 (USA)	Simulator with eye tracker	Car drivers	Cross-sectional, between-group (3): novice, younger and older drivers	Novice (n = 24, M _{age} = 16.3, M _{exp} = 0.5), younger (n = 24, M _{age} = 21.3, older (n = 24, M _{age} = 66.9)	Novice drivers were poorer at scanning for risks and acting in response to risks compared to younger and older drivers	N/A
Underwood ⁶ 2005 (UK)	Video with eye tracker	Car drivers	Cross-sectional, between-group (2): younger, older drivers; within-group (2): hazardous, non-hazardous parts of clip	Younger (n = 12, M _{age} = 37.8, M _{exp} = 19.3), older (n = 12, M _{age} = 68.4, M _{exp} = 42.1)	No significant differences in speed of hazard responses and fixation duration between driver groups. Older drivers perceived scenes as more hazardous than younger drivers	N/A
Sagberg ⁷ 2006 (Norway)	Video	Car drivers	Cross-sectional, between-group (4): novice – 1 month, 5 months, 9 months driving experience, experienced drivers	Novice: 1 month (n = 48, M _{age} = 22.5), 5 months (n = 44, M _{age} = 22.3), 9 months (n = 38, M _{age} = 23.7); experienced (n = 28, M _{age} = 47.0, M _{exp} = 27.1)	No significant differences in detection of hazards or responses times between novice and experienced drivers	N/A
Wallis ⁸ 2007 (Australia)	Video	Car driver	Cross-sectional, between-group (3): untrained and trained novices, experienced drivers	Trained novices (n = 25, M _{age} = 18.8, M _{exp} = 1.68), untrained novices (n = 27, M _{age} = 19.4, M _{exp} = 2.13), experienced (n = 17, M _{age} = 47.2, M _{exp} = 29.29)	Trained novice and experienced drivers responded significantly more often and faster, and were more liberal in their hazard perception responses than untrained novices. No significant differences in sensitivity	Fuzzy signal detection theory (Parasuraman et al., 2000)
Horswill ⁹ 2008 (Australia)	Video	Car drivers	Cross-sectional, between-group (2): novice, experienced drivers	Novice (n = 16, M _{age} = 19.0, M _{exp} < 3), experienced (n = 17, M _{age} = 41.9, M _{exp} = > 10)	Experienced drivers were significantly faster than novice drivers in detecting hazards, with no difference in accuracy	Multifactorial model of driving safety (Anstey et al., 2005)
Lee ¹⁰ 2008 (USA)	Test-Drive	Car drivers	Cross-sectional, between-group (2): novice teen, experienced adult drivers; within-group (3): hidden stop sign, hidden pedestrian, lane closure hazards	Novice teen (with parent; n = 42, M _{age} = 16.5, newly licensed), experienced (n = 42, M _{age} = 47.2)	Experienced drivers observed and recognised hazards more frequently than novice teen drivers.	N/A
Horswill ¹¹ 2009 (Australia)	Video	Car drivers	Cross-sectional, between-group (3): young adult, old adult, elderly adult drivers	Young (n = 22, M _{age} = 48.73), old (n = 34, M _{age} = 68.71), elderly (n = 23, M _{age} = 78.78)	The elderly were significantly slower at hazard perception than both old and young adults, with no difference between the latter two groups	N/A
Isler ¹² 2009 (NZ)	Video	Car drivers	Cross-sectional, between-group (2): young/less experienced, experienced drivers	Young/less experienced (n = 24, Age = 18 or 19 yrs, M _{exp} = 1.5), experienced (n = 8, M _{age} = 35.5, M _{exp} = 15.5)	Young/less experienced drivers detected fewer hazards than experienced drivers, and had significantly longer response times.	N/A

(continued on next page)

Table 1 (continued)

Study	Methodology	Road user group	Design	Participant groups	Conclusion	Theoretical framework
Jackson ¹³ 2009 (UK)	Video	Car drivers	Cross-sectional, between-group (2 × 2): novice, experienced drivers; freeze frame (FF) and cut-to-black (CB) clips	Novice (n = 41, M _{age} = 19.0, M _{exp} = 0.58), experienced (n = 39, M _{age} = 23.0, M _{exp} = 4.92)	Novice drivers were less accurate at hazard perception than experienced drivers in the CB clips, with no difference between driver groups for FF clips	Situation awareness (Endsley, 1995)
Liu ¹⁴ 2009 (Australia)	Simulator	Motorcyclists	Cross-sectional, between-group (4): inexperienced, experienced riders (with full car license), novice riders (with full car license and no motorbike license), novice riders (probationary car license and no motorbike license)	Inexperienced (n = 12, M _{age} = 33, M _{exp} = 0.67), experienced (n = 12, M _{age} = 41, M _{exp} = 14), novice – car license (n = 12, M _{age} = 34, M _{exp} = 0), novice – probationary car license (n = 12, M _{age} = 20, M _{exp} = 0)	Experienced riders were less likely to crash, performed better when approaching hazards and at lower speeds, compared to novice riders on a probationary license. There were few differences between inexperienced, and novice riders with full license	N/A
Smith ¹⁵ 2009 (Australia)	Video	Car drivers	Cross-sectional, between-group (2 × 2): novice, experienced, day testing (10am), night testing (3am)	Novice (n = 32, M _{age} = 19.88, M _{exp} = 1.65), experienced (n = 30, M _{age} = 31.78, M _{exp} = 14.41)	Experienced drivers had significantly faster and more accurate hazard detection compared to novice drivers. Novices were slower at night with no day/night differences for experienced drivers	N/A
Borowsky ¹⁶ 2010 (Israel)	Video with eye tracking	Car drivers	Cross-sectional, between-group (3): young, experienced and old car drivers; within-group (3): pedestrian, car-following, intersection-related hazards	Young (n = 21, Age = 19–20, M _{exp} = 0.22), experienced (n = 19, Age = 19–30, M _{exp} = 7.30), older (n = 16, Age = 65–72, M _{exp} = 37.50)	Differences in hazard perception were dependent on hazard location. No significant differences for pedestrian events. Old drivers were the slowest to respond to intersection and car-following events. Old and experienced drivers had wider visual scan patterns than young drivers	N/A
Huestegge ¹⁷ 2010 (Germany)	Static image with eye tracking	Car drivers	Cross-sectional, between-group (2): inexperienced, experienced; within-group (2): scene type (medium, highly dangerous)	Inexperienced (n = 20, M _{age} = 18, M _{exp} = 0.75), experienced (n = 20, M _{age} = 24, M _{exp} = 5.00)	Experienced drivers were significantly faster at detecting hazards compared to inexperienced drivers, for all scene types	N/A
Shahar ¹⁸ 2010 (UK)	Video	Car drivers	Cross-sectional, between-group (2): single-screen, three-screen; within-group (3): central, peripheral, lateral hazard location on screen	Single-screen (n = 20, M _{age} = 25.9), three-screen (n = 19, M _{age} = 23.2)	Hazard response times were significantly better in the three-screen condition, and were faster for hazards centrally positioned compared to peripherally or laterally. There was no significant difference in accuracy	N/A
Wetton ¹⁹ 2010 (Australia)	Video and static image	Car driver	Cross-sectional, between-group (2): novice and experienced drivers; within-group (4): ACT HPT (video), ACT Hazard detection task (static image), UK HPT (video), QLD HPT (video)	Novice (n = 24, M _{age} = 21.3, M _{exp} < 3.0), experienced (n = 24, M _{age} = 40.6, M _{exp} = + 10.0)	Experienced drivers had significantly faster response times than novices for the ACT HPT task, however novice drivers were faster on the ACT Hazard Change Detection task. ACT HPT scores correlated with existing UK and QLD HPTs. The ACT Hazard Change Detection task was not correlated with any other test	N/A
Cheng ²⁰ 2011 (Hong Kong)	Simulator with eye tracking	Motorcyclists	Cross-sectional, between-group (2): crash-free, crash-involved motorcyclists	Crash-free (n = 63, M _{age} = 35.4, M _{exp} = 15.2), crash-involved (n = 46, M _{age} = 37.8, M _{exp} = 17.4)	Crash-free motorcyclists were significantly faster at detecting hazards than crash-involved riders	N/A
Rosenbloom ²¹ 2011 (Israel)	Video	Motorcyclists and car drivers	Cross-sectional, between-group (2): motorcyclists, car drivers	Motorcyclists (n = 35, M _{age} = 28.5), car drivers (n = 25, M _{age} = 25.9)	Motorcyclists performed significantly better on a hazard perception test than car drivers	N/A
Scialfa ²² 2011 (Canada)	Video	Car drivers	Cross-sectional, between-group (2): young novice, young experienced drivers	Young novice (n = 29, M _{age} = 20.15, M _{exp} = 2.06), young experienced (n = 146, M _{age} = 21.3, M _{exp} = 5.8)	Novice drivers were significantly slower at detecting hazards than experienced drivers	N/A
Wetton ²³ 2011 (Australia)	Video	Car driver	Cross-sectional, between-group (2): learners, experienced drivers	Learners (n = 94, M _{age} = 18.02, M _{learn} = 0.35), experienced (n = 56, M _{age} = 47.12, M _{exp} = 28.56)	Experienced drivers were significantly faster at hazard perception than learner drivers	N/A

(continued on next page)

Table 1 (continued)

Study	Methodology	Road user group	Design	Participant groups	Conclusion	Theoretical framework
Belle ²⁴ 2012 (France)	Video	Motorcyclists	Cross-sectional, between-group (4): beginner, novice, experienced, expert (police) motorcyclists	Beginner ($n = 12$, $M_{age} = 21$, $M_{turn} = 17$ –28 hrs), novice ($n = 12$, $M_{age} = 23$, $M_{exp} = 0.66$), experienced ($n = 12$, $M_{age} = 32$, $M_{exp} = 13.00$), expert (police; $n = 12$, $M_{age} = 37$, $M_{exp} = 15.00$)	Expert and experienced motorcyclists performed significantly better on the hazard perception test than novice and beginner riders	Situation awareness (Endsley, 1995), Model of responding to risk (Grayson et al., 2003)
Borowsky ²⁵ 2012 (Israel)	Video with eye tracking	Car drivers	Cross-sectional, between-group (5): young inexperienced (control, active training, instructional training, hybrid training), experienced car drivers	Young, inexperienced (control $n = 10$, trained-active $n = 10$, instructional $n = 10$, hybrid $n = 10$, Age = 17–18 yrs, $M_{exp} = 0.14$), experienced ($n = 21$, $M_{exp} = 8.0$, Age = 23–29 yrs)	No significant difference in identifying pedestrian-related hazardous events between experienced and young, inexperienced (control) drivers. Experienced drivers had more efficient visual scanning patterns than inexperienced drivers	N/A
Bromberg ²⁶ 2012 (Israel)	Video and simulator	Car drivers	Cross-sectional, between-group (2): experienced and elderly, experienced car drivers; within-group (2): video, driving simulator task	Experienced ($n = 22$, $M_{age} = 31.0$, $M_{exp} = 13.0$), elderly, experienced ($n = 20$, $M_{age} = 68.0$, $M_{exp} = 43.0$)	Elderly drivers had longer response times than younger drivers, but with no significant difference in detection of pedestrians in the video technique. Slower driving speed of elderly drivers moderated hazard detection in the simulator task	N/A
Crundall ²⁷ 2012 (UK)	Simulator with eye tracking	Car drivers	Cross-sectional, between-group (3): learner, experienced, car driving instructor; within-group (3): environmental, behavioural, dividing and focussing attention hazards	Learner ($n = 14$, $M_{age} = 20.3$, $M_{com} = 0.63$), experienced ($n = 17$, $M_{age} = 33.0$, $M_{exp} = 16.4$), driving instructor ($n = 18$, $M_{age} = 48.5$, $M_{exp} = 30.0$)	Experienced drivers and instructors were significantly more accurate and faster at detecting hazards, with longer fixation durations than learner drivers. Environmental hazards were more likely to be missed by learner drivers	Situation awareness (Endsley, 1995)
Scialfa ²⁸ 2012 (Canada)	Static images	Car drivers	Cross-sectional, between-group (2): young novice, young experienced drivers	Young novice ($n = 25$, $M_{age} = 19.9$, $M_{exp} = 0.19$), young experienced ($n = 26$, $M_{age} = 20.4$, $M_{exp} = 5.10$)	Novice drivers were significantly slower and less accurate at hazard detection than experienced drivers	Situation awareness (Endsley, 1995)
Crundall ²⁹ 2013 (UK)	Video	Motorcyclists	Cross-sectional, between-group (3): novice, experienced, advanced motorcycle riders	Novice ($n = 20$, $M_{age} = 26.5$, $M_{exp} = 1.0$), experienced ($n = 21$, $M_{age} = 40.6$, $M_{exp} = 15.6$), advanced ($n = 20$, $M_{age} = 47.4$, $M_{exp} = 16.6$)	Advanced riders were faster at detecting hazards than novice riders followed by experienced riders, after adjusting for responses before the a priori hazard onset. No significant difference in accuracy. Advanced riders had greater internal attributions as to the cause of the hazard, compared to external attributions for novice riders	N/A
Horswill ³⁰ 2013 (Australia)	Video	Car drivers	Cross-sectional, between-group (2): experienced drivers (without training), police officers (with training)	Experienced (control; $n = 42$, $M_{age} = 49.27$, $M_{exp} = 30.54$), police officers ($n = 26$, $M_{age} = 47.04$, $M_{exp} = 29.27$)	Police officers were significantly faster and more accurate at detecting hazards than experienced car drivers	N/A
Lim ³¹ 2013 (Malaysia)	Video with eye tracking	Car drivers	Cross-sectional, between-group (2 × 2): novice – UK and Malaysian, experienced – UK and Malaysian drivers; within-group (2): UK, Malaysian film clips	UK novice ($n = 20$, $M_{age} = 18.9$, $M_{exp} = 0.76$), UK experienced ($n = 25$, $M_{age} = 21.0$, $M_{exp} = 3.8$), Malaysian novice ($n = 26$, $M_{age} = 18.8$, $M_{exp} = 0.82$), Malaysian experienced ($n = 27$, $M_{age} = 21.9$, $M_{exp} = 4.12$)	No significant difference in hazard perception or visual response patterns between novice and experienced drivers regardless of country of origin	N/A
Scialfa ³² 2013 (Canada)	Video and static image	Car drivers	Cross-sectional, within-group (2): static, dynamic HPT	Experienced drivers ($n = 56$, $M_{age} = 20.7$, $M_{exp} = 4.5$)	The correlation between static and dynamic HPT was low to moderate after controlling for simple reaction time. Both versions had good reliability	N/A

(continued on next page)

Table 1 (continued)

Study	Methodology	Road user group	Design	Participant groups	Conclusion	Theoretical framework
Castro ³³ 2014 (Spain)	Video	Car drivers	Cross-sectional, between-group (3): learner, novice and experienced car drivers; within-group (2 × 2): environmental, behavioural hazards and hazardous, quasi-hazardous clips	Learner (n = 14, M _{age} = 21.7, M _{exp} = 0, learning), novice (n = 16, M _{age} = 20.6, M _{exp} = 0.87), experienced (n = 14, M _{age} = 37.5, M _{exp} = 14.07)	Experienced drivers scored significantly higher on the HPT than novice and learner drivers, for both hazardous and quasi-hazardous clips. Scores were higher for behavioural hazards than environment hazards	Situation awareness (Endsley, 1995)
Meyer ³⁴ 2014 (Norway)	Video	Pedestrians	Cross-sectional, between-group (3): children 8–12, 13–17 yrs, adults 18yrs +	8–12 yrs (n = 259, M _{age} = 10.4), 13–17 yrs (n = 109, M _{age} = 14.2), 18 yrs + (n = 172, M _{age} = 43.4)	Young children had the slowest response times compared to teenagers and adults. Few significant differences in response rates dependent on scenario type. Younger children detected fewer hazards, with the largest differences for situations that involved pedestrians.	N/A
Vlaekveld ³⁵ 2014 (The Netherlands)	Video	Car driver	Cross-sectional, between-group (2 × 2): learner, professional drivers, task 1 (view video twice), task 2 (view video once); within-group (2): overt, covert hazards	Task 1: Learner (n = 30, M _{age} = 22.5, M _{learn} = 28.7 hrs), professional (n = 34, M _{age} = 44.3, M _{mileage} = 41,359) Task 2: Learner (n = 30, M _{age} = 21.4, M _{learn} = 35.6 hrs), professional (n = 30, M _{age} = 44.9, M _{mileage} = 54,187)	Professionals scored significantly higher on both tasks compared to learner drivers. Task 1 was moderately better at distinguishing professional and learner drivers due to greater differences in identifying overt latent hazards, however had a lower internal consistency than Task 2	Situation awareness (Endsley, 1995)
Mackenzie ³⁶ 2015 (UK)	Video with eye tracker and simulator	Car drivers	Cross-sectional, between-group (2): simulated video (non-driving), simulator (driving)	Driving (n = 17, M _{age} = 22.3, M _{exp} = 3.1), non-driving (n = 17, M _{age} = 22.3, M _{exp} = 2.6)	Those performing the non-driving task were significantly faster at hazard detection, and had wider visual search strategies than those performing the driving task, with no difference in accuracy	N/A
Malone ³⁷ 2015 (Germany)	Video	Car drivers	Cross-sectional, between-group (2): learner, experienced drivers; within-group (2): high, low ecological validity task	Learner (n = 35, M _{age} = 17.97, M _{exp} < 4 lessons), experienced (n = 31, M _{age} = 26.77, M _{exp} > 2)	Expert drivers outperformed learners when accurately identifying hazards, but with no significant difference in response time between tasks. Accuracy was marginally higher but not statistically significant in the low ecological validity task compared to the high ecological validity task	N/A
Meit ³⁸ 2015 (Israel)	Simulator with eye tracker	Pedestrians	Cross-sectional, between-group (4): children 7–9, 9–10, 10–13yrs, adults; within-group (4): zebra crossing, moving vehicles, obscured view parked car, obscured view road curvature	7–9 yrs (n = 9), 9–10 yrs (n = 7), 10–13 yrs (n = 11), adults (n = 20)	Younger children were less accurate at identifying hazards compared to older children and adults, and had longer response times when their field of view was obscured by parked vehicles. No other significant differences in response time between age groups or by hazard type	N/A
Rosenbloom ³⁹ 2015 (Israel)	Video	Pedestrians	Cross-sectional, between-group (3): children, adults, older people	Children (n = 158, M _{age} = 8.87), adults (n = 113, M _{age} = 25.36), older people (n = 88, M _{age} = 72.50)	Adults had the highest overall hazard perception score, followed by children and then older people. Children were the least likely to increase their field of view using left and right arrow keys	N/A
Yeung ⁴⁰ 2015 (Singapore)	Simulator with eye tracker	Car driver	Cross-sectional, between-group (3): young novice, young experienced and old experienced drivers	Young novice (n = 14, M _{age} = 21.93, M _{exp} = 1.43), young experienced (n = 14, M _{age} = 28.57, M _{exp} = 8.29), old experienced (n = 12, M _{age} = 56.58, M _{exp} = 29.5)	No effects of age or experience were found for response time and mean fixation duration despite older drivers being generally slower. The first fixation was significantly slower for older drivers in certain hazard scenarios, with less scanning compared to younger drivers	N/A

(continued on next page)

Table 1 (continued)

Study	Methodology	Road user group	Design	Participant groups	Conclusion	Theoretical framework
Castro ⁴¹ 2016 (Spain)	Video	Car drivers	Cross-sectional, between-group (3): learner, novice and experienced drivers; within-group (3 × 2): what, where, what next questions, gradual, abrupt onset hazards	Learner (n = 20, M _{age} = 20.3, M _{exp} = 0.00, lessons only), novice (n = 62, M _{age} = 22.3, M _{exp} = 2.00), experienced (n = 40, M _{age} = 37.1, M _{exp} = 4.00)	Experienced drivers performed significantly better than novice and learner drivers. There were greater differences between driver groups for abrupt hazards compared to gradual-onset hazards	Situation awareness (Endsley, 1995)
Crundall ⁴² 2016 (UK) a	Video	Car drivers	Cross-sectional, between-group (2): novice, experienced car drivers; within-group (3): long, intermediate, short clip length	Novice (n = 15, M _{age} = 20.8, M _{exp} = 2.1), experienced (n = 15, M _{age} = 20.8, M _{exp} = 3.7)	Experienced drivers were significantly more accurate at identifying hazards than novice drivers, regardless of clip length. Novice drivers were significantly poorer at hazard prediction in the longer clips	Situation awareness (Endsley, 1995)
Crundall ⁴² 2016 (UK) b	Video	Car drivers	Cross-sectional – between-group (2): novice, experienced car drivers; within-group (3): early, intermediate, late clip end point	Novice (n = 21, M _{age} = 20.5, M _{exp} = 1.8), experienced (n = 21, M _{age} = 23.9, M _{exp} = 6.4)	Experienced drivers were significantly more accurate at identifying hazards than novice drivers, regardless of clip occlusion point	Situation awareness (Endsley, 1995)
Crundall ⁴² 2016 (UK) c	Video	Car drivers	Cross-sectional, between-group (2): novice, experienced car drivers; within-group (2) – behavioural, environmental hazard type	Novice (n = 15, M _{age} = 20.7, M _{exp} = 0.9), experienced (n = 15, M _{age} = 20.8, M _{exp} = 3.4)	Experienced drivers were significantly more accurate at identifying all hazard types than novice drivers. Novice drivers were significantly poorer at predicting environmental hazards in which there were limited precursors	Situation awareness (Endsley, 1995)
Johnston ⁴³ 2016 (Canada)	Video	Car drivers	Cross-sectional, between-group (2): civilian, emergency medical services (EMS) drivers	Civilian (n = 20, M _{age} = 29.5, M _{exp} = 9.92), EMS (n = 28, M _{age} = 30.5, M _{exp} = 18.83)	EMS were significantly faster at hazard perception than civilian drivers, with no difference in accuracy	N/A
Malone ⁴⁴ 2016 (Germany)	Video	Car drivers	Cross-sectional, between-group (2 × 2): learner and experienced drivers, high and low ecological validity tasks; within-group (3): no hazard, materialised hazard, potential hazard	Learner (n = 101, M _{age} = 20.0, M _{exp} = 0, < 4 lessons), experienced (n = 49, M _{age} = 29.9, M _{exp} = 10.69)	Experienced drivers outperformed learners for no hazard/materialised hazard scenarios. No significant differences for potential hazards. High validity task was a better discriminator of driver groups for no hazard scenarios, with no effect of task type on performance for materialised and potential hazard scenarios	N/A
Vansteenkiste ⁴⁵ 2016 (Belgium)	Video with eye tracker	Bicyclists	Cross-sectional, between group (2): children, adult bicyclists	Children (n = 11, M _{age} = 8.36), adults (n = 17, M _{age} = 21.65)	No significant difference in accuracy or hazard fixation between adult and children bicyclists. Adult response times were significantly faster than children for only 4 out of 11 clips involving vulnerable road users	Situation awareness (Endsley, 1995)
Ventsislavova ⁴⁶ 2016 (Spain)	Video	Car driver	Cross-sectional, between-group (5): learners, novices – offenders and non-offenders, experience – offenders and non-offender drivers	Learners (n = 34, M _{age} = 19.3, M _{exp} = 0, learning) novice non-offenders (n = 36, M _{exp} = 3.92), novice offenders (n = 4, M _{age} = 23.8, M _{exp} = 7.75), experienced non-offenders (n = 54, M _{age} = 20.33), experienced offenders (n = 21, M _{age} = 41.9, M _{exp} = 20.67)	Hazard detection, situation awareness and degree of cautiousness were significantly higher for experienced drivers compared to novices and learners drivers. Offenders were less cautious and detected fewer hazards than non-offenders	Signal Detection Theory (Green and Swets, 2000); Situation awareness (Endsley, 1995)
Gugliotta ⁴⁷ 2017 (Spain)	Video	Car drivers	Cross-sectional, between-group (3): learner, novice and experienced drivers; within-group (4): what? Where? What next? Decision making questions	Learner (n = 20, Age = 18–27, M _{exp} = 1, lessons), novice (n = 62, M _{age} = 22.3, M _{exp} = 4.54), experienced (n = 40, M _{age} = 38.0, M _{exp} = 20.97)	Learner drivers were significantly less accurate at detecting hazards than novice and experienced drivers. Decision making questions were answered more accurately than situation awareness questions	Situation awareness (Endsley, 1995)
Zeuwis ⁴⁸ 2017 (Belgium)	Video with eye tracker	Bicyclists	Cross-sectional, between-group (2): children, adult bicyclists	Children (n = 75, M _{age} = 10.53, riding experience ≥ 2) adult (n = 41, M _{age} = 21.64)	Children had significantly delayed response times, with shorter fixation durations and dwell times compared to adults	Situation awareness (Endsley, 1995)

(continued on next page)

Table 1 (continued)

Study	Methodology	Road user group	Design	Participant groups	Conclusion	Theoretical framework
Feng ⁴⁹ 2018 (Canada)	Static images	Car drivers	Cross-sectional, between-group (2): young, older car drivers; within-group (2 × 2): target presence, traffic load	Young (n = 16, M _{age} = 24.4, M _{exp} = 5.1), older (n = 21, M _{age} = 70.6, M _{exp} = 44.1)	Older drivers were significantly less accurate and slower at hazard perception than younger drivers, with greater differences when there was a high traffic load.	N/A

Note: Key findings in bold indicates no significant difference ($p > .05$) between participant groups or methodology. M_{age} = mean age in years; M_{exp} = mean driving experience in years; M_{learn} = mean learning experience in years; M_{mileage} = mean driving mileage in kilometres; HPT, hazard perception test; ACT HPT, Australian Capital Territory Hazard Perception Test; UK HPT, United Kingdom Hazard Perception Test; QLD HPT, Queensland Hazard Perception Test. Sensitivity, number of participants in each group who responded to an event as a percentage of the group population; N/A, Not Applicable.

licensing) overlapped with experienced drivers ($M = 2.0$ – 29.9 years). Likewise for studies with a between-group design based on driver's age, there was an inconsistent categorisation of younger drivers ($M = 16.5$ – 48.7 years), and older drivers ($M = 47.2$ – 68.7 years), albeit with less overlap.

The use of a theoretical framework or model of driver behaviour was only observed in 14 studies (29%). Twelve studies were framed around the model of situation awareness^{13,24,27,28,33,35,41,42,45–48}, one study used signal detection theory⁴⁶ and one study used fuzzy signal detection theory⁸. Only two studies used models specific to driver behaviour – the model of responding to risk²⁴ and multifactorial model of driver safety⁹.

The four methodologies – video, static image, simulator and real-world test-drive, were used to capture 18 measures of hazard perception ability or dependent variables, with 12 hazard response types used singularly or in conjunction with other types. Across the 48 studies, the different combinations of methodologies, dependent variables and hazard response types resulted in 117 measures of hazard perception (see Table 2 for a summary of all recorded classes).

Data tables are organised by methodology and then by the dependent variable followed by hazard response type and stimulus type. Differences in response types were dependent on the research question regarding the temporal, spatial and/or decision-making differences in hazard perception ability. In order to better understand the differences in hazard perception ability of specific road user groups, the decision was made a priori to analyse road user groups independently. Results are presented for car drivers (see Table 3), motorcyclists (see Table 4), bicyclists (see Table 5) and pedestrians (see Table 6).

Differences in hazard perception ability between road users groups reflect the expected direction that novice/younger road users have poorer hazard perception skills than experienced/older road users, unless stated otherwise. Results with no statistically significant difference between groups were in the minority and are therefore highlighted in bold.

3.3. Hazard perception testing of car drivers

The results for studies of car driving populations ($n = 40$) indicate that all four methodologies i.e. video, static image, simulator and real-world test-drive, discriminated between driver groups based on age and/or experience, on at least one measure of hazard perception ability (see Table 3). Videos were employed in 30 studies (75% of car driver studies)^{2,3,6–9,11–13,15,16,18,19,22,23,25,26,30–33,35–37,41–44,46,47}, six studies used driving simulators (15%)^{1,5,26,27,36,40}, five used static images (13%)^{17,19,28,32,49} and two involved a test-drive (5%)^{1,10}.

Temporal measures (i.e. response times) were captured in 27 studies (68% of car driver studies)^{1–3,6–9,11,12,15–19,22,23,26–28,30,31,36,37,40,43,44,49}. Of these, 21 studies^{1,2,8,9,11,12,15–19,22,23,26–28,30,31,36,37,40,41,44,49} found a response time difference in the expected direction with novice/younger drivers significantly slower than experienced/older drivers at hazard perception. Six studies (15% of car driving studies)^{3,6,7,31,37,40} found no differences between driver groups. There were four studies that compared response times across methodologies. Novice drivers were slower than experienced drivers at hazard perception for video tests, but had faster response times for a static image hazard change detection test (Scialfa et al.³², 2013), and these two methodologies were not correlated (Wetton et al.¹⁹, 2010). Drivers were slower to respond to hazards in a simulator compared to a video test (Mackenzie and Harris³⁶, 2015) and the majority of elderly drivers who failed to respond to a video hazard, did initiate braking action in a simulator albeit later than the average brake time (Bromberg et al.²⁶, 2012).

Spatial measures that provide an accuracy score were found in 27 studies (68% of car driver studies)^{1,5–9,12,13,15–17,25,27,28,30,31,33,35–38,41,42,44,46,47,49} with 10 of these studies^{6–9,16,25,31,35–37} showing no significant differences between driver groups or between methodologies. Accuracy measures were primarily obtained using response types that also provided a temporal

Table 2
Summary of hazard perception test methodologies, dependent variables, hazard response types and stimulus types.

Methodology	Dependent variable	Hazard response type	Stimulus type
•Video	•Response time	•Press button	•Live video – staged
•Static images	•Accuracy	•Mouse click	•Live video – unstaged
•Simulator	•Total responses	•Touch screen	•Animated video
•Test-Drive	•Sensitivity	•Track-ball marker	•Image – taken from live footage
•Video and static image	•Hazard perception score	•Lever	•Animated image
•Video and simulator	•Cautiousness	•Verbal or written identification	•Hazard on real-world test drive
•Simulator and test-drive	•Time to first fixation	•Questionnaire – What? Where? What next? (WWW) format	
	•Fixation duration	•Questionnaire–multiple choice format	
	•Number of fixations	•Questionnaire–locus of control	
	•Horizontal variance of fixation	•Eye tracking	
	•Vertical variance of fixation	•Braking response	
	•Fixation until hazard response	•Acceleration/deceleration	
	•Safety index		
	•Situation criticality		
	•Locus of control		
	•Crashes		
	•Rider performance		
	•Speed		

measure such as mouse click, button press or eye fixation. Nine studies (33% of car driver studies)^{13,27,28,33,35,41,42,46,47} extended the measure of accuracy based on identification, to include a deeper understanding of decision-making about the hazard. This was largely based on SAGAT, with the majority used the WWW questionnaire format^{13,33,41,42,46,47} (Endsley, 1995, 1988). The remaining studies used a combination of touch screen, eye movement data, and questionnaire. All studies found a difference between driver groups based on age and experience. While a further two studies^{37,44} used verbal and written identification to provide accuracy scores, these were not based on a theoretical framework and were able to show a difference in the expected direction. Of the 10 studies that failed to find a difference in accuracy scores between driver groups, five studies also failed to find a difference in response times using button press^{6,7,8,9,16} and one using mouse click³¹. Furthermore, two studies^{37,44} directly compared accuracy across different hazard perception response types—a multiple choice questionnaire and response time task. There were no significant difference in scores but both response types were individually able to detect differences between novice and experienced drivers in the expected direction. This suggests that both response types are equally able to capture hazard identification. There was only one study that included simulator and test-drive methodologies, with driver errors on a real-world test drive associated with the number of hazards identified in both the simulator and on the test-drive (Quimby and Watts¹, 1981).

Eye tracking provided both temporal and spatial measures based on hazard response types such as time to first fixation and number of fixations on the critical stimuli, respectively ($n = 10$, 25% of car driver studies)^{3,5,6,16,17,25,27,31,36,40}. Two studies^{17,31} found no difference in time to first fixation between driver groups, and five studies^{3,6,17,25,40} found no difference in fixation duration. Eye tracking was able to provide additional information beyond traditional button press responses such as visual scanning patterns, with novice drivers having a smaller variance in horizontal and vertical fixations than experienced drivers^{16,25}.

There were three main stimulus types used in the car driver studies; live video either staged or unstaged, and animated video. All stimulus types were able to discriminate between novice and experienced drivers. There was a post hoc amendment of clips or hazardous events, based on their ability to discriminate between driver groups or uninterpretable data, in 16 studies (40% of car driver studies)^{5,8,16,17,22,23,26,28,32,33,35,36,41,43,46,47}. The criteria for inclusion/exclusion varied from quantitative measures such as an 85% response rate, to undefined criteria such as low detection rates. Only two studies reported pre- and post-amendment results^{23,28}.

3.4. Hazard perception testing of motorcyclists

There were six studies of motorcycling populations (see Table 4) with three using video^{21,24,29} and three using a simulator^{4,14,20}. Results were in line with those for car drivers, with both methodologies able to discriminate between rider groups. Of the studies using the video methodology, only one study used footage filmed from the motorcyclist's perspective²⁹ with the remaining studies presenting a car driver's view^{21,24}.

Temporal measures were captured in four studies^{4,20,24,29}, with only one study by Crundall et al.²⁹ (2013) failing to find a difference between novice and experienced riders. However by performing a post hoc amendment of clips to widen the hazard window allowing earlier responses, this then produced a statistically significant difference in the expected direction. This study was also the only one to capture a spatial measure based on accuracy, with no difference between novice and experienced riders. Furthermore this study was unique in that it also captured locus of control, which is the degree to which people attribute events to external or internal factors. It was measured on a continuum based on answers to the question 'Why did the hazard occur?' with novice riders making more external than internal attributions, and this pattern was reversed in experienced riders (Crundall et al.²⁹, 2013). The study by Bellet and Banet (2012)²⁴, was the only study based on the model of situation awareness, where both the temporal response and measure of the criticality of the hazard, were able to differentiate between riders based on experience.

The simulator study by Liu et al.¹⁴ (2009) examined additional rider behaviours beyond response times including crashes, rider performance (score calculated by the simulator) and speed measures. Results were inconsistent and depended on the scenario type (rural traffic, medium density or high density city traffic). There were no studies that compared methodologies.

3.5. Hazard perception testing of bicyclists

There were only two studies^{45,48} of bicyclist populations (see Table 5), and both used video methodology with eye-tracking variables as the hazard response type. There were a number of inconsistencies in results for response times, and visual scanning behaviour.

A temporal measure was captured in both studies, with Vansteenkiste et al.⁴⁵ (2016) finding no difference in response times between younger and older cyclists for 8 out of 13 clips. Differences were found for clips that involved interactions with other cyclists or pedestrians. Likewise, Zeuwts et al.⁴⁸ (2017) found significant differences in response times for clips that depicted covert hazards involving

Table 3
Results for studies that include hazard perception test methodologies for car drivers.

Methodology	Dependent variable	Hazard response type	Stimulus type	Study	Initial clip or hazardous event quantity (n)	Difference between participant groups or methodology	Inclusion/exclusion criteria for amending clip or hazardous event quantity	Post hoc clip or hazardous event quantity (n)	Difference between road user group or methodology
Video	Response time	Press button	Live, staged and unstaged	McKenna ² 1994	35	Yes: $p < .001$			
		Press button	Live	Chapman ³ 1998	13	No: $p > .05$			
		Press button	Live, unstaged	Underwood ⁶ 2005	10	No: $p > .05$			
		Press button	Live, unstaged	Sagberg ⁷ 2006	Block A 15 Block B 16		No: $p > .05$		
		Press button	Live, unstaged	Wallis ⁸ 2007	27	NR	90% response rate	12	Yes: $p < .01$
		Press button	Live, unstaged	Horswill ⁹ 2008	33	Yes: $p = .003$			
		Press button	Live, unstaged	Horswill ¹¹ 2009	33	Yes: $p = .011$			
		Press button	Live, unstaged	Borowsky ¹⁶ 2010	4 clips/23 hazardous events	NR	High support events with $\geq 30\%$ response rate or low support event with $\geq 20\%$ sensitivity		Yes: $p < .05$ (car following, urban area intersections; old slower than young and experienced drivers) No: $p > .05$ (pedestrian, car-following, residential event)
		Press button	Live, staged	Shahar ¹⁸ 2010	13	Task comparison Yes: $p = .05$ No: $p = .13$			
		Press button (task 1 only)	Animated	Malone ³⁷ 2015	20				
Video (cont'd)	Response time (cont'd)	Press button (task 1)	Animated	Malone ⁴⁴ 2016	20	Yes: $p = .04$ (materialised hazards) No: $p = .15$ (potential hazards) Yes: $p < .001$			
		Mouse click	Live, unstaged	Smith ¹⁵ 2009	Alternative forms 46 and 43				
		Mouse click	Live, unstaged	Wetton ²³ 2011	90	Yes: $p < .001$	Largest effect sizes/variety of traffic conflicts	60, then 15	Yes: $ps < .001$
		Mouse click	Live, unstaged	Horswill ³⁰ 2013	17	Yes: $p = .028$			
		Mouse click	Live, unstaged	Lim ³¹ 2013	40	No: $p > .05$			
		Mouse click	Animated	Isler ¹² 2009	20	Yes: $p < .05$			
		Touch screen	Live, unstaged	Scialfa ²² 2011	64	NR	85% + hit rate for experienced drivers Multiple hazards (ambiguous), 15% false alarm rate for experienced drivers	20	Yes: $p = .006$
		Touch screen	Live, unstaged	Johnston ⁴³ 2016	64	NR	85% + hit rate for drivers 15% false alarm rate	42	Yes: $p = .034$
		Press button	Live, unstaged	Underwood ⁶ 2005	10	No: $p > .05$			
		Press button	Live, unstaged	Horswill ⁹ 2008	33	No: $p > .05$			

(continued on next page)

Table 3 (continued)

Methodology	Dependent variable	Hazard response type	Stimulus type	Study	Initial clip or hazardous event quantity (n)	Difference between participant groups or methodology	Inclusion/exclusion criteria for amending clip or hazardous event quantity	Post hoc clip or hazardous event quantity (n)	Difference between road user group or methodology
Video (cont'd)	Accuracy (cont'd)	Verbal identification	Live, staged	Borowsky ¹⁶ 2010	23	NR	High support events – $\geq 30\%$ response rate or low support event with $\geq 20\%$ sensitivity	Pedestrian events (n = 4), car following (n = 2), intersection (n = 5)	Yes: $p < .05$ (car-following hazard, urban) No: $p > .05$ (pedestrian, car-following residential, intersection hazards)
		Verbal identification	Live, unstaged	Lim ³¹ 2013	40	No: $p > .05$			
		Verbal identification	Animated	Isler ¹² 2009	20	Yes: $p < .01$			
		Mouse click	Live, unstaged	Smith ¹⁵ 2009	Alternative forms 46 and 43	Yes: $p = .008$			
		Mouse click	Live, unstaged	Horswill ³⁰ 2013	17	Yes: $p = .034$			
		Press button and mouse click	Animated	Vlakveld ³⁵ 2014	13 in each of 2 tasks	Driver comparison No: Task 1 $p > .05$ (8 of 13 clips) No: Task 2 $p > .05$ (11 of 13 clips)			
		Press button (task 1) or verbal identification/questionnaire (task 2)	Animated	Malone ³⁷ 2015	20 for each task	Driver comparison Yes: $p < .05$			
Video (cont'd)	Accuracy (cont'd)	Press button (task 1) or verbal identification/questionnaire (task 2)	Animated	Malone ³⁷ 2015	20 for each task	Task comparison No: $p < .10$			
		Press button (task 1) or verbal identification/questionnaire (task 2)	Animated	Malone ⁴⁴ 2016	20 for each task	Driver comparison Yes: $ps \leq .001$ (no hazard/materialised hazard scenarios) No: $p > .05$ (potential hazards)			
		Press button (task 1) or verbal identification/questionnaire (task 2)	Animated	Malone ⁴⁴ 2016	20 for each task	Task comparison Yes: $p < .001$ (no hazards) No: $p > .05$ (materialised/potential hazards)			
Total responses		Press button	Live, unstaged	Underwood ⁶ 2005	10	Yes: $p < .05$			
		Press button	Live, staged	Shahar ¹⁸ 2010	13	Task comparison No: $p > .10$			
		Touch screen	Live, unstaged	Johnston ⁴³ 2016	64	NR	85% + hit rate for drivers, 15% false alarm rate	42	No: $p = .27$
Sensitivity		Press button	Live, unstaged	Sagberg ⁷ 2006	Block A 15 Block B 16	No: $p > .05$			
		Press button	Live, unstaged	Wallis ⁸ 2007	27	NR	90% response rate	12	No: $p > .05$

(continued on next page)

Table 3 (continued)

Methodology	Dependent variable	Hazard response type	Stimulus type	Study	Initial clip or hazardous event quantity (n)	Difference between participant groups or methodology	Inclusion/exclusion criteria for amending clip or hazardous event quantity	Post hoc clip or hazardous event quantity (n)	Difference between road user group or methodology	
Video (cont'd)	Sensitivity (cont'd)	Press button	Live, staged	Borowsky ¹⁶ 2010	4 clips/23 hazardous events	NR	High support events – ≥ 30% response rate or low support event with ≥ 20% sensitivity	Pedestrian (n = 4), car following (n = 2), intersection (n = 5)	No: p > .05 (all hazard types)	
		Press button Questionnaire (WWW)	Live, unstaged Live, unstaged	Borowsky ²⁵ 2012 Ventsislavova ⁴⁶ 2016	63 24	No: p > .05 NR	Discrimination indice less than 0.20	20	Yes: p = .001 (experience), Yes: p = .012 (offender status) Task 1 Yes: p < .001 Task 2 Yes: p < .05 Drivers in Task 1 compared to task 2 No: p = .41	
	Hazard perception score	Press button and mouse click	Animated	Vlakveld ³⁵ 2014	13 in each of 2 tasks	NR	NR	Less than 7% response rate	12 in each of 2 tasks	
		Questionnaire (WWW)	Live, staged and unstaged	Jackson ¹³ 2009	25 in each task	Driver comparison Yes: p = .04 (cut to black task)				
	Video (cont'd)	Hazard perception score (cont'd)	Questionnaire (WWW)	Live, staged and unstaged	Jackson ¹³ 2009	25 in each task	Task comparison Yes: p = .02 (novices) No: p > .05 (experienced)			
			Questionnaire (WWW)	Live, unstaged	Castro ³³ 2014	28	NR	Correlation-item total and total score r < .30	17	Yes: p = .049, hazardous clips clips Yes: p = .01, quasi-hazardous clips Yes: p = .003
		Cautious-ness	Questionnaire (WWW)	Live, unstaged	Castro ⁴¹ 2016	32	NR	Discrimination indice < than 0.30	21	
			Eye tracking	Live, staged and unstaged	Grundall ⁴² 2016 a	40	Yes: p < .001			
		Time to first fixation	Questionnaire (WWW)	Live, staged and unstaged	Grundall ⁴² 2016 b	40	Yes: p < .005			
			Questionnaire (WWW)	Live, staged and unstaged	Grundall ⁴² 2016 c	40	Yes: p < .001			
Fixation duration		Questionnaire (WWW)	Live, unstaged	Ventsislavova ⁴⁶ 2016	24	NR	Discrimination indice < 0.20	20	Yes: p = .001 (experience) No: p = .051 (offender status)	
		Questionnaire (WWW)	Live, unstaged	Gugliotta ⁴⁷ 2017	16	NR	Discrimination indice < than 0.30	11	Non-offenders Yes: p = .003, Offenders No: p = .597 Yes: p = .001 (experience), p = .044 (offender status)	
Number of fixations		Questionnaire (WWW)	Live, unstaged	Ventsislavova ⁴⁶ 2016	24	NR	Discrimination indice < 0.20	20		
		Eye tracking	Live, unstaged	Lim ³¹ 2013	40	No: p > .05 No: p = .07				
Horizontal variance of fixation	Eye tracking	Live	Chapman ³ 1998	13						
	Eye tracking	Live, unstaged Live, unstaged Live, unstaged	Underwood ⁶ 2005 Borowsky ²⁵ 2012 Borowsky ²⁴ 2012	10 63 63	No: p > .05 No: p > .05 Yes: p < .05					
Video (cont'd)	Horizontal variance of fixation	Eye tracking	Live, staged	Borowsky ¹⁶ 2010	4 clips/23 hazardous events	NR	High support events – ≥ 30% response rate or low support event with ≥ 20% sensitivity	Pedestrian (n = 4), car following (n = 2), intersection (n = 5)	Yes: p < .05 (car-following, intersection-related events) No: p > .05 (pedestrian events)	

(continued on next page)

Table 3 (continued)

Methodology	Dependent variable	Hazard response type	Stimulus type	Study	Initial clip or hazardous event quantity (n)	Difference between participant groups or methodology	Inclusion/exclusion criteria for amending clip or hazardous event quantity	Post hoc clip or hazardous event quantity (n)	Difference between road user group or methodology
Static images	Response time	Press button	Image taken from live footage	Huestegge ¹⁷ 2010	90	NR	Low detection scenes	NR	Yes: $p < .001$
		Touch screen	Image taken from live footage	Scialfa ²⁸ 2012	120	Yes: $p = .006$	85% response rate and experienced 200 ms faster than novice	21	Yes: $p = .001$
	Accuracy	Press button Touch screen	Animated Image taken from live footage	Feng ⁴⁹ 2018 Scialfa ²⁸ 2012	96 120	Yes: $p = .045$ NR	85% response rate and experienced 200 ms faster than novice	21	Yes: $p = .004$
	Total responses	Press button Press button	Animated Image taken from live footage	Feng ⁴⁹ 2018 Huestegge ¹⁷ 2010	96 90	Yes: $p < .001$ No: $p > .05$			
Static images (cont'd)	Time to first fixation	Eye tracking	Image taken from live footage	Huestegge ¹⁷ 2010	90	NR	Low detection scenes	NR	No: $p = .19$
	Time from fixation until hazard response	Eye tracking	Image taken from live footage	Huestegge ¹⁶ 2010	90	NR	Low detection scenes	NR	Yes: $p = .002$
	Number of fixations	Eye tracking	Image taken from live footage	Huestegge ¹⁷ 2010	90	NR	Low detection scenes	NR	No: $p > .05$
	Fixation duration	Eye tracking	Image taken from live footage	Huestegge ¹⁷ 2010	90	NR	Low detection scenes	NR	No: $p > .05$
Video and static image	Response time	Touch screen	Live, unstaged and image taken from live footage	Wetton ¹⁹ 2010	68 ACT HPT 64 ACT Change detection task	Yes: $p = .009$ (ACT HPT) Yes: $p = .04$ (ACT Hazard detection task)			
		Touch screen	Live, unstaged and image taken from live footage	Wetton ¹⁹ 2010	68 ACT HPT 64 ACT Change detection task	Correlation Yes: $ps \leq .001$ (ACT HPT and QLD, UK HPT) No: $ps \geq .77$ (ACT Hazard detection task and ACT, QLD and UK HPT)			
	Hazard perception score	Touch screen	Live, unstaged and image taken from live footage	Scialfa ³² 2013	100 static images 65 video clips	Correlation No: $p = .074$	85% hit rate 15% + false alarm rate	53 static Images 46 dynamic clips	Correlation No: $p = .056$
	Response time	Press button	Animated	Mackenzie ³⁶ 2015	6	Task comparison Yes: $p = .042$	> 70% response rate	3 hazardous events	Yes: $p < .001$
Video and simulator	Response time (cont'd)	Press button	Live, staged	Bromberg ²⁶ 2012	8 (video only)	NR			
	Accuracy	Press button	Animated	Mackenzie ³⁶ 2015	6	Task comparison No: $p = .20$			
	Total responses	Press button	Live, staged	Bromberg ²⁶ 2012	8 (video only)	Yes: $p < .05$ (4 out of 8 clips)	50% response rate, experienced pressed + 3 more than elderly	8 pedestrian-related events	No: $p > .05$ (5 out of 8 pedestrian related events)
	Horizontal variance of fixation	Braking response	Live, staged	Bromberg ²⁶ 2012	6 (simulator)	NR	Sufficient braking response	3 hazardous events	No: $p > .05$ (2 out of 3 events)
Video and simulator (cont'd)	Eye tracking	Eye tracking	Animated	Mackenzie ³⁶ 2015	6	NR	Courses with no hazards	4	Yes: $p \leq .001$
	Vertical variance of fixation	Eye tracking	Animated	Mackenzie ³⁶ 2015	6	NR	Courses with no hazards	4	Yes: $p \leq .001$

(continued on next page)

Table 3 (continued)

Methodology	Dependent variable	Hazard response type	Stimulus type	Study	Initial clip or hazardous event quantity (n)	Difference between participant groups or methodology	Inclusion/exclusion criteria for amending clip or hazardous event quantity	Post hoc clip or hazardous event quantity (n)	Difference between road user group or methodology
Simulator	Response time	Braking response	Live, staged	Yeung ⁴⁰ 2015	4	No: $p > .05$	Not interpretable	14	Yes: $p < .001$
	Risk recognition	Eye tracking	Animated	Pradhan ⁵ 2005	16	NR			
	Time to first fixation	Eye tracking	Animated	Grundall ²⁷ 2012	9	Yes: $p < .001$			
	Sensitivity	Eye tracking	Live, staged	Yeung ⁴⁰ 2015	4	Yes: $p < .05$ (3 of 4 scenarios only)			
	Fixation duration	Eye tracking	Animated	Grundall ²⁷ 2012	9	Yes: $p < .05$			
	Fixation duration (cont'd)	Eye tracking	Animated	Grundall ²⁷ 2012	9	Yes: $p < .001$			
Simulator (cont'd)	Fixation duration	Eye tracking	Live, staged	Yeung ⁴⁰ 2015	4	No: $p > .05$			
	Time to first fixation	Eye tracking	Live, staged	Yeung ⁴⁰ 2015	4	Yes: $p < .05$ (2 of 4 scenarios only)			
	Horizontal variance of fixation	Eye tracking	Live, staged	Yeung ⁴⁰ 2015	4	Yes: $p < .05$ (2 of 4 scenarios only)			
Simulator and test drive	Response time	Lever	Live video, staged (simulator)	Quimby ¹ 1981	60		Response within four seconds of hazard, and 90% response rate	7	Age comparison Yes: $p = .025$ (simulator)
	Total responses	Lever	Live video, staged (simulator)	Quimby ¹ 1981	60 (video) and 45 (test -drive)	Correlation with driver errors on test-drive Yes: $p < .05$			
Test-Drive	Total responses	Verbal identification	Hazards on test drive	Quimby ¹ 1981	60 (video) and 45 (test -drive)	Correlation with driver errors on test-drive Yes: $p < .05$			
	Safety index	Braking response and speed	Hazardous traffic or locations, on real-world test drive	Quimby ¹ 1981	45	Age comparison Yes: $p < .025$ (test drive)			
Test-Drive	Total response	Driver behaviour (braking, speed change)	Live staged hazards	Lee ¹⁰ 2008	3	Yes: $p < .0001$			
	Fixation	Eye tracking (based on glances seen in video of driver behaviour)	Live staged hazards	Lee ¹⁰ 2008	3	Yes: $p < .0002$			

Note. Differences that were significantly different ($p < .05$) were in the expected direction with novice/young/inexperienced drivers poorer at hazard perception than experienced/older drivers, unless otherwise stated; Differences that were not statistically significant ($p > .05$) are in bold; Press button includes space bar responses; NR, not reported; Response time, time from hazard onset until driver response; Sensitivity, number of participants in each group who responded to an event divided by the group population; ACT HPT, Australian Capital Territory Hazard Perception Test; UK HPT, United Kingdom Hazard Perception Test; QLD HPT, Queensland Hazard Perception Test; WWW, What? Where? What next? questionaire; Safety index, difference between forward visibility and overall calculated stopping distance in response to hazard.

Table 4
Results for studies that include hazard perception test methodologies for motorcyclists.

Methodology	Dependent variable	Hazard response type	Stimulus type	Study	Initial clip or hazardous event quantity (n)	Difference between participant groups or methodology	Inclusion/exclusion criteria for amending clip or hazardous event quantity	Post hoc clip or hazardous event quantity (n)	Difference between road user group or methodology
Video	Response time	Press button	Live, unstaged	Belle ²⁴ 2012	25	Yes: $p < .001$	Responses before a priori hazard onset	14	Yes: $p = .051$
		Press button	Live, unstaged	Crundall ²⁹ 2013	14	No: $p > .05$	Responses before a priori hazard onset	14	No: $p > .05$
	Accuracy	Verbal identification	Live, unstaged	Crundall ²⁹ 2013	14	Yes: $p < .05$		14	
	Total responses Situation criticality	Press button Verbal assessment	Live, unstaged	Crundall ²⁹ 2013 Belle ²⁹ 2012	14 25	No: $p > .05$ Yes: $p < .001$			
Locus of control over hazard	Questionnaire on locus of control	Live, unstaged	Crundall ²⁹ 2013	14	Yes: $p < .01$ internal locus Yes: $p < .05$ external locus No: $p > .05$ general measure of locus of control Yes: $p < .05$				
Simulator	Hazard perception score	Mouse click	Live, unstaged	Rosenbloom ²¹ 2011	10	Yes: $p < .05$			
	Response time	Press button	Live, staged and unstaged	Horswill ¹⁴ 2003	8	Yes: $p = .028$			
Simulator (cont'd)	Time to first fixation	Eye tracking	Animated	Cheng ²⁰ 2011	1	Yes: $p < .001$			
	Crashes	Rider behaviour (resulting in crash)	Animated	Liu ¹⁴ 2009	24 (3 scenarios only)	Yes: OR 2.9 [CI 1.05-8.9] ("Touring" scenario only)			
	Rider performance	Rider behaviour (computed by simulator)	Animated	Liu ¹⁴ 2009	16 ("Path" and "Avenue" scenarios only)	Yes: OR 2.6 [CI 1.19-5.66] ("Avenue" scenario only)			
	Speed	Acceleration/ deceleration	Animated	Liu ¹⁴ 2009	16 ("Avenue" and "Touring" scenarios only)	Yes: CI 0.7-13.2 and CI 2.6-20.4 ("Avenue" scenario only)			

Note: Differences that were significantly different ($p < .05$) were in the expected direction with novice/young/inexperienced drivers poorer at hazard perception than experienced/older riders; Differences that were not statistically significant ($p > .05$) are in bold; Press button includes space bar responses; OR, odds ratio; CI, confidence intervals; Locus of control, internal or external attribution as to why the hazard occurred; Situation criticality, subjective assessment of risk and self-evaluation of ability to manage risk and avoid a crash.

Table 5
Results for studies that include hazard perception test methodologies for bicyclists.

Methodology	Dependent variable	Hazard response type	Stimulus type	Study	Initial clip or hazardous event quantity (n)	Difference between participant groups or methodology	Inclusion/exclusion criteria for amending clip or hazardous event quantity	Post hoc clip or hazardous event quantity (n)	Difference between road user group or methodology
Video	Response time	Mouse click	Live, unstaged	Vansteenkiste ⁴⁵ 2016	13	No: $p > .05$ (7 out of 13 clips) Yes: $p < .001$ (7 out of 11 clips)			
	Accuracy	Verbal identification	Live, unstaged	Zeuwts ⁴⁸ 2017	11	No: $p = .57$			
	Total response	Mouse click	Live, unstaged	Vansteenkiste ⁴⁵ 2016	13	No: $ps > .05$ (for 10 out of 11 clips) No: $p > .05$ Yes: $p = .049$ (car clips only) No: $p = .552$			
	Number of fixations	Eye tracking	Live, unstaged	Zeuwts ⁴⁸ 2017	11	Yes: $p < .001$			
	First fixation	Eye tracking	Live, unstaged	Zeuwts ⁴⁸ 2017	11	No: $p > .05$			
	Fixation duration	Eye tracking	Live, unstaged	Vansteenkiste ⁴⁵ 2016	13	Yes: $p < .001$			
		Eye tracking	Live, unstaged	Zeuwts ⁴⁸ 2017	11	Yes: $p < .001$			

Note: Differences that were significantly different ($p < .05$) were in the expected direction with novice/younger cyclists poorer at hazard perception than experienced/older cyclists; Differences that were not statistically significant ($p > .05$) are in bold.

other road users that were hidden from view, with younger cyclists slower than older cyclists for 7 out of 11 clips.

Spatial measures of accuracy (Vansteenkiste et al.⁴⁵, 2016) and total responses (Zeuwts et al.⁴⁸, 2017) suggest there were no differences between younger and older cyclists with both equally able to correctly identify the hazards. Eye tracking was used in both studies however no differences were found in number of fixations, with mixed findings for fixation duration.

Both studies used video filmed from a bicyclist's perspective, with similar instructions to press the button when detecting a hazard that required action such as braking or change of direction to avoid a crash. There were no comparisons between methodologies or post hoc amendment of clips.

3.6. Hazard perception testing of pedestrians

There were only three studies^{34,38,39} testing hazard perception of pedestrian populations (see Table 6) with two using video and one a simulator, and all using a button press response type. There were methodological differences between these studies with the use of clips filmed from a pedestrian's perspective^{38,39} and from a car driver's perspective³⁴. The definition of a hazard varied between studies from a situation that may prevent crossing the street as a pedestrian³⁸ to a hazard that required an abrupt braking or avoidance manoeuvre as a car driver³⁴.

Results for temporal measures were largely in the expected direction with younger pedestrians slower to respond to hazards than older pedestrians, however this was dependent on scenario type. Meyer et al.³⁴ (2014) used a hazard perception test designed for drivers and found younger children as pedestrians ($M_{age} = 10.4$) took significantly longer to respond to 6 out of 10 hazard scenarios than older children as pedestrians ($M_{age} = 14.2$) and adults ($M_{age} = 43.4$). This difference was particularly significant for those scenarios that involved vulnerable road users such as a pedestrian at a crossing that is hidden by an on-coming vehicle. Meir et al.³⁷ (2015) also found differences in response times between age groups for scenarios where parked vehicles limited the field of view, but found no differences when the scenario included a visible pedestrian crossing, moving vehicles or where the curvature of the road limited the field of view.

Spatial responses as measured by sensitivity, or accuracy as part of a total hazard perception score, were inconsistent and dependent on scenario type. Meyer et al.³⁴ (2014) found 7 out of 10 scenarios had no difference in response rate with younger and older pedestrians equally able to identify the hazard. Younger pedestrians were significantly poorer at identifying hazards when they involved other vulnerable road users such as pedestrians and cyclists, including scenarios where a pedestrian at a crossing is hidden from view. In contrast, Meir et al.³⁸ (2015) found younger pedestrians were less likely to report hazards than older pedestrians for scenarios that included a pedestrian crossing, where there were moving vehicles, and when the field of view was limited by parked cars or curvature of the road. Rosenbloom et al.³⁹ (2015) found hazard detection scores were in the expected direction with younger pedestrians less accurate at identifying hazards for a range of road crossing hazard scenarios, for which the majority included a pedestrian crossing.

There were no methodological comparisons, or post hoc amendments of clips.

4. Discussion

This systematic review was the first to synthesise the literature on hazard perception test methodologies across a range of road user groups – car drivers, motorcyclists, bicyclists and pedestrians. The aim was to examine the ability of test methodologies – video, static image, simulator and real-world test-drive, to discriminate between participants based on age and/or experience. The results demonstrated considerable

Table 6
Results for studies that include hazard perception test methodologies for pedestrians.

Methodology	Dependent variable	Hazard response type	Stimulus type	Study	Initial clip or hazardous event quantity (n)	Difference between participant groups or methodology	Inclusion/exclusion criteria for amending clip or hazardous event quantity	Post hoc clip or hazardous event quantity (n)	Difference between road user group or methodology
Video	Response time	Press button	Live, unstaged	Meyer ³⁴ 2014	10	Yes: $p < .028$ (6 of 10 clips, majority include vulnerable road users) No: $p > .05$ (7 out of 10 clips)			
	Sensitivity Hazard detection score	Press button Press button	Live, unstaged Live, staged	Meyer ³⁴ 2014 Rosenbloom ³⁹ 2015	10 10	Yes: $p < .001$ (7 out of 10 clips including pedestrian crossing scenarios) No: $p > .05$ (zebra crossing; moving vehicles; road curvature limits field of view)			
Simulator	Response time	Press button	Animated	Meir ³⁸ 2015	18 (4 groups of clips)	Yes: $p < .001$ (parked car limits field of view) Yes: $p \leq .05$ (zebra crossing; presence of moving vehicles; parked cars and road curvature limits field of view)			
	Sensitivity	Press button, track-ball marker, written identification	Animated	Meir ³⁸ 2015	18 (4 groups of clips)	No: $p > .05$ (no-zebra-crossing; no moving vehicles)			

Note: Differences that were significantly different ($p < .05$) were in the expected direction with young pedestrians poorer at hazard perception than older pedestrians; Differences that were not statistically significant ($p > .05$) are in bold; Press button includes space bar responses; Sensitivity, number of participants in each group who responded to an event divided by the group population; Hazard detection score, final score (sum of scores based on when hit occurs within hazard response window – higher scores for early responses) divided by potential hazards in clip.

heterogeneity in the methodologies reviewed, with inconsistencies across a range of findings. However there was support for all methodologies, based on their ability to discriminate between participant groups on at least one measure of hazard perception. The majority of findings were in the direction of novice/younger road users being significantly poorer at hazard perception than experienced/older road users. Of the remaining studies, the majority found no significant differences between road user groups based on age and/or experience.

4.1. Methodological review

The review revealed that the most commonly used methodology was videos using live footage. The majority were unstaged and filmed from the road users perspective; and for car drivers, this format matches that used in driver licensing tests. However, both unstaged and staged videos were equally able to discriminate between novice/younger and experienced/older road users. Given the hours of footage required to capture unstaged traffic hazards, the findings of this review suggest that there may be benefits to using staged footage to test hazard perception skills, allowing for a wider range of scenarios to be examined. Static image tests were less commonly used and only studied in car driving populations. A comparison of static image and video tests found both were able to discriminate between drivers based on age and experience. However, differences in response times for a static image hazard detection test were in the opposite direction to that predicted in that novice/younger drivers were faster than experienced/older drivers, while response times for a video test were in the expected direction (Wetton et al.¹⁹, 2010). Given the increased crash risk for novice/younger drivers, this lends support for the argument that static image tests may not be representative of the dynamic nature of real-world driving, and lack ecological validity to measure hazard perception. However, both methodologies were correlated with self-reported driver errors in young adult drivers (Scialfa et al.³², 2013). Furthermore, there was no correlation between static image and video tests suggesting these methodologies may be tapping into different aspects of hazard perception (Wetton et al.¹⁹, 2010). Static image tests have the benefit of explicit response times that can be calculated from stimulus onset while videos contain subjectively pre-determined hazard windows that may fail to capture early responses. The brevity of responding to a static image could also reduce testing time and allow a larger variety of hazardous scenarios to be tested. Licensing tests may benefit from using a combination of video and static images, given both are able to discriminate between road user groups based on age and/or experience, and this combined approach could provide a more global assessment of hazard perception skills.

Compared to video and static images, driving simulators and motorcycle simulators can be considered to better emulate the on-road experience by introducing the physical and cognitive demands of safe driving. However, the use of simulators is largely restricted to research and this is likely due to the impracticality of mass testing as well as the costs associated with set up. However, a comparison of video and simulator methodologies in a younger driver population found hazard perception was slower in the simulator method compared to the video method, suggesting the simulator may more accurately reflect the greater cognitive load experienced while driving (Mackenzie et al.³⁶, 2015). This has implications for hazard perception tests based on the video methodology that require appropriately timed responses to hazardous scenarios in order to pass the test. Whilst this finding suggests that video tests may overestimate the temporal response to hazards, it does not necessarily invalidate the method; which is based on the ability to discriminate between road user groups of different ages and experience.

The most naturalistic method of measuring hazard perception skills is to use real-world test-drives. However as with simulators, test-drives are costly, time consuming and there is poor control of the hazards and hazard types that may present themselves. Importantly, the present

review identified that Quimby and Watts¹ (1981)'s study, in which poor results on a hazard perception test correlated with errors on a test drive and were significantly related to age, has never been replicated. Given the assumption that the hazard perception tests can discriminate between novice/younger and experienced/older drivers, it may be timely to investigate the relationship between video test performance and on-road driving to ensure this underlying premise remains sound.

Across the four methodologies, there was a range of response types that captured temporal, spatial and decision-making responses. Temporal responses alone do not identify the hazard unless it is unambiguous, so the addition of spatial responses provided a measure of accuracy. However, the majority of studies found no difference in spatial accuracy between novice/younger and experienced/older participants, suggesting an equal ability to identify hazards despite differences in response times. The review suggests that measures of spatial accuracy add little value over and above temporal measure in discriminating between novice/younger and experienced/older road users. However, the extension of spatial measures to include decision-making was found to discriminate between novice/younger and experienced/older drivers, while providing an additional measure of hazard perception.

Based on the theoretical framework of situation awareness (Endsley, 1988; Endsley, 1995), questionnaires that probed the deeper understanding of the hazard were able to successfully discriminate between road user groups, despite not providing the traditional temporal response. The What? Where? What next? (WWW) format was used exclusively in car driver studies and had the unique ability to explore decision-making, a crucial aspect of safe driving once a hazard is perceived. A recent study published after the search date directly compared a traditional hazard perception test that captured accuracy and response times using a mouse click, and a hazard prediction test based on the question 'What happens next?' (Ventsislavova et al., 2019). In this hazard prediction test, the SAGAT was implemented where the clips were occluded prior to the hazard developing (Endsley, 1988). Participants provided a short answer that describes the hazard, its location and how the situation would develop. Results suggest that the WWW test was better at discriminating between novice and experienced drivers than the hazard perception test (Ventsislavova et al., 2019). While the use of open-ended questions has limitations for mass testing due to the time and cost involved in marking responses, studies have demonstrated that implementing a WWW test through a multiple-choice test and response time test can equally discriminate between novice/younger and experienced/older drivers (Malone and Brünken³⁷, 2015; Malone and Brünken⁴⁴, 2016). This was supported by a recent study published after the search date, in which a multiple-choice and free-response hazard prediction test based on SAGAT were equally able to discriminate between drivers based on experience (Ventsislavova and Crundall, 2018). This suggests that further research is required to better understand the role of hazard prediction and its association with crash risk, and how this can best be assessed using questionnaires. There may be benefits in including questionnaires as part of driver licensing to extend the testing beyond response times and assess decision-making.

The locus of control questionnaire was used in a single motorcyclist study and uniquely measured the rider's internal and external attributions for why the hazard occurred (Crundall et al.²⁹, 2013). The overall measure of locus of control questionnaire failed to discriminate between novice, experienced and advanced riders, however internal attributions were shown to increase with riding experience with a corresponding decrease in external attributions. Studies into the relationship between locus of control and crash risk have been inconsistent. Özkan and Lajunen (2005) found greater internal attributions were associated with more self-reported crash and violations and may reflect overconfidence. In contrast, internal attributions were associated with improved driver behaviour on a real-world test drive following defensive driver training (Huang and Ford, 2012). These inconsistent results suggest future research is needed to better understand the

relationship between locus of control and hazard perception in road users, and this may improve hazard perception training programmes.

The inclusion of eye tracking was expected to provide an additional understanding of visual scanning patterns with novice/younger road users having a narrower field of view and scanning less often than experienced/older road users. However, results for fixation duration, number of fixations, and time to first fixation were inconsistent. Further, from a testing perspective, the use of eye tracking requires additional time and training in order to calibrate. [Huestegge and colleagues¹⁷ \(2010\)](#) also highlight the technological issues with eye tracking, in that 12.75% and 14.94% of novice and experienced driver trials respectively, were discarded due to poor quality eye movement data. There may be alternative approaches to measuring the field of view such as that used by [Shahar et al.¹⁸ \(2010\)](#). Specifically, videos of hazardous scenarios were viewed on a traditional single computer screen, or on three computer screens, one centrally located and two laterally positioned, that increased the field of view. The three-screen condition encouraged wider and more frequent scanning, with the lateral screens contributing to faster average response times. The use of multiple screens to increase the field of view and encourage more efficient visual search strategies may be an alternative to eye tracking when considering mass testing and improvements to driver licensing tests.

4.2. Methodology by road user group – drivers

The majority of studies were in car driving populations and the four methodologies (video, static image, simulator and real-world test drive) were equally able to discriminate between novice/younger and experienced/older drivers on at least one measure of hazard perception. While the video methodology matches the format currently used in most hazard perception testing for driver licensing, the findings from this review suggest there may be opportunities to enhance testing of novice drivers. This could be through the inclusion of static images, questionnaires that capture a deeper understanding of hazards, and widening the field of view to assess visual search strategies. While simulators may not be feasible for use in mass testing due to cost, there may be an argument for increasing the cognitive load while testing e.g. distractors or dual tasks to better replicate the challenges when perceiving hazards during real-world driving.

4.3. Methodology by vulnerable road users

Studies of motorcyclist populations used video and simulator methodologies, with a consistent pattern of results to those in car drivers, and were in the expected direction. While the limited number of studies suggests further research is needed, methodologies that included a temporal measure such as video tests were equally discriminative of motorcyclists as car drivers, with novice/younger riders slower at hazard perception than experienced/older riders. Furthermore, the assessment of locus of control while limited to a single study, found a shift from external to internal attributions with increasing age and riding experience. Given the known reduction in crash risk for novice car drivers that pass a hazard perception test as part of the licensing process, the studies in this review suggest there is an opportunity to use a similar video test with temporal responses to assess hazard perception skills in novice motorcyclists. This could be enhanced with questionnaires that explore the role of attributions in rider behaviour.

Hazard perception testing as part of the motorcyclist licencing process is not universally implemented, and where used, it is the same test as that used for driver licensing ([United Kingdom Government, 2018](#); [Austroads, 2018a,b](#)). Traffic authorities in Australia have recommended that a motorcycle-specific hazard perception test be implemented as part of a GDL system, however to date, Western Australia is the only state in Australia to implement this recommendation

([Austroads, 2018a,b](#)). Motorcycling licensing schemes may also benefit from assessing hazard perception skills using scenarios from the motorcyclist's perspective to increase face validity.

The two studies of hazard perception in bicyclist populations used the video methodology with eye tracking. Both studies used video filmed from the cyclist's perspective which was displayed on a computer screen, and with consistency in instructions to respond to hazards that require the cyclist to brake or change direction. However, these studies failed to discriminate between novice/younger and experienced/older bicyclists on at least one measure of hazard perception. For selected clips containing covert or latent hazards, results were in the expected direction with older bicyclists faster at hazard perception and with more efficient visual search strategies compared to younger bicyclists. However the inconsistency in results across a range of dependent variables suggests the current methodologies are limited in their ability to assess hazard perception skills in bicyclist populations.

The studies of pedestrian populations used the video and simulator methodology, and as with bicyclist studies there were inconsistencies across a number of dependent variables. In particular, there were considerable differences in the type of scenarios and instructions for responding to a hazard. One study used clips filmed from the driver's perspective with only two scenarios featuring a pedestrian crossing ([Meyer et al.³⁵, 2014](#)). This study required participants to press the space bar when they detected a dangerous situation, to represent braking or an avoidance manoeuvre. However in contrast, [Rosenbloom et al.³⁹ \(2015\)](#) utilised videos filmed from a pedestrian perspective with a focus on hazards that may prevent crossing the road. While both studies found younger participants were generally slower at hazard perception than older participants, it could be argued the inconsistencies in scenario selection and instructions for responding may limit findings with regard the hazard perception ability of this road user group. Further inconsistencies were found in the study by [Meir et al.,³⁸ \(2015\)](#), in which there was no difference in response times between younger and older participants. This study attempted to replicate the pedestrian experience by viewing animated videos projected onto a dome while standing upright, and gave a 180° field of view from the pedestrian's perspective rather than the traditional view on a computer screen. The lack of response time differences may have been influenced by the task instructions for younger participants who were asked to press the button when they identified a hazard that may prevent them crossing the road. Rapid response times that were similar to older participants may reflect crossing decision-making and cautious behaviour near roads, rather than the detection of a hazard. Instructions to younger participants have been shown to bias responses in a road risk task ([Underwood et al., 2007](#)). Despite these inconsistencies, younger participants had difficulty in recognising hazards that were hidden, those involving pedestrian crossings and hazard scenarios that included other vulnerable road users. Given 54% of road fatalities worldwide involve pedestrians, bicyclists and motorcyclist, findings from this review suggests more research is required to develop test methodologies that better capture hazard perception in these populations ([World Health Organisation, 2018](#)). This may lead to opportunities for training and assessment that will improve outcomes for these road user groups.

4.4. Methodological limitations and inconsistencies

While this review found support for all methodologies to capture differences between participant groups based on age and experience, a number of studies failed to find a difference in the key measure of response time. An examination of these studies highlighted key inconsistencies, namely with the categorisation of participants by age and experience, and the range of different hazard types included in the scenarios. The problems created by varied categorisation of road users was highlighted in a cross-cultural study by [Lim et al.³¹ \(2013\)](#), who found no difference in response times. In this study, experienced drivers had an average of four years of driving experience, which is much lower

than the majority of studies for which this level of experience would have classified the drivers as novices. Across all studies, the mean years since licensing for novices ranged from 0.1–4.5 years, and 2–29 years since licensing for experienced drivers. There were also broad and overlapping ranges for categorising participants based on age, with young drivers aged from 17–49 years and older drivers aged 47–69 years. Given the interrelated effect of age and experience, these inconsistencies create challenges when comparing methodologies across studies. More importantly, given that the findings are based on such broad and overlapping age and experience categorisations, it may be possible that studies or measures are not capturing true between-group differences. The fundamental basis of hazard perception testing is to discriminate between those that have developed appropriate hazard perception skills and those that have not. It is important to note that road user experience and cognitive maturation that comes with age both play a critical role in the development of hazard perception skills. In order to determine if categorising participants by age or experience is the best determinant of the effectiveness of hazard perception testing, studies need to re-examine these definitions in accordance with driving performance safety levels. Furthermore, given the wealth of knowledge that exists around the developmental and experiential stages of road users, it seems feasible to establish a gold standard for age and experience categorisation that would enable comparisons between studies to better determine best practice, not only for hazard perception testing, but for other road user interventions as well as policy decisions.

Across the 49 studies, there was a range of classifications used to categorise hazardous scenario types. This included materialised hazards requiring immediate evasive action and potential hazards that included an anticipatory cue. Potential hazards have been further categorised, for example, into overt or covert, based on behaviour (i.e. a pedestrian walking towards the road) or the environment (i.e. parked truck obscuring the pedestrian). The problem of varied scenario type was highlighted in a study by Sagberg and Bjørnskau⁷ (2006), in which there were no difference in overall response times between novice and experienced drivers. However, a detailed clip analysis found novice drivers were significantly slower than experienced drivers for complex scenarios and hazards that required immediate evasive action, but not potential hazards. This was supported by Malone and Brünken⁴⁴ (2016) and suggests that highly ambiguous scenarios, where the hazard may or may not develop, are equally challenging to identify regardless of age or experience. Furthermore, Borowsky et al. (2010, 2012) found no differences in hazard perception for scenarios that involved detecting pedestrians and bicyclists, which suggests consistency across driver groups in the rapid detection of vulnerable road users. These findings suggest scenario type plays an important role in the discrimination between road user groups of different age and experience. Further research is needed to determine which types of hazardous scenarios serve this purpose best, and how these translate to safety on the road. For car drivers, this could enhance driver-licensing tests and be extended to hazard perception testing of other road users.

There were also inconsistencies in the selection of scenarios with one third of studies conducting a post hoc amendment that altered the final number of scenarios to be tested or analysed. Post hoc amendments saw the removal of clips that failed to discriminate between road user groups, using various inclusion/exclusion criteria. The majority of studies failed to report findings for a priori scenario selections. Furthermore, the study by Crundall et al.²⁹ (2013) found a difference in response times between novice and experienced motorcyclists only after a post hoc amendment of the hazard response window to allow for earlier responses from the more experienced riders. While the use of cut-off scores is common practice in test development, the lack of consistency in the criteria used has the potential to bias results by only including clips that elicit responses in the hypothesised direction. For the clips that are discarded, there is little understanding of why the scenario failed to find a difference between road user groups and how this relates to road safety. For example, the lack of discriminative

ability of a scenario may be because it is difficult for both novice/younger and experienced/older road users to perceive the hazard, resulting in comparable response times. This scenario may represent a crucial skill for road safety regardless of age and experience. Further research is needed to determine which scenarios are better at discriminating between novice/younger and experienced/older road users and how these relate to crash risk.

In addition to the inconsistencies in classifying and selecting scenarios for hazard perception testing, this review has highlighted the challenges in accurately capturing hazard perception using a range of hazard response types. Is the button press, mouse click or eye fixation for example, accurately assessing the participant's ability to detect hazards in the scenario or other critical elements within their visual field of view? How well does this response type reliability measure and assess the skills that are necessary to respond to real-world traffic events in order to avoid a crash? Future research is needed to better understand how response types that measure hazard perception through temporal, spatial or decision-making information are associated with accurate real world hazard perception and crash risk. This research could inform current hazard perception testing for licensing, particularly in assessing decision-making after a hazard is detected.

Another area of inconsistency is in the application of theory or models of driver behaviour in the development of hazard perception tests. Less than one third of studies cited a theoretical foundation. The benefit of applying a theoretical framework can be seen in the development of questionnaires in the WWW format that changed the direction of test methodologies away from measuring a response time to capturing differences in decision making once the hazard has been detected. Given the inconsistencies outlined previously, future studies would benefit from incorporating a theoretical framework in test development to ensure methodologies can also capture the cognitive processes involved in hazard perception.

4.5. Limitations at the outcome level

There were a number of shared limitations across the studies reviewed. The considerable heterogeneity in the studies limits the comparison of methodologies in order to empirically determine the most effective. For many studies both age and experience of participants were not reported, making it difficult to compare samples. The inconsistent and somewhat arbitrary categorisation of participant groups in a research context may therefore bias results.

There is also the risk of bias due to the post hoc removal of scenarios that do not discriminate between participant groups. There is little understanding of the content of the discarded clips as they are often not reported, and in particular why they failed to discriminate between participant groups. Whilst the clips that are retained may have high face validity, their measurement of hazard perception performance must relate to on-road behaviour. For drivers, this leads to the question, if a novice and experienced driver had similar responses, does this indicate they are both safe drivers? That is, hazard perception tests should assess core hazard perception skills that all drivers should possess as well as hazard perception skills that only experienced drivers possess.

The lack of a theoretical framework leads to inconsistencies in the inclusion/exclusion criteria that are used to determine the final clip selection. These questions could be addressed in future research and importantly; the categorisation of participants and scenario selection should be linked to on-road hazard perception performance, guided by a theoretical framework.

4.6. Limitations at the review level

This review has several limitations in that all the studies were cross-sectional designs based on the inclusion criteria. A number of longitudinal studies were excluded as hazard perception test methodologies were used to gather pre- and post-intervention scores for training

programmes. These studies provided a measure of short-term validity and reliability, however there are a limited number of studies that have investigated the effect of hazard perception training on future hazard perception or crash risk (McDonald et al., 2015). A recent study by Thomas and colleagues (2016) found that while a hazard perception training programme reduced the crash risk of male novice drivers, it increased crash risk for female novice drivers following training. There is an opportunity for a systematic review of hazard perception training studies to determine the utility of such training, a line of research which is currently being investigated by this team. Without longitudinal studies in the area of hazard perception, the link between test methodology and its predictive validity of crash outcomes over the months and years following licensure cannot be established. There may also be publication bias in that studies that did not find a significant result may not have been published, or hazards were defined using alternative terminology such as 'risk' and therefore would not be identified using the present search strategy.

4.7. Future directions

It is important for future research to resolve the inconsistencies in the study of hazard perception testing. Specifically, there should be greater consensus around the categorization of participants by age and by experience, and this should be driven by the vast literature which identifies the decline in young novice driver's crash risk over time, and mechanisms by which this occurs. This will allow a better comparison of methodologies across studies with the purpose of determining best practice methods to discriminate those with appropriate hazard perception skills. Second, there should be consensus around the post hoc amendments of clips using an agreed exclusion and inclusion criteria, and this may remove the bias that may be present in studies that use a more stringent criterion. The issue of clip selection is complex given the number of variables within any hazardous scenario such as location, level of ambiguity or number of hazards. By discarding clips that fail to discriminate between road user groups, little is understood about those scenarios in which novice/younger road users are as skilled at hazard perception as experienced/older road users. More importantly for drivers, further research is needed to better understand how this translates to real-world driver behaviour and crash risk. The current hazard perception testing for licensing is usually based on a cut-off score with binary responses for a range of scenarios (detected hazard or did not detect hazard). There may be a case for investigating hazard perception as a continuous variable, using a range of scenarios both ambiguous and unambiguous that are commonly experienced in real traffic conditions.

The lack of correlation between static and video hazard perception tests and the ability of questionnaires to explore decision-making suggests different methodologies may be tapping into distinct aspects of hazard perception. This highlights an opportunity for future research to investigate expanding the hazard perception licensing tests to incorporate measures of different components of hazard perception in order to better discriminate between novice and experienced drivers. Furthermore, motorcycle GDLs would benefit from hazard perception testing that is specific to the motorcycle riders experience on the road. The current methodologies may not be adequately assessing the experience of bicyclist and pedestrians, and the development of new hazard perception test methodologies may provide opportunities for testing and training for these vulnerable road user populations.

4.8. Conclusion

In summary, this systematic review was the first to synthesise the literature on hazard perception testing methodologies across road users. There was consensus that all four methodologies – video, static image, simulator and test-drive, were able to discriminate between novice/younger and experienced/older road users on at least one measure of hazard perception. The majority of studies found results were in the

expected direction; with novice/younger road users poorer at perceiving and responding to hazards than experienced/older road users.

Despite a high degree of heterogeneity in the studies, this review determined that the video methodology that includes temporal responses has wide application in testing hazard perception skills across road user groups. Spatial responses were inconsistent in discriminating between novice/younger and experienced/older road users however were necessary inclusions when response types failed to identify the actual hazard, such as button press, or when scenarios were ambiguous. The review determined that staged and unstaged video footage has similar discriminability of hazard perception skill. As such, the inclusion of staged footage would allow testing of a wider range of predefined scenarios that could be empirically linked to safe driver behaviour. Eye tracking provided valuable information regarding visual search strategies however results were inconsistent in discriminating between novice and experienced drivers and the use of this technology in mass testing may be limited. The review suggested that static image tests may measure a different component of the hazard perception skill and the inclusion of these enhance current testing by examining a wider range of scenarios and hazard perception skills. While simulators and real-world test drives may represent the most naturalistic method for testing hazard perception with the added cognitive load of steering and braking, application is largely limited to research due to the time and cost associated with testing.

It is important for future research to resolve the inconsistencies around the categorisation of drivers by age and experience with the view to developing a gold standard that can be applied across studies. This will ensure true between-group differences are captured and allow for more accurate comparison of results between studies. In addition, research is needed to address the varied criteria for selecting types of hazard scenarios and how these are linked to crash risk. This should include those scenarios that failed to discriminate between road user groups but may demonstrate hazard perception skills important for road user safety. Furthermore, greater application of a theoretical framework is needed to ensure that test methodologies are effectively capturing the cognitive processes involved in hazard perception. More importantly, for driver licensing purposes, future longitudinal research is needed to better understand the relationship between hazard perception testing and its validity in discriminating between those with appropriate hazard perception skills in real-world driving and those without. It is recommended that hazard perception testing including a temporal measure should be extended to motorcyclists with the development of motorcyclist-specific hazard perception tests as part of a GDL system, that best capture the rider's experience. Finally, future research is needed to develop methodologies that assess hazard perception skills in pedestrian and bicyclist populations in order to enhance education and training that will improve outcomes for these vulnerable road users.

Conflict of interest

There is no conflict of interest to declare for this paper.

Author contributions

All of the listed authors contributed significantly to the development, research and written components of the presented manuscript. The weight of the contribution is signified by the order in which the authors names have been presented.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References¹

- *Anstey, K., Wood, J., Lord, S., Walker, J., 2005. Cognitive, sensory and physical factors enabling driving safety in older adults. *Clin. Psychol. Rev.* 25 (1), 45–65. <https://doi.org/10.1016/j.cpr.2004.07.008>.
- Australian Government, 2018. Driver's Licence Application. Retrieved from <https://www.australia.gov.au/information-and-services/transport-and-regional/drivers-licence-application>.
- Austrroads, 2018a. Australian Driver Licensing. Retrieved from: https://austrroads.com.au/drivers-and-vehicles/registration-and-licensing/australian-driver-licensing#Table_2.
- Austrroads, 2018b. A Discussion Paper on Elements of Graduated Licensing Systems for Motorcycle Riders. Retrieved from: https://austrroads.com.au/publications/road-safety/ap-r469-14/media/AP-R469-14-Elements_of_GLS_for_Motorcycle_Riders.pdf.
- Bates, L.J., Allen, S., Armstrong, K., Watson, B., King, M.J., Davey, J., 2014. Graduated driver licensing: an international review. *Sultan Qaboos Univ. Med. J.* 14 (4), e432–e441.
- *Bellet, T., Banet, A., 2012. Towards a conceptual model of motorcyclists' Risk Awareness: a comparative study of riding experience effect on hazard detection and situational criticality assessment. *Accid. Anal. Prev.* 49, 154–164. <https://doi.org/10.1016/j.aap.2011.10.007>.
- *Bromberg, S., Oron-Gilad, T., Ronen, A., Borowsky, A., Parmet, Y., 2012. The perception of pedestrians from the perspective of elderly experienced and experienced drivers. *Accid. Anal. Prev.* 44 (1), 48–55. <https://doi.org/10.1016/j.aap.2010.12.028>.
- *Borowsky, A., Oron-Gilad, T., Meir, A., Parmet, Y., 2012. Drivers' perception of vulnerable road users: a hazard perception approach. *Accid. Anal. Prev.* 44 (1), 160–166. <https://doi.org/10.1016/j.aap.2010.11.029>.
- *Borowsky, A., Shinar, D., Oron-Gilad, T., 2010. Age, skill, and hazard perception in driving. *Accid. Anal. Prev.* 42 (4), 1240–1249. <https://doi.org/10.1016/j.aap.2010.02.001>.
- Boufous, S., Ivers, R., Senserrick, T., Stevenson, M., 2011. Attempts at the practical on-road driving test and the hazard perception test and the risk of traffic crashes in young drivers. *Traffic Inj. Prev.* 12 (5), 475–482. <https://doi.org/10.1080/15389588.2011.591856>.
- *Castro, C., Padilla, J.L., Roca, J., Benítez, I., García-Fernández, P., Estévez, B., ... Crundall, D., 2014. Development and validation of the Spanish hazard perception test. *Traffic Inj. Prev.* 15 (8), 817–826. <https://doi.org/10.1080/15389588.2013.879125>.
- *Castro, C., Ventsislavova, P., Peña-Suarez, E., Gugliotta, A., Garcia-Fernandez, P., Eisman, E., Crundall, D., 2016. Proactive listening to a training commentary improves hazard prediction. *Saf. Sci.* 82, 144–154. <https://doi.org/10.1016/j.ssci.2015.09.018>.
- *Chapman, P., Underwood, G., 1998. Visual search of driving situations: danger and experience. *Perception* 27 (8), 951–964. <https://doi.org/10.1068/p270951>.
- *Cheng, A.S., Ng, T.C., Lee, H.C., 2011. A comparison of the hazard perception ability of accident-involved and accident-free motorcycle riders. *Accid. Anal. Prev.* 43 (4), 1464–1471. <https://doi.org/10.1016/j.aap.2011.02.024>.
- *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. *Accid. Anal. Prev.* 45, 600–609. <https://doi.org/10.1016/j.aap.2011.09.049>.
- *Crundall, D., van Loon, E., Stedmon, A.W., Crundall, E., 2013. Motorcycling experience and hazard perception. *Accid. Anal. Prev.* 50, 456–464. <https://doi.org/10.1016/j.aap.2012.05.021>.
- *Crundall, D., 2016. Hazard prediction discriminates between novice and experienced drivers. *Accid. Anal. Prev.* 86, 47–58. <https://doi.org/10.1016/j.aap.2015.10.006>.
- Drive Right Netherlands, 2019. What to Expect During the CBR Theory Test? Retrieved from <https://www.driverright.nl/theory/theory-test/>.
- Endsley, M., 1988. Situation awareness global assessment technique (SAGAT). *IEEE Proceedings of the National Aerospace and Electronics Conference* 789–795.
- Endsley, M., 1995. Toward a theory of situation awareness in dynamic systems. *Hum. Factors: J. Hum. Factors Ergonom. Soc.* 37 (1), 32–64. <https://doi.org/10.1518/001872095779049543>.
- *Feng, J., Choi, H., Craik, F.I.M., Levine, B., Moreno, S., Naglie, G., Zhu, M., 2018. Adaptive response criteria in road hazard detection among older drivers. *Traffic Inj. Prev.* 19 (2), 141–146. <https://doi.org/10.1080/15389588.2017.1373190>.
- Foy, H., Runham, P., Chapman, P., 2016. Prefrontal cortex activation and young driver behaviour: a fNIRS study. *PLOS ONE* 11 (5), e0156512. <https://doi.org/10.1371/journal.pone.0156512>.
- Grayson, G., Maycock, G., Groeger, J., Hammond, S., Field, D., 2003. Risk, Hazard Perception and Perceived Control. Transport Research Laboratory. Retrieved from <https://trl.co.uk/sites/default/files/TRL560.pdf>.
- Grayson, G.B., Sexton, B.F., 2002. The Development of Hazard Perception Testing (TRL Report 558). Retrieved from <https://trl.co.uk/reports/TRL558>.
- Green, D., Swets, J., 2000. *Signal Detection Theory and Psychophysics*. Peninsula Publishing, Los Altos Hills, California.
- *Gugliotta, A., Ventsislavova, P., Garcia-Fernandez, P., Peña-Suarez, E., Eisman, E., Crundall, D., Castro, C., 2017. Are situation awareness and decision-making in driving totally conscious processes? Results of a hazard prediction task. *Transp. Res. Part F: Traffic Psychol. Behav.* 44, 168–179. <https://doi.org/10.1016/j.trf.2016.11.005>.
- Horswill, M., 2016a. Improving fitness to drive: the case for hazard perception training. *Aust. Psychol.* 51 (3), 173–181. <https://doi.org/10.1111/ap.12132>.
- Horswill, M., 2016b. Hazard perception in driving. *Curr. Dir. Psychol. Sci.* 25 (6), 425–430. <https://doi.org/10.1177/0963721416663186>.
- *Horswill, M., Helman, S., 2003. A behavioral comparison between motorcyclists and a matched group of non-motorcycling car drivers: factors influencing accident risk. *Accid. Anal. Prev.* 35 (4), 589–597. [https://doi.org/10.1016/s0001-4575\(02\)00039-8](https://doi.org/10.1016/s0001-4575(02)00039-8).
- Horswill, M., Hill, A., Wetton, M., 2015. Can a video-based hazard perception test used for driver licensing predict crash involvement? *Accid. Anal. Prev.* 82, 213–219. <https://doi.org/10.1016/j.aap.2015.05.019>.
- *Horswill, M.S., Marrington, S.A., McCullough, C.M., Wood, J., Pachana, N.A., McWilliam, J., Raikos, M.K., 2008. The hazard perception ability of older drivers. *J. Gerontol.: Ser. B: Psychol. Sci. Soc. Sci.* 63 (4), 212–218. <https://doi.org/10.1093/geronb/63.4.P212>.
- Horswill, M.S., McKenna, F.P., 2004. Drivers' hazard perception ability: Situation awareness on the road. In: Banbury, S., Tremblay, S. (Eds.), *A Cognitive Approach to Situation Awareness: Theory and Application*. Ashgate Publishing Ltd, UK, pp. 155–175.
- *Horswill, M., Pachana, N., Wood, J., Marrington, S., McWilliam, J., McCullough, C., 2009. A comparison of the hazard perception ability of matched groups of healthy drivers aged 35 to 55, 65 to 74, and 75 to 84 years. *J. Int. Neuropsychol. Soc.* 15 (05), 799. <https://doi.org/10.1017/s1355617709990312>.
- *Horswill, M.S., Taylor, K., Newnam, S., Wetton, M., Hill, A., 2013. Even highly experienced drivers benefit from a brief hazard perception training intervention. *Accid. Anal. Prev.* 52, 100–110. <https://doi.org/10.1016/j.aap.2012.12.014>.
- Huang, J., Ford, J., 2012. Driving locus of control and driving behaviors: Inducing change through driver training. *Transp. Res. Part F: Traffic Psychol. Behav.* 15 (3), 358–368. <https://doi.org/10.1016/j.trf.2011.09.002>.
- *Huestegge, L., Skottke, E.M., Anders, S., Müssele, J., Debus, G., 2010. The development of hazard perception: Dissociation of visual orientation and hazard processing. *Transp. Res. Part F: Traffic Psychol. Behav.* 13 (1), 1–8. <https://doi.org/10.1016/j.trf.2009.09.005>.
- International Traffic Forum, 2018. Road Safety Annual Report 2018. Retrieved from https://www.itf-oecd.org/sites/default/files/docs/irtad-road-safety-annual-report-2018_2.pdf.
- *Isler, R., Starkey, N., Williamson, A., 2009. Video-based road commentary training improves hazard perception of young drivers in a dual task. *Accid. Anal. Prev.* 41 (3), 445–452. <https://doi.org/10.1016/j.aap.2008.12.016>.
- *Jackson, L., Chapman, P., Crundall, D., 2009. What happens next? Predicting other road users' behaviour as a function of driving experience and processing time. *Ergonomics* 52 (2), 154–164. <https://doi.org/10.1080/00140130802030714>.
- *Johnston, K.A., Scialfa, C.T., 2016. Hazard perception in emergency medical service responders. *Accid. Anal. Prev.* 95 (Part A), 91–96. <https://doi.org/10.1016/j.aap.2016.06.021>.
- *Lee, S., Klauer, S., Olsen, E., Simons-Morton, B., Dingus, T., Ramsey, D., Ouimet, M., 2008. Detection of road hazards by novice teen and experienced adult drivers. *Transp. Res. Rec.: J. Transp. Res. Board* 2078 (1), 26–32. <https://doi.org/10.3141/2078-04>.
- *Lim, P.C., Sheppard, E., Crundall, D., 2013. Cross-cultural effects on drivers' hazard perception. *Transp. Res. Part F: Traffic Psychol. Behav.* 21, 194–206. <https://doi.org/10.1016/j.trf.2013.09.016>.
- *Liu, C., Hosking, S., Lenné, M., 2009. Hazard perception abilities of experienced and novice motorcyclists: an interactive simulator experiment. *Transp. Res. Part F: Traffic Psychol. Behav.* 12 (4), 325–334. <https://doi.org/10.1016/j.trf.2009.04.003>.
- *Malone, S., Brünken, R., 2015. Hazard perception assessment – How much ecological validity is necessary? *Proc. Manuf.* 3, 2769–2776. <https://doi.org/10.1016/j.promfg.2015.07.713>.
- *Malone, S., Brünken, R., 2016. The role of ecological validity in hazard perception assessment. *Transp. Res. Part F: Traffic Psychol. Behav.* 40, 91–103. <https://doi.org/10.1016/j.trf.2016.04.008>.
- *Mackenzie, A.K., Harris, J.M., 2015. Eye movements and hazard perception in active and passive driving. *Vis. Cognit.* 23 (6), 736–757. <https://doi.org/10.1080/13506285.2015.1079583>.
- Mayhew, D.R., Simpson, H.M., Pak, A., 2003. Changes in collision rates among novice drivers during the first months of driving. *Accid. Anal. Prev.* 35 (5), 683–691. [https://doi.org/10.1016/S0001-4575\(02\)00047-7](https://doi.org/10.1016/S0001-4575(02)00047-7).
- McDonald, C., Goodwin, A., Pradhan, A., Romoser, M., Williams, A., 2015. A review of hazard anticipation training programs for young drivers. *J. Adolesc. Health* 57 (1), S15–S23. <https://doi.org/10.1016/j.jadohealth.2015.02.013>.
- *McKenna, F.P., Crick, J.L., 1994. Hazard Perception in Drivers: A Methodology for Testing and Training (Contractor Report 313). Retrieved from <https://trlco.uk/sites/default/files/CR313.pdf>.
- *Meir, A., Oron-Gilad, T., Parmet, Y., 2015. Are child-pedestrians able to identify hazardous traffic situations? Measuring their abilities in a virtual reality environment. *Saf. Sci.* 80, 33–40. <https://doi.org/10.1016/j.ssci.2015.07.007>.
- *Meyer, S., Sagberg, F., Torquato, R., 2014. Traffic hazard perception among children. *Transp. Res. Part F: Traffic Psychol. Behav.* 26 (Pt A), 190–198. <https://doi.org/10.1016/j.trf.2014.07.007>.
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D., 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Plos Med.* 6 (7), e1000097. <https://doi.org/10.1371/journal.pmed.1000097>.
- Özkan, T., Lajunen, T., 2005. Multidimensional Traffic Locus of Control Scale (T-LOC): factor structure and relationship to risky driving. *Pers. Individ. Differ.* 38 (3), 533–545. <https://doi.org/10.1016/j.paid.2004.05.007>.
- Parasuraman, R., Masalonis, A., Hancock, P., 2000. Fuzzy signal detection theory: Basic postulates and formulas for analyzing human and machine performance. *Hum.*

¹ References marked with an * were included in the systematic review.

- Factors. *J. Hum. Factors Ergonom. Soc.* 42 (4), 636–659. <https://doi.org/10.1518/001872000779697980>.
- *Pradhan, A.K., Hammel, K.R., DeRamus, R., Pollatsek, A., Noyce, D.A., Fisher, D.L., 2005. Using eye movements to evaluate effects of driver age on risk perception in a driving simulator. *Hum. Factors* 47 (4), 840–852.
- *Quimby, A.R., Watts, G.R., 1981. *Human Factors and Driving Performance (Laboratory Report 1004)*. Retrieved from <https://trl.co.uk/sites/default/files/LR1004.pdf>.
- *Rosenbloom, T., Perlman, A., Pereg, A., 2011. Hazard perception of motorcyclists and car drivers. *Accid. Anal. Prev.* 43 (3), 601–604. <https://doi.org/10.1016/j.aap.2010.08.005>.
- *Rosenbloom, T., Mandel, R., Rosner, Y., Eldror, E., 2015. Hazard perception test for pedestrians. *Accid. Anal. Prev.* 79, 160–169. <https://doi.org/10.1016/j.aap.2015.03.019>.
- *Sagberg, F., Bjørnskau, T., 2006. Hazard perception and driving experience among novice drivers. *Accid. Anal. Prev.* 38 (2), 407–414. <https://doi.org/10.1016/j.aap.2005.10.014>.
- *Scialfa, C.T., Deschenes, M.C., Ference, J., Boone, J., Horswill, M.S., Wetton, M., 2011. A hazard perception test for novice drivers. *Accid. Anal. Prev.* 43 (1), 204–208. <https://doi.org/10.1016/j.aap.2010.08.010>.
- *Scialfa, C.T., Borkenhagen, D., Lyon, J., Deschênes, M., Horswill, M., Wetton, M., 2012. The effects of driving experience on responses to a static hazard perception test. *Accid. Anal. Prev.* 45, 547–553. <https://doi.org/10.1016/j.aap.2011.09.005>.
- *Scialfa, C.T., Borkenhagen, D., Lyon, J., Deschênes, M., 2013. A comparison of static and dynamic hazard perception tests. *Accid. Anal. Prev.* 51, 268–273. <https://doi.org/10.1016/j.aap.2012.12.006>.
- Senserrick, T., Williams, A., 2015. Summary of Literature of the Effective Components of Graduated Driver Licensing Systems. Retrieved from https://www.parliament.vic.gov.au/images/Attachment_to_Submission_no_75_-_Literature_Review_of_Effective_Components_of_Graduated_Licensing_Schemes.pdf.
- *Shahar, A., Alberti, C.F., Clarke, D., Crundall, D., 2010. Hazard perception as a function of target location and the field of view. *Accid. Anal. Prev.* 42 (6), 1577–1584. <https://doi.org/10.1016/j.aap.2010.03.016>.
- *Smith, S., Horswill, M., Chambers, B., Wetton, M., 2009. Hazard perception in novice and experienced drivers: the effects of sleepiness. *Accid. Anal. Prev.* 41 (4), 729–733. <https://doi.org/10.1016/j.aap.2009.03.016>.
- Thomas, F.D., Rilea, S.L., Blomberg, R.D., Peck, R.C., Korbela, K.T., 2016. Evaluation of the Safety Benefits of the Risk Awareness and Perception Training Program for Novice Teen Drivers (DOT HS 812 235). Retrieved from <http://www.anstse.info/Resources%20PDF%27s/NEW%20Resources/NHTSA%20Risk%20Awareness%20PerceptionTraining%20NoviceTeen%20Drivers.pdf>.
- *Underwood, G., Phelps, N., Wright, C., van Loon, E., Galpin, A., 2005. Eye fixation scanpaths of younger and older drivers in a hazard perception task. *Ophthalmic Physiol. Opt.* 25 (4), 346–356. <https://doi.org/10.1111/j.1475-1313.2005.00290.x>.
- Underwood, J., Dillon, G., Farnsworth, B., Twiner, A., 2007. Reading the road: the influence of age and sex on child pedestrians' perceptions of road risk. *Br. J. Psychol.* 98 (1), 93–110. <https://doi.org/10.1348/000712606x104409>.
- United Kingdom Government, 2018. *Riding a Motorcycle, Moped or Motor Tricycle*. Retrieved from: <https://www.gov.uk/ride-motorcycle-moped>.
- *Vansteenkiste, P., Zeuwts, L., Cardon, G., Lenoir, M., 2016. A hazard-perception test for cycling children: an exploratory study. *Transp. Res. Part F: Traffic Psychol. Behav.* 41 (Part B), 182–194. <https://doi.org/10.1016/j.trf.2016.05.001>.
- *Ventsislavova, P., Gugliotta, A., Peña-Suarez, E., Garcia-Fernandez, P., Eisman, E., Crundall, D., Castro, C., 2016. What happens when drivers face hazards on the road? *Accid. Anal. Prev.* 91, 43–54. <https://doi.org/10.1016/j.aap.2016.02.013>.
- Ventsislavova, P., Crundall, D., 2018. The hazard prediction test: a comparison of free-response and multiple-choice formats. *Saf. Sci.* 109, 246–255. <https://doi.org/10.1016/j.ssci.2018.06.004>.
- Ventsislavova, P., Crundall, D., Baguley, T., Castro, C., Gugliotta, A., Garcia-Fernandez, P., et al., 2019. A comparison of hazard perception and hazard prediction tests across China, Spain and the UK. *Accid. Anal. Prev.* 122, 268–286. <https://doi.org/10.1016/j.aap.2018.10.010>.
- *Vlakveld, W.P., 2014. A comparative study of two desktop hazard perception tasks suitable for mass testing in which scores are not based on response latencies. *Transp. Res. Part F: Traffic Psychol. Behav.* 22, 218–231. <https://doi.org/10.1016/j.trf.2013.12.013>.
- *Wallis, T.S.A., Horswill, M.S., 2007. Using fuzzy signal detection theory to determine why experienced and trained drivers respond faster than novices in a hazard perception test. *Accid. Anal. Prev.* 39 (6), 1177–1185. <https://doi.org/10.1016/j.aap.2007.03.003>.
- *Wetton, M.A., Horswill, M.S., Hatherly, C., Wood, J.M., Pachana, N.A., Anstey, K.J., 2010. The development and validation of two complementary measures of drivers' hazard perception ability. *Accid. Anal. Prev.* 42 (4), 1232–1239. <https://doi.org/10.1016/j.aap.2010.01.017>.
- *Wetton, M.A., Hill, A., Horswill, M.S., 2011. The development and validation of a hazard perception test for use in driver licensing. *Accid. Anal. Prev.* 43 (5), 1759–1770. <https://doi.org/10.1016/j.aap.2011.04.007>.
- World Health Organisation, 2018. *Global Status Report on Road Safety 2018*. Retrieved from https://www.who.int/violence_injury_prevention/road_safety_status/2018/en/.
- *Yeung, J.S., Wong, Y.D., 2015. Effects of driver age and experience in abrupt-onset hazards. *Accid. Anal. Prev.* 78, 110–117. <https://doi.org/10.1016/j.aap.2015.02.024>.
- *Zeuwts, L., Vansteenkiste, P., Deconinck, F.J.A., Cardon, G., Lenoir, M., 2017. Hazard perception in young cyclists and adult cyclists. *Accid. Anal. Prev.* 105, 64–71. <https://doi.org/10.1016/j.aap.2016.04.034>.