Austroads

Technical Report
AP-T388-26

Literature Review on Hazard Perception Tests

Literature Review on Hazard Perception Tests

Prepared by

Simon Raftery, Lisa Wundersitz

Project Manager

Matin Nabavi Niaki

Publisher

Austroads Ltd.
Level 9, 570 George Street
Sydney NSW 2000 Australia
Phone: +61 2 8265 3300
austroads@austroads.gov.au
www.austroads.gov.au

Abstract

The purpose of this report is to identify the characteristics of effective Hazard Perception Tests (HPTs) for novice drivers. Evidence suggests that strong performance on HPTs positively correlates with safe driving, indicating HPTs have an important role to play in ensuring drivers have the skills necessary to safely respond to hazards on the road.

This literature review provides an analysis of current literature to identify the characteristics of HPTs, offering an overview of the strengths and limitations of different testing paradigms including the different approaches to hazard perception testing and how hazard perception is measured. The review also considers evidence regarding links between hazard perception testing and safety outcomes (e.g. crashes), procedural aspects of hazard perception (e.g. instructions, number of times the test can be taken and the inter-test interval), interventions for improving hazard perception, and the potential use of HPTs to identify individuals who may have an increased risk of crashing.

Representing a significant contribution to research in Hazard Perception Tests for novice drivers, this report provides key insights into the characteristics of HPTs and suggests further research into this technology will be an important component in the move to reduce road trauma across Australia and New Zealand.

Keywords

Hazard Perception Test, hazard prediction, hazard anticipation, hazard identification, road safety, HPT characteristics.

ISBN 978-1-923617-14-8

Austroads Project No. SAG6410

Austroads Publication No. AP-T388-26

Publication date February 2026

Pages 29

About Austroads

Austroads is the association of Australasian transport agencies.

Austroads' purpose is to support our member organisations to deliver an improved Australasian road transport network. To succeed in this task, we undertake leading-edge road and transport research which underpins our input to policy development and published guidance on the design, construction and management of the road network and its associated infrastructure.

Austroads provides a collective approach that delivers value for money, encourages shared knowledge and drives consistency for road users.

Austroads is governed by a Board consisting of senior executive representatives from each of its 11 member organisations:

- Transport for NSW
- Department of Transport and Planning (Transport Victoria)
- Queensland Department of Transport and Main Roads
- Main Roads Western Australia
- Department for Infrastructure and Transport South Australia
- Department of State Growth Tasmania
- Department of Logistics and Infrastructure Northern Territory
- City and Environment Directorate, Australian Capital Territory
- Department of Infrastructure, Transport, Regional Development, Communications, Sport and the Arts
- Australian Local Government Association
- NZ Transport Agency Waka Kotahi.

Acknowledgements

The authors would like to thank the project control group members: Linda Ferguson (TfNSW), Vanessa Cattermole (QTMR), Rugiyya Fradley (DIT), Portia Niarchos (DIT), Ragu Ganeshasundaram (MRWA), Louise Purcell (DTP) and Emily Morrison (DSG).

Summary

The purpose of this report is to identify the characteristics of effective Hazard Perception Tests (HPTs) for novice drivers through a review of literature.

The introduction of HPTs into licensing systems in recent years reflects a recognition that higher order skills such as visual search and hazard perception are more critical to safe driving than the vehicle handling skills that were traditionally emphasised in driver education and training (Isler and Starkey 2012). The rationale behind the tests is that novices who do not respond appropriately to video hazards may not respond appropriately to real on-road hazards and therefore may have a higher risk of crashing. Furthermore, it was anticipated that including hazard perception testing in the licensing process would encourage learners and instructors to focus more on scanning for and detecting hazards during training.

Hazard perception can be defined simply as the ‘ability to read the road and anticipate forthcoming events’ (McKenna et al. 2006, p2). This encompasses the detection of hazards, the appraisal of the threat posed by the hazard, and the selection of an appropriate response (Grayson et al. 2003). The ability to perceive threats and respond appropriately is perhaps one of the most important skills a driver can develop and deficiencies in this ability have been identified as a risk factor for crashes (Boufous et al. 2011). Differences in hazard perception have been observed across different driver demographics with inexperienced drivers and crash-involved drivers demonstrating poorer hazard perception than experienced and crash-free drivers (Horswill et al. 2020; Scialfa et al. 2011).

A comprehensive literature review was undertaken to identify the characteristics of HPTs as they relate to measurement and response styles, test instructions, format (e.g. video, CGI, still image), and test item selection. The review also examined the links between HPT performance and safety, and procedural aspects of testing, such as implications of failure, repeat testing intervals, and the utility of HPTs for identifying drivers requiring intervention.

Literature regarding the development or evaluation of HPTs, or which was otherwise relevant to other aspects of the review were obtained from the Scopus, Web of Science, PsycINFO, and Transportation Research Information Database (TRID) academic databases, with preference given to literature published in 2013 or later. Older research was included where appropriate. Only research published in English with full-text available was included.

Several methods have been used to measure hazard perception, with the primary methods being the presentation of hazardous scenarios via a video clip and recording either reaction time or ability to predict what happens next. While there are some exceptions, the preponderance of evidence suggests both approaches are valid methods for assessing an individual’s ability to detect hazards in the road environment. A common criticism of the standard HPT based on reaction time measures is that it is insufficient to determine the stimulus (hazard) to which the person has responded, while other factors such as sensitivity to risk can influence results. The hazard prediction test overcomes these issues by requiring the subject to predict how the scenario unfolds and may yield a more robust measure of hazard perception skill.

Modern HPTs use video clips to assess hazard perception skill, with both real-world footage and CGI clips proving suitable to assess hazard perception, though some argue that test items should be based on real examples of hazards. While real-world videos may be more valid representations of hazards, they can be limited in the types of hazards they can show (e.g. only daylight driving, high quality clips, and only those hazards encountered while recording footage). The hazards which are not captured cannot be evaluated in the test. While CGI clips may be staged, the major benefit is that the types of hazards that can be portrayed is virtually limitless and the ability to produce a wide variety of examples offers greater opportunity to evaluate a range of hazard scenarios.

Performance on HPTs has been associated with crash involvement, where drivers who pass the HPT on the first attempt have a lower risk of crashing than drivers who require at least 3 attempts to pass, particularly males and drivers from remote areas. The limitations of this literature review relate to HPT research as an emerging field of inquiry. While, as noted, some research exists to suggest HPT is a useful tool in addressing road safety issues, there is limited evidence in the field. Thus, there is significant scope for further investigation in this area, especially to better understand how HPT may be a useful tool in assessing fitness to drive and age-related deficits in driving ability, as well as the complex factors impacting drivers who repeatedly offend or who are living with medical conditions affecting capacity.

Hazard perception testing contributes to improved road safety outcomes by ensuring a minimum level of performance deemed necessary to drive safely. Understanding the characteristics of effective Hazard Perception Tests (HPTs) for novice drivers is critical to ensure drivers are well-equipped to prepare for and respond to driving hazards, particularly with respect to higher order skills such as visual search and hazard perception. This report offers a comprehensive analysis of current literature relating to HPTs and represents an important step along the path to reducing road trauma across transport networks in Australia and New Zealand.

Contents

1. Introduction

1.1 Purpose

The purpose of this report is to provide a comprehensive review of Australian and international literature identifying the characteristics of effective Hazard Perception Tests (HPTs).

This report is useful for practitioners considering the possibilities of HPTs to improve road safety outcomes.

Representing a significant contribution to research in Hazard Perception Tests for novice drivers, this report offers important insights into the characteristics of HPTs and suggests further research into this technology will be an important component in reducing road trauma across Australia and New Zealand.

1.2 Scope

This report provides a comprehensive literature review addressing the use and effectiveness of HPTs for novice drivers, providing a comprehensive summary of the characteristics of HPTs, including:

- test instructions
- test type (hazard perception vs. hazard prediction)
- hazard perception measurement and response style/format (e.g. reaction time)
- test format/style (e.g. video, computer generated imagery, virtual reality, and augmented reality)
- hazard scenario selection
- timing of HPT in the licensing scheme
- repeat testing after failure and measures for managing people who fail the HPT
- HPT validity
- the safety benefits of hazard perception testing.

The review also discusses evidence of the relationship between the development and implementation of HPTs and road safety outcomes, including crashes and association with other safety-relevant issues such as unsafe driving behaviours (e.g. driving styles and traffic offences), with consideration given to the utility of HPT scores to identify unsafe drivers.

Additionally, the review considers the implications of failing an HPT, including re-testing, procedural aspects of hazard perception (e.g. instructions, number of times the test can be taken and the inter-test interval), interventions for improving hazard perception, and the potential use of HPTs to identify individuals who may have an increased risk of crashing.

1.3 Background

The traffic environment is complex and dynamic, comprising interactions between a diverse range of road users and vehicles of different sizes, often travelling at different speeds. To safely navigate this environment, road users must constantly monitor and evaluate their own behaviour, as well as the movements of other road users. While improvements in road infrastructure and vehicle safety have produced substantial safety benefits, human behaviour and decision making are a critical factor in most crashes (Dingus et al. 2016). Literature indicates that hazard perception – the ability to anticipate, identify, and react to dangerous situations in the driving environment – can predict crash involvement (Boufous et al. 2011; Horswill et al. 2015; Horswill et al. 2020; Wells et al. 2008).

The introduction of Hazard Perception Tests (HPTs) into licensing systems in recent years reflects a recognition that higher order skills such as visual search and hazard perception are more critical to safe driving than the vehicle handling skills that were traditionally emphasised in driver education and training (Isler and Starkey 2012). The rationale behind the use of hazard perception tests is that novices who do not respond appropriately to video hazards may not respond appropriately to actual road hazards and therefore may have a higher risk of crashing. Furthermore, including HPTs in the licensing system encourages drivers and instructors to prioritise hazard perception as a core component of safe driving practice, and to explore driving skills and strategies during the learning process to help learner drivers respond as safely as possible to unpredictable scenarios.

Hazard perception testing contributes to safety by ensuring a minimum hazard perception skill level deemed necessary to drive safely. Methods for improving hazard perception performance include the gaining of more driving experience, though this will be dependent on the quality of the experience gained. Hazard perception training has been shown to improve hazard perception skill and performance on HPTs and may lead to improvement in driving behaviour, though the effects on crashes are limited. There is a risk that hazard perception training may increase the likelihood of people passing the hazard perception test without a concomitant reduction in crash risk, though this requires further investigation.

2. Methodology

This methodology describes the steps to identify and compile literature on the development, evaluation, and implementation of HPTs. Literature published during and after 2013 was prioritised, though relevant older research was included where appropriate, particularly relating to literature that addressed issues not covered in more recent studies. This review included only research published in English, and with full-text available.

The methodology included a comprehensive search of the following academic databases for peer-reviewed journal articles, conference papers, books/book chapters, and technical or research reports:

- Scopus
- Web of Science
- PsycINFO
- Transportation Research Information Database (TRID).

Databases covering relevant areas to the development or use of hazard perception tests are prioritise by the methodology. The first 3 databases capture peer-reviewed academic literature (e.g. journal articles) related to engineering, computer science, education and medicine. TRID captures a wider range of transport related research, including unpublished or 'grey' literature such as government and consultant technical reports, which are not always indexed by other databases. It is important to capture 'grey' literature as it is a common method of publishing driver training and hazard perception test literature. These databases are among those recommended as preferred databases for conducting literature searches in a systematic review of such research databases (Gusenbauer and Haddaway 2020).

Relevant literature is identified by a search query using relevant key words and Boolean operators. As an example, the initial search may use the following queries:

> *Traffic AND ('Hazard perception test' OR 'Hazard detection test' OR 'hazard prediction' OR 'hazard identification') AND 'road safety' AND ('benefit' OR 'evaluation' OR 'evidence' OR 'design').*
>
> *('Hazard perception test' OR 'Hazard detection test' OR 'hazard prediction' OR 'hazard identification') AND 'driver' AND ('benefit' OR 'evaluation' OR 'evidence' OR 'design').*

Additional terms or queries were used to focus on specific aspects of HPT and to address all areas within the scope of the literature review, including:

- test instructions
- hazard perception measurement and response style/format (e.g. reaction time)
- test format/style (e.g. video, computer generated imagery, virtual reality, and augmented reality)
- hazard scenario selection
- repeat testing after failure.

Titles and abstracts were scanned to identify relevant papers for the review. Additionally, an ancestry approach was used to identify other relevant research from the reference lists of included studies.

3. Characteristics of HPTs

This section provides a comprehensive literature review on use and effectiveness of HPTs for novice drivers. It covers characteristics of HPTs and discusses evidence of the relationship between the development and implementation of HPTs and road safety outcomes. Additionally, the review considers the implications of failing an HPT as well as interventions to improve Hazard Perception skills.

This section discusses the various aspects and characteristics of hazard perception testing. Each characteristic is addressed explaining the methods and approaches to operationalisation in research and practice.

As a starting point it is helpful to have some guidance on the effective design for an HPT. Wetton et al. (2011) identified 5 principles for the development of a hazard perception test. These principles offer a useful guide when considering the characteristics of the different hazard perception tests that have been used in both research and the licensing process. These principles are:

- HPTs should only measure hazard perception skills.
- HPTs should be able to discriminate between individuals on the basis of differences in hazard perception skill, not response time. There should be sufficient anticipatory cues such that those with good hazard perception are able to determine if a hazard is unfolding.
- Hazard scenarios should present genuine driving situations. Staged situations, either in the real world or CGI, may not represent true or real-life hazards as they may lack important precursors or cues to enable someone to determine if a hazard is unfolding, which may negatively impact the measure of hazard perception. Also, tests based on staged scenarios may be biased more towards drivers who are similar to those creating the test (e.g. same level of driving experience).
- Instructions should be unambiguous and define what a hazard is and the situations for which a response is required.
- HPTs should be able to identify and classify inappropriate responses and be able to identify people who are trying to cheat.

3.1 Hazard perception

Hazard perception refers to a motorist’s (driver or rider) ability to anticipate, identify, and react to dangerous situations in the driving environment. This encompasses the detection of hazards, the appraisal of the threat posed by the hazard, and the selection of an appropriate response (Grayson et al. 2003). The ability to perceive threats and respond appropriately is perhaps one of the most important skills a motorist can develop and deficiencies in this ability have been identified as a risk factor for crashes (Boufous et al. 2011). Differences in hazard perception have been observed across different driver demographics, with inexperienced drivers and crash-involved drivers demonstrating poorer hazard perception than experienced drivers and crash-free drivers (Horswill et al. 2020; Scialfa et al. 2011).

The successful detection and avoidance of hazards is dependent upon a range of sub-skills. Based on analysis of different models of hazard perception Dreßler et al. (2017) identified the following components of hazard perception skill:

1. Observation – scanning the road and surrounds for hazards.
2. Localisation – identifying relevant objects in the environment, including attending to places where hazards may appear.
3. Identification – the visual processing of objects and drawing on mental models to understand what they are and how they may behave.
4. Assessment of hazard – relating to urgency (how likely) and intensity (how serious).
5. Self-assessment of abilities – the greater an individual perceives their ability to cope with a hazard the less dangerous a situation is perceived.
6. Assessment of risk – appraisal of risk in relation to the individual's own level of risk acceptance.
7. Decision making – describes cognitive processes encompassing many of the previous components (assessment of hazard, self-assessment of abilities, and assessment of risk) and the driver's repertoire of behaviour and anticipated consequences of behaviour. The activation of automatic and routine behaviours is essential for coping with hazards.
8. Action – encompasses the execution of the behaviour and the competence of vehicle operation and manoeuvring.

With the exception of the *action* component, each of these is relevant to hazard perception and will likely shape the outcomes of HPTs regardless of the manner in which they are measured. For example, the first 3 skills relate to hazard detection, looking in the right direction and identifying to objects to assess if they are hazards. The next 4 skills relate to hazard appraisal and are influenced by factors such as the individual's sensitivity to risk and the perceived ability to manage the situation. Together, these skills may impact HPT outcomes based on the speed at which hazards are identified (skills 1-3), while during the appraisal and decision-making stages (skills 4-7) those with a greater sensitivity to risk may respond earlier than those who wait longer for the hazard to unfold before they respond. In this study, Dreßler et al. (2017) suggests that those with more acute risk perception (skills 1-3) are able to make decisions about how to respond to hazards more quickly (skills 4-7) and decrease crash risk, whereas those with a delayed response to an unfolding hazard are at greater risk of an adverse event.

3.2 Measurement of Hazard Perception

Hazard perception is measured using tests designed to determine a person's ability to identify hazards in a traffic scene and can be measured as the amount of time it takes for the person to identify the hazard (reaction time), the accuracy of hazard detection (based on identifying the hazard or not), or the ability to accurately predict how the traffic scene will unfold. The method used to measure hazard perception is dependent on the type of test employed.

Test types

The 2 main types of tests are the classic HPT which is primarily based on reaction time and accuracy, and the hazard prediction test, which draws on the individual's ability to predict how the scenario will unfold. Tests usually consist of several items based on scenarios and hazard types, with a final hazard perception or prediction score determined based on the individual's performance for the whole test. This section provides a discussion of the different testing methods, response formats, and the process of how items are selected or created for inclusion in the test.

3.2.1 Hazard Perception Test

Hazard Perception Tests present individuals with a traffic scene and require the respondent to indicate when they have detected a hazard. Methods of indicating detection in HPTs usually involve clicking a mouse, pressing a button, or tapping a touch screen at a point when either a crash is likely or when the camera car (i.e. the car which the candidate is ’driving‘) may be required to perform some action (slow down, brake, turn) to avoid a collision with another road user (Crundall et al. 2016; Crundall et al. 2021; Malone and Brünken 2016; Wetton et al. 2010; Wetton et al. 2011). Reaction time is considered a reasonable measure of hazard perception as those with good hazard perception should identify potential hazards sooner and respond earlier than those with poorer hazard perception skills. Reaction time is usually measured from the point at which the hazard is first identifiable, and the average time taken to respond to all hazards across all test items is used as the measure of hazard perception skill (see, inter alia, Molone and Brünken 2015; Manley et al. 2020; Moran et al. 2019; Wetton et al. 2010).

Another approach to response time is the use of a ‘response window’ which provides a defined period during which input from the test taker is recorded as a response to the hazard (Crundall 2016). As with standard reaction time, the response window is specified based on advice from experts (but may be refined based on the results of pilot testing) and commences with the first onset of the hazard and concludes at a point when either a collision would have occurred or a response is considered too late to have avoided a collision (Crundall et al. 2021). In some cases, the response window is divided into segments, with responses awarded a score based on the segment in which it was recorded (Crundall 2016). For example, Crundall (2016) divided the response window into 5 segments with a score ranging from one to 5 such that 5 points is awarded for earlier responses and lower scores awarded in descending order. A score of zero is given where a response is not made during the response window. These points can then be used to provide a hazard perception score based on performance on all items within a hazard perception test. This approach has been adopted in the official HPT in the United Kingdom (Crundall 2016).

Reaction time appears to be a valid measure of hazard perception because during real-world driving, the speed at which someone identifies and reacts to a hazard may be the difference between either crashing or avoiding a crash (Malone and Brünken 2016). However, within the realm of hazard perception testing, using reaction time as a measure of hazard perception is not without issue. First, perhaps one of the largest limitations of the reaction time measure is that it is not possible to determine the stimulus to which the person is responding and, while the reaction may be fast, the accuracy of the response may be low. Also, responses could be made for an entirely different element than that expected by the test (Crundall 2016; Wetton et al. 2010). Second, differential thresholds for risk may influence reaction time such that those with a low threshold for risk may respond to stimuli with a small potential for a crash, while those with a higher risk threshold may delay responding until they perceive the risk to be more certain (Dreßler et al. 2017; Egea-Caparros et al. 2016; Ventsislavova et al. 2022). Several studies have demonstrated that those who rate a hazard perception item as having lower risk also have slower response times and experienced drivers tend to rate clips as more hazardous than novice drivers (Castro et al. 2021; Ehsani et al. 2020; Ventsislavova et al. 2019).

A limitation of response measures based on scoring windows is that the responses that occur outside of the window are not counted, which may lead to discounting early responding and potentially punish safe drivers who detect and respond to a hazard sooner. When responses are recorded in milliseconds, having made a response that is not recorded the respondent may attempt another response (if they are aware that their first response was not recorded which is unlikely) which could result in a lower hazard perception score. There is also the possibility that response windows are too small and do not afford drivers sufficient time to make a response. The careful determination of the response window should reduce the likelihood of such an issue. Furthermore, investigations of the potential for this issue to impact hazard perception scores have found that it has not affected the results of the tests being used (Crundall 2016; Crundall et al. 2021).

Finally, another limitation of reaction time measures is the potential for respondents to make multiple responses to game the system (i.e. cheat), while in other cases repeated responses may reflect a genuine response to different, perhaps incorrect (i.e. hazard may not eventuate), hazardous stimuli. In the case of cheating, hazard perception testing should have the ability to identify those making repeated responses. Such indicators may be an excessive number of responses to an item as compared to the average responses for an item or a genuine baseline established during pilot testing. Handling cases where genuine attempts result in multiple responses present a different problem. In other words, should the first response be accepted, which may be outside of the response window or the first correct response? Crundall et al. (2021) opted to use the first response within the response window (which would also result in the highest hazard perception score). A more simplified approach to resolving both problems would be to limit the number of recorded responses per item to one.

In order to overcome the limitations of standard reaction time tests regarding hazard perception accuracy, some HPTs incorporate a spatial element requiring respondents to indicate the hazard to which they are responding. There are several methods for achieving this. One approach is to use a standard reaction time response with the clip stopping at the time the input is made and the respondent then indicating the location of the hazard that they are responding to using a touch screen or mouse pointer (Vlakveld 2014). Another approach is to have respondents make one input, such as touching or clicking on the hazard, and recording the response time and accuracy simultaneously (Hill et al. 2019; Scialfa et al. 2011; Wetton et al. 2011). A less common approach to measure both reaction time and accuracy involves the use of eye tracking devices. Eye tracking monitors visual scanning behaviour and also determines how fast the respondent identifies the hazard (time to first fixation), as well as how long they fixate on that hazard. Eye tracking has been employed in several studies of hazard perception and has demonstrated the ability to differentiate groups based on reaction measures, for example time to first fixation (Moran et al. 2019; Sun et al. 2023; Yang et al. 2021). While the inclusion of spatial/accuracy measures have been found to increase the accuracy of hazard perception measures, it has been suggested that the inability of these measures to distinguish between groups (i.e. both novice drivers and experienced drivers can accurately identify hazards, the primary difference is in how fast they do so) adds little additional benefit over simple reaction time measures (Moran et al. 2019).

3.2.2 Hazard prediction test

A second method is the hazard prediction test which has been developed based on theories of situational awareness (Jackson et al. 2009). Hazard prediction tests are thought to overcome the limitations of reaction-based hazard perception tests by removing the need to respond to the hazard. Hazard prediction tests, therefore, are not influenced by individual differences in risk sensitivity and they also measure the accuracy of the hazard response (Crundall 2016).

Hazard prediction tests adopt a largely similar methodology to HPTs, presenting the individual with video clips of traffic scenarios. Rather than rely on reaction times to measure perception, hazard prediction tests stop and occlude the clip at around the time a hazard can first be identified. The person undertaking the test then responds to questions designed to gauge their level of perception, asking them to identify the hazard and its location, and predict what happens next (i.e. explain how the scene unfolds). Using this method to gauge hazard perception is less susceptible to differences in risk sensitivity as, rather than reacting to the presence of a hazard (which may be delayed due to a high threshold for risk), respondents are required to describe how they believe the scene will unfold, including the expected movements of other road users, and thus identify potential hazards in a more descriptive manner rather than a potentially ambiguous reaction time response. Answer formats have included written or typed responses (Gugliotta et al. 2017; Ventisislavova et al. 2018) and verbal responses (Crundall 2016; Horswill et al. 2020), though more recent hazard perception tests have used multiple choice response options (Castro et al. 2021; Goodge et al. 2024; Ventsislavova and Crundall 2018). The tests have variously been scored giving 2 points for a completely correct answer, one point for a partially correct answer or answer with less detail, and zero points for a wrong answer (Castro et al. 2014; Gugliotta et al. 2017; Wu et al. 2021). Others have scored one point for a correct answer and zero points for a wrong answer (Muela et al. 2021; Ventislavova et al. 2018; Ventsislavova et al. 2022). A hazard prediction score is then calculated based on performance across all items. Ehsani et al. (2020) used a different method and asked how likely the driver was to be in a crash with answers scored on a Likert scale ranging from 1 (not likely) to 5 (very likely). Alternatively, hazard prediction scores can be simply based on the number or proportion of hazards correctly identified (Crundall et al. 2021). Studies using the hazard prediction paradigm have demonstrated the ability to discriminate between novice and experienced drivers, with experienced drivers found to have better hazard prediction and accuracy compared to novices (Castro et al. 2021; Crundall et al. 2021; Ehsani et al. 2020; Gugliotta et al. 2017; Horswill et al. 2020; Muela et al. 2021; Ventsislavova et al. 2022; Wu et al. 2021).

A key element of the hazard prediction test is stopping the clip at a point where there is sufficient evidence for those with good hazard perception skills, who have been attending to the correct areas of the scene, to be able to predict the onset of the hazard. At this point the clip is either paused with an image remaining visible to respondents, or occluded (i.e. replaced with a black screen) such that no image is available. Of these methods, occlusion has been found to have the ability to discriminate between novices and experts (Castro et al. 2014; Crundall 2016; Jackson et al. 2009; Wu et al. 2021), while paused images have not (Jackson et al. 2009; Vogel et al. 2003). The occlusion method is likely more effective, as drivers are required to attend the scene and use cues to identify hazards to develop an accurate prediction of how the scene will unfold before the scene is occluded. Experienced drivers with better hazard prediction can be expected to outperform novices as they are more likely to identify relevant cues and have a better understanding of how the traffic scene will unfold. A paused image on the screen is likely less effective because the act of pausing alerts the respondent that something important is about to unfold and, with the image still available to view, they have more time to examine the scene and identify relevant cues. Thus, all drivers, regardless of experience, should be able to predict hazards when a paused image remains visible. Additionally, the point at which the clip is occluded can influence test results. If the occlusion occurs too soon and before relevant cues are present, there is insufficient information for drivers to accurately predict hazards, while stopping the clip at a point after the hazard has become apparent is insufficient to differentiate between drivers of varying experience (Ventsislavova and Crundall 2018).

The second aspect of the hazard prediction test is the subject's response used to measure hazard prediction. Various versions of hazard prediction tests have included several questions to measure the subject's ability to identify the hazard (e.g. What is the hazard? Where is it located?) and all ask the question: What happens next? (Castro et al. 2014; Castro et al. 2019; Ventisilavova et al. 2016). However, it appears that simply asking subjects to identify what happens next is sufficient to measure hazard prediction (Castro et al. 2021; Goodge et al. 2024; Muela et al. 2021; Ventsislavova et al. 2018). The nature of these questions is such that answers are often recorded as a free response, with subjects either writing down or verbalising their answers. This answer modality requires each answer to be coded for accuracy and correctness, a time-consuming approach, which limits the utility of the hazard prediction test using these response modalities for official hazard perception testing purposes. More recently, studies adopting the hazard prediction test have used multiple choice answers to measure hazard prediction, which would simplify the computation of hazard prediction scores and make the testing more suitable for mass testing for official purposes (Ventsislavova and Crundall 2018). A potential drawback for adopting multiple choice responses is that they introduce the potential for subjects to guess the correct response. While this has not been examined, it does not appear that this has negatively impacted hazard prediction tests, as those adopting multiple choice answers have retained the ability to distinguish between novice and experienced drivers (Ventislavova et al. 2018).

3.2.3 Hazard Perception Test versus hazard prediction test, which is best?

The evidence suggests that each type of test provide a valid assessment of hazard perception skill though there are benefits and limitations to both approaches.

Ventsislavova et al. (2019) examined the efficacy of a standard HPT compared to a hazard prediction test, both using the same items, with drivers in Spain, China, and the UK. The tests utilised 10 hazard clips from each country (i.e. 30 clips total) recorded during real driving. Each test was used with different samples of novice and experienced drivers from each country. The Hazard Perception Test was based on reaction time with participants' scores based on response made during a response window, while the hazard prediction test occluded each clip and participants were required to type their answer to the question 'what happens next?', with one point awarded for a correct response. The results of the Hazard Perception Test found no effect of experience on accuracy (number of hazards identified, measured by correct response) or response times, whereas the hazard prediction test found experienced drivers were more likely than inexperienced drivers to predict hazards. Comparisons between the 2 test types demonstrated that participants scored higher on the Hazard Perception Test than the hazard prediction test, suggesting that driving experience may only be beneficial for the hazard prediction test.

The findings suggest that the hazard prediction paradigm may be a more robust measure of hazard perception skill. However, the findings that the standard Hazard Perception Test was unable to distinguish between groups based on driving experience runs counter to a large body of research demonstrating this capability for tests based on reaction time (e.g. Horswill et al. 2020; Malone et al. 2016; Moran et al. 2019; Rosenbloom et al. 2011; Scialfa et al. 2012). The inability of the HPT used by Ventsislavova et al. (2019) to discriminate between driver groups may be more suggestive that the test itself was not a valid measure; the validity of the test was not assessed prior to the study.

A more recent study has also addressed this issue. Using a 10-minute CGI video clip containing 10 hazards Crundall et al. (2021) assessed both hazard perception and hazard prediction for experienced and novice drivers from the UK, with half the participants assigned to the hazard perception (reaction time) measure and the other half to the hazard prediction measure. The latter was measured by occluding the clips at the onset of hazards and asking participants to provide a multiple-choice response to the question 'what happens next?'. For both test types, experienced drivers were found to perform better than novice drivers, with no indication that one test was better at discriminating between groups than the other, although the effect size was greater for the response time measure used in the Hazard Perception Test. Accuracy for both experienced and novice drivers was higher in the Hazard Perception Test (i.e. in terms of the number of hazards to which correct responses were made) than the hazard prediction test. Additionally, based on the performance of individual items, Crundall et al. (2021) suggest that some hazards are better measured by response-time measures, while others, perhaps those that are less obvious, are better measured by hazard prediction.

These 2 studies directly comparing hazard perception and hazard prediction measures using the same test items have produced conflicting results, with one indicating hazard prediction is the stronger test (Ventsislavova et al. 2019) and the other suggests Hazard Perception Tests are better (Crundall et al. 2021). While both are arguably valid measures of the ability to detect hazards, the hazard prediction paradigm appears to overcome some of the limitations of reaction-time measures (e.g. confounding effects of risk sensitivity, limited ability to determine the stimulus to which the response is made, potential for multiple responses and attempts to cheat), although is not itself without limitations (e.g. selection of occlusion point can influence test outcomes). Of the 2 methods, it appears that participants found the standard Hazard Perception Test measuring reaction time the easiest, while the hazard prediction test proved challenging, even for experienced drivers. However, both have been proven effective measures of hazard perception skill, the use of either mode for official testing may be a matter of preference or ease with which they can be scored or graded. Hazard perception testing using reaction time measures are more suitable to this approach than free-text responses, although the use of multiple-choice responses simplifies scoring of hazard prediction testing.

Kroll et al. (2020) developed a Hazard Perception Test for emergency service drivers in the UK incorporating a variety of response types, including both standard reaction time measures and hazard prediction measures in the same test for different items. The test was found to effectively discriminate between novice and experienced drivers. There may be benefits for adopting Hazard Perception Tests using different response types in order to realise the benefits and overcome the limitations of each approach. As noted above, Crundall et al. (2021) found that some items were better when measured by the hazard perception (reaction time) format, while others were better when measured using the hazard prediction format. Others have suggested that hazard perception in general may be enhanced through the inclusion of multiple test formats and response types (Moran et al. 2019).

3.3 Test format

The 2 main formats for Hazard Perception Tests involve the use of video or still images (i.e. pictures) of traffic scenes taken from the driver's perspective. With the advent of better technologies including high-definition video cameras and faster, more capable computers, videos (real-world and CGI) have become the predominant mode to measure hazard perception performance.

It has been suggested that HPTs should be based on real driving situations to avoid potential biases and ensure all hazards unfold with naturally occurring cues (Wetton et al. 2011). While this is a methodologically sound and valid approach, there are some limitations to the use of such footage. First, as there is a need for high quality footage, hazards occurring during low-light conditions due to poor weather or at night may be excluded (Ehsani et al. 2020). Second, many hours (i.e. hundreds) of footage are needed to produce a sufficient pool of clips for a test (Horswill et al. 2011). As hazards are naturally occurring, it may not be possible to obtain clips or examples of some types of important hazards as they may not eventuate during filming. As such, the test is made only from the pool of hazards available.

One of the main concerns over the use of CGI clips is that they are staged and important cues as to the onset of hazards may be missing (Crundall et al. 2021). However, the benefits of using CGI are that there is greater control over the type of hazards included (they can be specified), it is not necessary to record hundreds of hours of footage (although time and money savings from not filming real driving may be diminished by clip development times and costs), and there is virtually no limit to the number or variety of scenarios that can be developed, ensuring a diverse pool of clips for testing and repeat testing. Additionally, the quality of each clip can be controlled by the development team, and clips occurring under conditions that may be somewhat rare and difficult to capture by real world driving (e.g. night driving on highways with infrequent traffic) are possible. The visual quality of current CGI hazard perception clips is very high, and some are almost indistinguishable from real driving footage. For an example, see the UK hazard perception test used for licensing (https://www.driving-theory-test.com/hazard-perception?qt- hpt_mini_pager_test=1).

While there are benefits and limitations to using both CGI and real-world driving footage, HPTs based on both formats have demonstrated suitable performance for measuring hazard perception and distinguishing between groups based on driving experience (Moran et al. 2019).

The third format that has been used to measure hazard perception is still images (Lyon et al. 2011; Moran et al. 2019; Scialfa et al. 2012; Sun et al. 2024; Yang et al. 2021). The method involves presenting a series of images of traffic scenes and recording reaction time to an input indicating the presence of a hazard (Lyon et al. 2011) or to identify the hazard in the scene by touching or clicking on it (Scialfa et al. 2012; Yang et al. 2021). Sun et al. (2024) employed a different approach and presented static images for 3 seconds and required respondents to indicate whether the scene contained a hazard by pressing ‘F’ on the keyboard if no hazard was present or ‘J’ if one was present, measuring hazard perception by reaction time. All studies using still images demonstrated differences between experienced and less-experienced drivers, suggesting they are suitable to use for hazard perception (Lyon et al. 2011; Moran et al. 2019; Scialfa et al. 2012; Sun et al. 2024; Yang et al. 2021) although there is some evidence that they may be less effective for measuring hazard perception in some scenarios. Yang et al. (2021) found that reaction time for items involving vehicles were faster than those involving pedestrians or covert hazards. It is possible that static images do not provide sufficient cues to rapidly identify these hazards, particularly if they take longer to unfold as may be the case for covert hazards. It is possible that when using still images respondents focus more on objects on the road, which are perhaps more immediate hazards, than objects (e.g. pedestrians) that appear at the side of the road and are not a hazard until they move on to the roadway.

When driving, the road is a dynamic environment and always changing. As such, video-based hazard perception tests can be considered more ecologically valid than still images. Indeed, the video format is the most adopted mode for modern HPTs (Moran et al. 2019).

3.4 Selection of HPT items

Hazard perception clips for video-based HPTs are generally obtained from video recording of natural driving in an area largely representative of the typical driving environment for the jurisdiction in which the HPT will be used (see inter alia Ehsani et al. 2020; Horswill et al. 2011; Scialfa et al. 2011; Muela et al. 2021). Segments of the video with a potential hazard are then identified by experts such as researchers, traffic psychologists, and sometimes driving instructors (Castro et al. 2014; Castro et al. 2021; Horswill et al. 2011; Muela et al. 2021; Moran et al. 2019; Sagberg and Bjørnskau 2006; Ventsislavova et al. 2021). While there is no control over the types of hazards encountered during the filming of naturalistic driving, some tests endeavour to include hazards based on prior research and crash frequency (Dreßler et al. 2017; Horswill et al. 2011). Some tests include only actual driving hazards where the camera car was required to slow, brake, or perform an avoidance manoeuvre (e.g. Ehsani et al. 2020; Horswill et al. 2011; Sawada et al. 2021) while others also identify quasi-hazardous or non-hazardous situations where there is a potential conflict but the hazard did not unfold and the camera car was not required to alter speed or direction (e.g. Ventsislavova et al. 2016). The latter scenario is intended to provide test items that do not require a response from participants. Once clips are identified they are again reviewed by experts and a pool of items (clips) selected based on specified criteria and expert assessment and consensus. Criteria can include whether the hazard is present (i.e. the camera car was forced to take action to avoid a collision) and unambiguous, that only one hazard is present, and that clips are of good video quality with clear image, good lighting, suitable cues present to detect the hazard (e.g. Ehsani et al. 2020; Hill et al. 2019; Horswill et al. 2011; Moran et al. 2019). Experts often also identify the earliest point at which a hazard can be predicted based on the cues present in the clip, which is used as the start of the scoring window to determine response time or, for hazard prediction tests, the occlusion point and response options (e.g. Castro et al. 2021; Chou et al. 2014; Crundall et al. 2021; Hill et al. 2019; Ventsislavova et al. 2021). These clips are then often pilot tested to determine their validity to measure hazard perception, with the most common method comparing the results of experienced drivers versus novice drivers (Horswill et al. 2011; Moran et al. 2019; Scialfa et al. 2011; Ventsislavova et al. 2021).

Following pilot testing, final clip selection and HPT formation may be based on several factors. First, individual clips that show statistically significant differences in performance between novice and experienced drivers such that experienced drivers perform better (i.e. faster reaction time, better hazard prediction) than novices (e.g. Hill et al. 2019; Horswill et al. 2011). Second, based on the hits, misses, and false alarm rates for each item where hits are accurate detection of hazards, misses are failing to detect a hazard, and false alarms are responding to an incorrect hazard. For example, if there are items for which a large number of experienced drivers do not perform correctly the item may be removed. A typical cut point for items may be hazards that are successfully identified by 85% of the experienced driver cohort (Scialfa et al. 2011; Scialfa et al. 2012). Another method of selection may be based on a threshold for differences between response times, for example, selecting clips where the difference in mean response times between experienced or novice drivers is at least 200ms (Lyon et al. 2011; Scialfa et al. 2012). Where necessary, adjustment of the scoring window may also occur based on the performance of participants. Once a final set of clips has been identified it is common to evaluate the psychometric properties of the test using each item's correlation to the total HP score and other items in the test (Scialfa et al. 2014) and reliability assessed by Cronbach's alpha (see inter alia, Crundall et al. 2021; Wu et al. 2021).

Tests involving the use of CGI clips are typically based on common crash scenarios for novice drivers based on analysis of relevant crash data. Tests developed for licensing purposes may also specify that particular types of hazards are also included, such as those involving vulnerable road users (pedestrians, cyclists, or motorcyclists). Experts may then help develop scripts for the clip, including the inclusion of cues for hazards (Crundall et al. 2021). Tests and clips may also be validated using pilot tests, again comparing the results of novice and experienced drivers on performance (Dreßler et al. 2017). Clips may be further refined or amended based on findings of statistical analysis and in some cases based on feedback from participants or experts (Dreßler et al. 2017; Moran et al. 2019).

The length of video clips is another consideration with the HPTs identified in the literature having clip lengths ranging from 7 seconds to 60 seconds for tests involving the presentation of a series of clips and scenarios (e.g. Castro et al. 2014; Crundall 2016; Ehsani et al. 2020; Hill et al. 2019; Malone and Brünken 2016; Manley et al. 2020; Wu et al. 2021). One study employed a 10-minute CGI video clip containing 10 hazards arguing that a continuous video is more akin to the actual driving experience (Crundall et al. 2021). All have proven effective at measuring hazard perception based on the differences in performance for experienced versus novice drivers. The presentation of the hazard within the clip should vary across clips so that individual's do not always expect or predict the hazard to occur at the same time as this could negatively impact the measurement of hazard perception.

3.5 Test instructions

An important aspect of the HPT is the instructions provided to candidates, particularly regarding the definition of a hazard and the circumstances under which a response should be made. As individuals may differ in what they consider a hazard based on their sensitivity to risk, or their belief in their ability to identify and negotiate hazards as they arise, it is important to provide unambiguous definitions of what constitutes a hazard. Many hazard perception tests adopt the term 'traffic conflict' often defined as an event that may result in a collision between the driver's car and another road user if they do not slow down, stop, or alter course (e.g. Crundall 2016; Malone and Brünken 2015; Manley et al. 2020; Wetton et al. 2010). It follows then, that the hazards contained within the test items must conform to the definition provided. Any definition that refers to 'traffic conflicts' or 'other road users' rules out the inclusion of hazards that do not involve such conflicts, such as corners, changes in road surface conditions, and changes in forward visibility due to weather events (e.g. heavy rain, fog, sun glare), and potentially animals or wildlife.

The instructions should also include information about how to make an appropriate response (e.g. press button as soon as a hazard is detected) and discourage cheating by describing inappropriate responses (e.g. repetitive button pressing) and the consequences (e.g. item or test failure). It is also common practice for HPTs to provide examples or practice scenarios to enable candidates to familiarise themselves with the test and response interface (see inter alia, Crundall 2016; Castro et al. 2021; Botzer et al. 2019; Horswill et al. 2011; Sagberg and Bjørnskau 2006; Sawada et al. 2021; Scialfa et al. 2014; Sun et al. 2024; Ventisislavova et al. 2016, etc.). Tests used as part of the licensing process should be able to be understood by people of varying English language comprehension abilities (Wetton et al. 2011).

3.6 Timing of HPT

The introduction of HPTs into the licensing process is intended to improve safety outcomes. One question that needs to be addressed relates to the time in the Graduate Licensing Scheme (GLS) at which the HPT is best taken to yield the best effect. Research of HPTs largely focuses on how they are conducted (test type, style, etc.) and content (e.g. items) and do not tend to consider the HPT within the broader licensing context and this issue has largely remained unaddressed in the literature. Current practice in Australian jurisdictions with HPT is that the HPT is required in order to progress to solo driving, such as to obtain a provisional licence. Two studies (Castro et al. 2021; Muela et al. 2021) included a group of non-drivers with both finding this group to have poorer performance on hazard prediction tests than both novice and experienced drivers. This suggests that implementing HPT prior to starting driving, such as in order to obtain a Learner's permit, may not yield a safety benefit; measuring a skill before the individual has time to practice and develop that skill may not accurately reflect future ability. Furthermore, the safest period of driving for young drivers is when they undertake supervised driving with the risk of being involved in a crash increasing with the commencement of solo driving (Kloeden 2008). As such, the current approach adopted throughout Australia of requiring the HPT in order to progress to a provisional licence or solo driving appears to be a sensible approach.

3.7 Implications of HPT Failure

Hazard perception testing used as part of the licensing process is intended to ensure that those progressing to unsupervised driving meet a minimum level of expectation regarding the ability to detect hazards and be safe road users. The nature of the test is such that people with poor hazard perception skills will fail and be prevented from progressing to unsupervised driving until their hazard perception is at a level deemed acceptable for safe road use. Within a licensing scheme HPT failure has broader implications, such as how long someone should be required to wait before re-taking the test and how hazard perception can be improved in the interim. Furthermore, it is possible that performance on a licensing HPT could be used to identify drivers who possess characteristics (e.g. risk-taking propensity) and be indicative of a greater risk of crash involvement or other unsafe behaviours that may require further intervention. These issues are addressed below.

3.7.1 Time delay to re-test after failed attempt

The primary concern of research regarding hazard perception has focused on the validity of HPTs, the use of HPTs with special interest groups (e.g. medical populations), and investigating the association between hazard perception and safety outcomes (e.g. crashes or driving behaviour). As such, the evidence has supported the use of hazard perception testing in the licensing process, however, the literature has not addressed procedural aspects such as whether subjects should be required to wait some time before retaking the test or be allowed to retake it immediately. There is no real guidance in the literature regarding the length of time before re-taking the hazard perception test. In Australia there is no standard approach. In the ACT there is a 24-hour lockout period before the test can be retaken and in WA the test may only be taken once per day. In Queensland and Tasmania there are unlimited attempts and no lock-out period. South Australia allows unlimited attempts although it is not clear if there is a lockout period, while in Victoria candidates are allowed 2 attempts per day. There does not appear to be a lockout period in NSW.

The inter-test interval for someone who has failed a Hazard Perception Test is an important consideration within licensing schemes. There is the potential for repeat testing to result in a practice effect rather than a learning effect, such that subsequent tests may measure improvement in the individual's ability to take the test (e.g. response time) rather than reflecting improvements in hazard perception. Is it possible for an individual's hazard perception ability to improve from one test to the next in the absence of more experience or some other intervention (e.g. training)? It must also be asked: If an individual's hazard perception skill is not sufficient to pass the first time, how likely is this skill to have changed in the next test if there has been no change in experience or other improvement in hazard perception? While there has been no systematic study of the potential confounding of practice effects and repeat testing, there is some evidence. Horswill et al. (2021) examined the effects of a brief training intervention on hazard perception (measured using reaction time) on a sample of Australian University students. This study compared the performance of a treatment group (those who received the training) before and after training, to a control group (no training) with a second test undertaken after an interval of one hour. While the treatment group was found to have a larger improvement in median reaction time compared to the control group, of interest is that the control group also demonstrated a reduction in reaction time by a median of 0.12 seconds (Horswill et al. 2021). The potential for a practice effect exists though the effect of this, particularly regarding HPT for licensing, is unknown.

Another possibility for consideration is that some people may attempt to memorise items or, having experienced the same item on a number of attempts, may have learned the correct response by trial and error. While there is some possibility that this could lead to a small improvement in hazard perception – learning to pass the items may prompt some subconscious learning of hazard perception, such as scanning for and identifying hazards – it is likely that the HPT, for these people, would potentially be a measure of memory recall rather than hazard perception skill. This risk can be minimised by randomly generating a hazard perception test from a larger pool of items or having several versions of the HPT using different items. Additionally, a lock-out period from failure to next attempt could help minimise this effect by reducing or removing the ability for repeated attempts in a short period of time, making it less likely that items can be memorised, and adding a mandatory period between tests during which the individual has the opportunity to improve their hazard perception by either gaining more driving experience, practice, or undertaking hazard perception training.

Taken at face value, failure on an HPT indicates a deficit in hazard perception. Ideally, people who fail the HPT would gain additional driving experience before retaking the test, which should result in improved hazard perception and increase both safety and likelihood of passing the test. It is not clear from the literature the amount of experience required to improve hazard perception, although the evidence clearly indicates that drivers with greater driving experience have better hazard perception (although this would be dependent on the quality of the experience). Rather than assume hazard perception skill will improve naturally with time and experience (although experienced drivers perform better on hazard perception than novice drivers, at the individual level hazard perception skill will be variable), a more reliable means would be to ensure people develop the necessary skill by undertaking hazard perception training.

3.7.2 Interventions for improving hazard perception performance

Failing an HPT suggests that an individual's ability to detect hazards requires further development, which usually comes with increased driving experience. People who fail a licensing HPT should be encouraged to obtain more experience, though there is some evidence that other factors, such as confidence, risk sensitivity, the perceived risk of the situation, and belief in own driving ability may also influence hazard perception scores (e.g. Castro et al. 2021; Egea-Caparrós et al. 2016; Ehsani et al. 2020; Farrand et al. 2001). Simply relying on the passage of time may not be sufficient to ensure the adequate development of hazard perception skills and the amount of time or experience necessary for this to occur is difficult to determine due to various definitions of experience adopted throughout the literature. An alternative approach is to accelerate the improvement in hazard perception through training.

There is a large body of research addressing the effectiveness of hazard perception training covering a variety of training programs and modalities (e.g. computer based, driving simulators, virtual reality), though there are common themes and approaches to training, including instruction, feedback, and practice (see inter alia Fisher et al. 2017; Horswill et al. 2021; Krishnan et al. 2019; Plumert et al. 2021; Prabhakharan et al. 2024; Yamani et al 2016). The general approach to hazard perception training is to expose trainees to hazardous traffic scenarios requiring them to identify hazards or perform some other action (e.g. braking in a driving simulator), provide feedback on behaviour and performance, provide additional information and explanation of the hazardous situation and appropriate responses (including theory or demonstrations of expert driver behaviour), and allowing trainees to practice hazard perception further in additional hazard scenarios.

In many cases the hazard scenarios are similar in type and presentation as those used in HPTs for licensing and have used photos or videos based on either real-world footage or CGI to present hazards from the driver's perspective. In general, evidence indicates that hazard perception training appears to improve the performance of hazard perception tasks for trained drivers, with reported improvements in road scanning behaviour, increased monitoring of critical areas (i.e. locations where hazards or cues can be found; Katrahmani et al. 2017; Meir et al. 2014; Petzoldt et al. 2013; Pradhan et al. 2007), improved reaction times (Horswill et al. 2021; van der Kint et al. 2024; Wetton et al. 2013), and greater likelihood of detecting hazards (Fisher 2008; Glassman et al. 2022; Horswill et al. 2021; Muttart et al. 2019; Šeibokaitė et al. 2022; Thomas et al. 2017; Zafian et al. 2016). The literature has largely focussed on the effects of hazard perception training for improving hazard perception as measured using HPTs, which are often similar, if not identical, to the content of the training. The evidence suggests that hazard perception training appears to improve performance on hazard perception testing, though the evidence is confounded by a number of factors such as the effects of driving experience and exposure and the inability to determine if improvements in HPT performance are due to training or a test-related practice effect (i.e. improvements may be due to familiarity with undertaking the test rather than improved hazard perception per se). A small number of studies have examined the effect of hazard perception training on real world safety outcomes, including driving behaviour and crashes (Horswill et al. 2022; Roberts et al. 2021; Thomas et al. 2016).

Horswill et al. (2022) conducted a randomised controlled trial to examine the effect of a 6-week multi-session online self-paced hazard perception training program based on videos of real-world crashes on the driving behaviours of trainees. Over the course of the program trainees completed one 30 minute training session per week consisting of 5 types of exercise, including hazard prediction, crash analysis (e.g. why a crash occurred and how to avoid it), commentary of driving scenarios (i.e. a voice over commentary of the driving environment by an experienced driver with examples of crashes to illustrate the expert's comments), hazard perception practice with comparisons to expert drivers (feedback showing trainee vs. expert responses), and real-world transfer (practising different skills during real-world driving). Real-world driving behaviour was measured using dashcams with accelerometers installed to the participants' own vehicles for a one-month baseline period prior to training and 2 months post training. An untrained control group also had their driving monitored over the same period. This study found that, compared to the control group, trained drivers had significant improvements in heavy braking events (i.e. fewer) and speeding behaviour (i.e. amount of time spent speeding and magnitude of speeding decreased). This effect was observed for post-training measures taken an average of 8 weeks after the completion of training.

Two other studies examined safety-relevant outcomes of the Risk Awareness and Perception Training (RAPT) program developed in the USA (Roberts et al. 2021; Thomas et al. 2016). The RAPT program has undergone numerous refinements since it was first developed and the following studies examined different versions of the program.

Roberts et al. (2021) examined the effect of RAPT on the crash involvement of young drivers based on socio-economic status. The version of RAPT they used consisted of one session during which trainees completed 3 modules (pre-test, 8-minute training session, and post-test) during which they were shown 9 driving scenarios and asked to indicate where they would look to identify potential hazards. The 5,306 participants who had recently passed their driving test to progress to a provisional or unrestricted licence were randomly assigned to the training or control group who undertook the pre-test only. During the training phase they were exposed to scenarios in which they were found to typically make safety-related mistakes. They were provided feedback and given an explanation on how to mitigate the mistake. A separate control group completed the pre-test module only. Crashes and moving violations for around one year following the training were obtained from the California DMV. This study found a protective effect of RAPT for trained drivers for the first 12 months following training. Analysis indicated a positive association between poverty rates and crashes such that the number of crashes increased as the level of poverty increased. More specifically, the number of crashes increased by 3% with every 1% increase in poverty. However, among trained drivers, for every 1% increase in poverty the number of crashes reduced by 5%. Furthermore, at higher levels of poverty (greater than 30%, low SES) trained participants had a lower number of crashes compared to untrained participants but, when poverty was low (i.e. <5%, higher SES), no differences between trained and untrained drivers were observed. It should be noted, however, that this study failed to control for the potential confounding of driving exposure such that the results may simply reflect a lack of driving opportunity for young drivers with low socio-economic status. These findings suggest RAPT training had a protective effect among drivers of low SES.

In a separate study Thomas et al. (2016) examined the effect of RAPT-3 on crash involvement for trained drivers. The RAPT-3 training involved one 40-minute session during which trainees were presented with 9 driving scenarios in which there is an inherent risk of a collision with another vehicle or pedestrian; hazards were either visible or hidden (e.g. obscured by another vehicle). Trainees were shown a plan view (i.e. top-down) of each scenario accompanied by explanations of the risky elements. They then viewed sequences of snapshots (i.e. still images) from the driver's perspective and were required to click on the relevant areas to detect a hazard. If the trainee correctly identified the critical areas, the training proceeded to the next scenario – if not the scenario was repeated from the beginning (i.e. top-down view and explanations). Four trials were allowed for each scenario. Participant crash and violation data were obtained from the California DMV for the 12-month period following licensure (RAPT was undertaken by young drivers aged 16-18 who had just passed their on-road driving test to obtain a provisional or unrestricted licence). This study found that drivers receiving the RAPT-3 training demonstrated significant improvements in their hazard perception test performance post-training, though the effect of training on crashes was mixed. No main effect of training was found for crash involvement, but RAPT-3 trained males had a significantly lower crash rate than control group males; females in the RAPT trained group were observed to have a higher crash rate than control group females, although this difference was not statistically significant. Additionally, training was found to have no effect on the time to first crash. No effect of training on traffic violations was detected. While training appears to have improved hazard perception, with the exception of males, this does not appear to have translated into reductions in crash involvement. It is possible that young males, who are generally observed to have a higher crash risk than young females (Austroads and Kloeden 2009; Jones 2016), benefit more from improved hazard perception than do young females, for whom hazard perception improves but crash risk remains unchanged. Indeed, statistical analysis undertaken in this study indicates that the crash rates for female participants were comparable for both groups.

Using crash involvement as a measure of efficacy for hazard perception training is not without limitations. While it might be expected that people who have improved their hazard perception through training should be less likely to be involved in crashes and therefore have fewer crashes, a more accurate operationalisation would be to determine the number of crashes that are successfully avoided. However, crashes that do not occur are not recorded in any official database and it would be difficult to objectively record this data, particularly for very large samples. Another issue with the use of crash involvement relates to the nature of the crashes that occur. There are a number of scenarios in which study participants may be involved in a crash which does not reflect their hazard perception ability. First, if the participant is in a vehicle that is struck by another driver where their (study participant) hazard perception is not relevant (e.g. rear end). Second, some crash types and causes may not be addressed by training (e.g. single vehicle crashes due to slippery road conditions), or participants may be involved in crashes due to high speeds or impaired driving (e.g. DUI, distraction, or fatigue). Third, it is possible that in some circumstances participants correctly identify and react to a hazard, but a crash may result if there is insufficient time between detecting a hazard and responding or where the participant fails to select an appropriate response. In some cases, manoeuvring to avoid a collision may also result in a separate collision (e.g. due to loss of control while swerving to avoid striking an animal in the road). Finally, many studies examining the association between some intervention and crashes are limited due to insufficient frequency of crashes occurring (which may or may not be exacerbated by a successful intervention reducing the number of crashes). A suitable number of crashes will be necessary to detect a statistically significant effect, which may require substantially larger sample sizes or longer observation periods than adopted in some studies.

Hazard perception training appears to have a generally positive effect on hazard perception skills, particularly on the performance of HPTs. There is some evidence that training can lead to improvements in real-world driving behaviours (e.g. speed and braking) but limited evidence that it can reduce crash involvement. As such, the benefit of training appears to be useful for improving performance on Hazard Perception Tests but the effect on real-world driving and safety outcomes requires further investigation. These findings have implications for HPT in the licensing process. If training increases the likelihood of people passing the licensing HPT but is limited in terms of improving safety, then referral to hazard perception training may be counterproductive. However, given the small number of studies investigating the effects of hazard perception training on real-world safety, more research in this area is warranted.

3.7.3 Using failed HPT attempts to identify individuals at risk of crashing or other unsafe behaviour

There is evidence that performance on Hazard Perception Tests can be used to distinguish between drivers based on level of driving experience. It is also possible that differences in hazard perception performance may be able to identify other groups with an increased risk of crashing. First, there is some evidence that HPT performance is associated with crash involvement (see section 5), though the strongest evidence of this indicates that the ability to distinguish between groups based on safety is only apparent for those who take at least 3 attempts (fail twice) to pass the HPT and that these people may require assistance or interventions to improve their safety (Boufous et al. 2011).

Some studies have found evidence that people with driving violations have poorer hazard perception than drivers with no violations. For example, Wu et al. (2021) found that people with driving violations are worse at identifying the location of hazards than those with no violations, though this did not impact their performance on hazard prediction. Similarly, Sun et al. (2021) found that taxi drivers with violations had significantly higher reaction times on an HPT (i.e. slower to detect hazards) and also scored higher on other behavioural measures that may indicate an increased risk of crashing, including anxiety, distress reduction, and risky driving styles. These findings offer some suggestion that it may be possible to identify motorists more likely to commit driving offences based on their HPT results.

Other studies provide some indication that HPT results may also be indicative of fitness to drive. In an examination of drivers' ability to anticipate hazards following a traumatic brain injury (TBI), Preece et al. (2011) found that drivers with a TBI were 0.75s slower at anticipating hazards than a control group of drivers with no TBI and while the severity of TBI was not related to hazard perception performance, lower HPT scores were associated with increased duration of post-traumatic amnesia. Studies have also demonstrated declines in hazard perception performance for older drivers (aged 60+) such that older drivers demonstrate poorer hazard perception than younger drivers (Folli et al. 2023), while reaction time has been found to reduce with age for drivers over the age of 65 (Horswill et al. 2011), and hazard perception performance has been found to deteriorate with declining cognitive function and visual ability (Folli et al. 2023). Glen et al. (2016) also found that people with visual deficits had impaired hazard perception, though this was due to difficulties identifying hazards that appear in the area affected by the visual deficit. It is unlikely that an HPT would be needed to identify people with a visual deficit.

While there is evidence of a relationship between HPT performance and risky driving, relying solely on HPT results to label an individual as unsafe may not be appropriate as other factors could be contributing to the increased risk (Boufous et al. 2011). Repeated failures on an HPT may be a suitable indicator to refer drivers to an intervention such as hazard perception training. Rather, HPTs may be best suited as a screening tool for drivers referred to interventions or assessment based on other reasons or behaviours (e.g. fitness to drive assessment) and used to identify whether those presenting for intervention have deficits in hazard perception that should also be addressed.

4. HPT Validity

There are 2 primary methods for assessing the validity of an HPT to measure hazard perception skill. The first is to compare the performance on the HPT with another established measure of hazard perception skill or some other relevant construct (e.g. crash involvement). The second method is to assess differences in HPT performance between groups who are assumed to differ in hazard perception ability. The most common method adopted throughout the literature is to investigate the ability of a given HPT to differentiate between novice and experienced drivers. It is assumed that experienced drivers have more highly developed hazard perception skills than novices so a test could be considered valid if the scores can be used to discriminate between the 2 groups or if the experienced drivers are found to outperform novice drivers. It is worth noting that the development of HPTs adheres to these principles and uses these criteria in the selection of test items (see discussion above). The validity of HPTs based on establishing differential performance based on experience is summarised below; the association between HPT scores and real-world safety outcomes, including crash involvement, are discussed in Section 5.

Throughout the literature studies following either the hazard perception or hazard prediction paradigms using either video (real or CGI) or still images have demonstrated the ability to distinguish between novice and experienced drivers based on measures of reaction time or accuracy (Castro et al. 2014; Crundall 2016; Crundall et al. 2021; Endriulaitiene et al. 2022, Gugliotta et al. 2017; Hill et al. 2019; Horswill et al. 2020; Manley et al. 2020; Sun et al. 2019; Tūské et al. 2019; Ventsislavova et al. 2022, Vlakveld 2014; Wu et al. 2021). However, some studies have failed to produce this effect (Sagberg and Bjørnskau 2006; Underwood et al. 2013; Ventsislavova et al. 2019). There are more studies demonstrating group-based differences in hazard perception performance than those that do not, which may be due to differences in methodologies or test characteristics. For example, Sagberg and Bjørnskau (2006) found no differences based on level of driving experience, though there was a large discrepancy between group sizes (130 novices vs. 28 experienced drivers) with the majority of participants novices with less than 9 months of driving experience. The categorisation of novice versus experienced drivers and young versus older drivers was also variable across studies. Novice drivers could range from either newly licensed drivers or learner drivers waiting to take a driving test (Sagberg and Bjørnskau 2006; Sumer et al. 2007; Ventsislavova et al. 2016) to those with less than 8 years driving experience (Gugliotta et al. 2017). The driving experience of experienced drivers ranged from a minimum of one year (Ventsislavova et al. 2018) to 15 years (Horswill et al. 2011). Moran et al. (2019) noted ranges of 0.1 to 4.5 years for novices and 2 to 29 years for experienced drivers.

5. Safety Effects of HPTs

Evidence for the safety effects of HPTs can be found in 3 studies linking the performance of a licensing HPT on crash involvement. Wells et al. (2008) undertook a large study of young learner drivers from the UK examining driver cohorts recruited from November 2001 to August 2005 investigating how they undertake driver training and testing, and their experiences as new drivers, including crashes, offences, attitudes and other behaviours. Over this period, an HPT measuring reaction time was incorporated into the licensing theory test. A random sample of young drivers waiting to take their practical test were selected every 3 months and sent a series of self-report questionnaires addressing their learning and driving experience. Four driving experience questionnaires were sent at 6 (~10,000 responses), 12 (~3,600 responses), 24 (~1,000 responses), and 36 (~250 responses) months after passing the practical driving test. 20,512 questionnaires were sent at the 6-month mark with a 49% response rate with some attrition over the course of the study with around 250 responses received for the questionnaire sent after 36 months. When investigating the effect of hazard perception testing Wells et al. (2008) were able to compare those who were licenced without taking the HPT (as it was not implemented at the time) to those who completed the HPT as part of licensing, and also able to further investigate the effect of differences in HPT for the latter cohort. After controlling for confounding factors, including age, sex, and driving experience and exposure, they found that compared to the non-HPT cohort, those who had passed the HPT reported 3% fewer crashes that did not involve manoeuvring at low speeds (i.e. crashes that might be influenced by hazard perception) and for which they accepted some liability in their first year of driving. When comparing drivers in the HPT cohort, for those with higher HPT scores the rate of non-low speed manoeuvre crashes was 4.5% lower than the lowest scoring HPT group (who still passed the test). These findings demonstrate that the introduction of the HPT appear to have positive effects on safety as intended, though the actual effect appears somewhat small. Although this was a large, detailed study with a randomly selected sample, final data was influenced by a potential self-selection bias and not entirely representative of the broader public (higher proportion of females, younger people slightly over-represented), and crash experience was self-reported. One of the main limitations for the use of self-reported crash is the potential bias due to under-reporting of actual crashes. Official records on the other hand provide a more reliable measure, but only for crashes that are reported to police, crashes of lower severity (i.e. no injury to road users) may be excluded while other crashes may also not be reported (Agran et al. 1990; Macpherson et al. 2024). Wells et al. (2008) found respondents reported less serious crashes that would not appear in official crash records, though the potential for under-reporting of more serious crashes remains a limitation of this study.

In Australia, Boufous et al. (2011) investigated the impact of hazard perception testing on newly licensed drivers as part of the DRIVE study undertaken in NSW. A large sample (N=20,822) of young drivers aged 17 to 24 who had held their P1 license for 12 months or less were recruited via mail during June 2003 to December 2004 and completed a baseline questionnaire (driving experience, risk taking, demographic characteristics) that was linked with records on the number of HPT attempts and police reported crashes for the 12 months following completion of the HPT measuring reaction time. After controlling for sociodemographic characteristics, learning experience, and behavioural factors, the number of failures on the HPT was found to be associated with the risk of being involved in a crash. Compared to drivers who passed the HPT on the first attempt, for those who had failed at least twice (i.e. minimum 3 attempts to pass), the risk of crashing was 1.83 times greater (95%CI 1.27-2.63). More detailed analysis demonstrated that males who failed at least twice had an increased crash risk of 2.5 times (95%CI 1.5-4.1) but no increase was observed for females. Young drivers from remote areas showed 5.5 times increase in crash risk (95%CI 1.63-18.71). These findings suggest performance on the HPT is related to safety outcomes but the ability to discriminate between safe and unsafe drivers is only apparent after at least the third attempt to pass the test. Findings also suggest that people who fail repeatedly, particularly young males, may require support to lower their risk of future crashing. It would appear that those who fail the HPT at least twice possess certain characteristics that are related to test failure and that investigation and understanding of these characteristics and how they relate to performance on HPT (e.g. language comprehension) and crash risk (e.g. risk-taking propensity) will aid the development of appropriate interventions for these drivers.

More evidence of the association between HPT performance and safety outcomes is provided by Horswill et al. (2015) who recruited a large sample of young drivers who had completed an HPT as part of the licensing process in Queensland (Horswill et al. 2015). Using self-reported data regarding HPT score and crash involvement (number of crashes) at the time of taking the HPT (Time 1) and self-reported crash involvement in the 12 months following the test (measured T2, 12 months later) Horswill et al. (2015) sought to determine whether performance on the HPT was related to crash involvement. After adjusting for confounding factors (age, sex, driving exposure) Horswill et al. (2015) found that for every one second increase in HP score, the risk of being involved in a retrospective crash (i.e. crashes prior to taking the HPT measured at T1) increased by 10% and a 21% increase in crashes in the 12 months following completion of the HPT. To investigate the association between HPT and crashes further, Horswill et al. (2015) calculated a 'pass mark' that maximised the sensitivity and specificity of predicting crash involvement and found that those who failed the test were 17% more likely to have been involved in a crash prior to undertaking the HPT and had a 25% increase in the likelihood of having a crash in the 12 months following the test. These findings further support the association between HPT performance and crash risk, though the effect is not strong due to 'noise' in the self-reported crash data (Horswill et al. 2015). Other limitations of this study are the reliance on self-reported data for HP scores and crashes and while the samples at T1 and T2 were large, their nature (i.e. self-selected participants) may limit the generalisability of findings. In contrast to these findings, as part of their evaluation of the benefits of RAPT-3 hazard perception training for novice drivers (see Section 3.8 for more detail), Thomas et al. (2016) also examined the correlation between pre- and post-training HPT scores on police-reported crash involvement for the 12 months following completion of training. They found that HPT scores at either pre- or post-training were not correlated with crash involvement.

Other evidence for the association between hazard perception testing and safety outcomes is provided by Botzer et al. (2019) who examined the relationship between HPT scores and hard braking events. A small sample of students (N=33, 26 females, mean age 24.3) completed a brief hazard perception test using reaction times as a measure of hazard perception and driving behaviour was measured using a smartphone application to detect harsh braking events as recorded by GPS and accelerometer data for a period of 6 weeks. They found that there was a negative relationship with hazard perception scores and harsh braking, indicating that worse hazard perception was associated with an increase in the number of harsh braking events. They also found that males had more harsh braking events than females. Botzer et al. (2019) suggest that harsh braking may be the result of poor hazard perception skill impacting the ability to identify hazards such that they are identified and reacted to later. However, while their evidence clearly shows a relationship, harsh braking may be explained by other factors, including differences in traveling speed or risk sensitivity. Furthermore, in some cases harsh braking may also be an appropriate response and indicative of successful hazard detection and avoidance. Four studies have demonstrated a link between HPT performance and safety outcomes, particularly the risk of being involved in a crash. One study failed to find a correlation between HPT scores and crash involvement. The studies examining crash outcomes had large samples but were not without limitations, primarily related to the possibility of self-selection bias in the sample (all studies) and reliance on self-reported data (Horswill et al. 2015; Wells et al. 2008). The findings of Boufous et al. (2011) appear to provide the best estimate of the effects as they are based on official records of both HPT results and crash involvement. Botzer et al. (2019) provide some evidence that performance on HPTs transfers to real-world reaction time, though more research in this area is required.

5.1 Mechanisms through which HPTs influence safety

There have been no studies explicitly examining the mechanisms through which the testing of hazard perception skills influence safety. As a measure of ability, the HPT itself may be indicative of a level of risk, but it does not modify that risk. As such, there are likely 2 key mechanisms through which HPTs influence safety.

First, the preceding discussion has highlighted the links between hazard perception skill and a range of driving-related outcomes. On the surface, measuring hazard perception skill and ensuring a minimum level of proficiency in order to progress through a graduated licensing scheme, particularly from supervised to unsupervised driving (i.e. learners to provisional/probationary licence) might be expected to ameliorate the future risk of crashing to some extent. At some point, however, all drivers meet this minimum threshold to obtain their provisional/probationary licence though there is some evidence that even after meeting this criterion, drivers with poor hazard perception skills (as indicated by multiple attempts to pass an HPT) maintain a higher risk of crashing than those with better performance. Thus, while the HPT is intended to ensure a minimum level of hazard perception competence, the act of taking or passing the test does not necessarily improve this skill or reduce risk for some drivers.

The second, and likely main mechanism through which an HPT in the licensing process influences safety is by delaying the progression to unsupervised driving for drivers with deficits in this skill. The greatest period of crash risk for newly licenced young drivers is the first 3 to 6 months following the progression to unsupervised driving (Kloeden 2008; Wells et al. 2008). Thus, depending on the quality of the experience gained, any measure that prolongs the supervised driving period may be beneficial.

6. Conclusion

Several methods have been used to measure hazard perception, with the primary methods being the presentation of hazardous scenarios via a video clip and recording either reaction time or ability to predict what happens next. While there are some exceptions, the preponderance of evidence suggests both approaches are valid methods for assessing an individual's ability to detect hazards in the road environment. A common criticism of the standard HPT based on reaction time measures is that it is insufficient to determine the stimulus (hazard) to which the person has responded, while other factors such as sensitivity to risk can influence results. The hazard prediction test overcomes these issues by requiring the subject to predict how the scenario unfolds and may yield a more robust measure of hazard perception skill.

Modern HPTs use video clips to assess hazard perception skill, with both real-world footage and CGI clips proving suitable to assess hazard perception, though some argue that test items should be based on real examples of hazards. While real-world videos may be more valid representations of hazards, they can be limited in the types of hazards they can show (e.g. only daylight driving, high quality clips, and only those hazards encountered while recording footage) and some hazards may be difficult to capture and then not be assessed. While CGI clips may be staged, the types of hazards that can be portrayed is virtually limitless and it is possible to produce a variety of examples.

Performance on HPTs has been associated with crash involvement, suggesting drivers who pass the HPT on the first attempt have a lower risk of crashing than drivers who require at least 3 attempts to pass, particularly males and drivers from remote areas, though the evidence is limited. Differences in HPT performance have also been noted for driving offenders, medical populations, and age-related deficits. HPT may be a useful tool in assessments of fitness to drive, though this requires further investigation.

Hazard Perception Tests contribute to safety by ensuring a minimum level of performance deemed necessary to drive safely, while failing the test delays the commencement of unsupervised driving for drivers who are potentially less safe than others. Methods for improving hazard perception performance include the gaining of more driving experience, though this will be dependent on the quality of the experience gained. Hazard perception training can improve hazard perception skill and performance on HPTs and may lead to improvement in driving behaviour, though the effects on crashes are limited. There is a risk that hazard perception training may increase the likelihood of people passing the hazard perception test without a concomitant reduction in crash risk, though this requires further investigation.

References

Agran PF, Castillo DN and Winn DG (1990) 'Limitations of data compiled from police reports on pediatric pedestrian and bicycle motor vehicle events,' *Accident Analysis & Prevention*, 22(4):361-370.

Austroads (2009) 'The crash and offence experience of newly licensed young drivers,' AP-R331-08, Austroads, Sydney, NSW.

Botzer A, Musicant O and Mama Y (2019) 'Relationship between Hazard-Perception Test scores and proportion of hard-braking events during on-road driving – An investigation using a range of thresholds for hard-braking,' *Accident Analysis & Prevention*, 132: Article 105267, doi:10.1016/j.aap.2019.105267.

Boufous S, Ivers R, Senserrick T and 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 Injury Prevention*, 12(5):475-482.

Castro C, Padilla J-L, Doncel P, Garcia-Fernandez P, Ventsislavova P, Eisman E and Crundall D (2019) 'How are distractibility and Hazard Prediction in driving related? Role of driving experience as moderating factor,' *Applied Ergonomics*, 81, Article 102886, doi:10.1016/j.apergo.2019.102886.

Castro C, Padilla J-L, Roca J, Benítez I, García-Fernández P, Estévez B, López-Ramón MF and Crundall D (2014) 'Development and validation of the Spanish Hazard Perception Test, *Traffic Injury Prevention,* 15(8):817-826, doi:10.1080/15389588.2013.879125.

Castro C, Ventsislavova P, Garcia-Fern P and Crundall D (2021) 'Risky decision-making and Hazard Prediction are negatively related and could be assessed independently using driving footage,' *Psychology Research and Behaviour Management,* 14:857-876, doi:10.2147/PRBM.S305979.

Chou W-S and Chuang P-C (2014) 'A study on the analysis and design of drivers' Hazard Perception Test,' *Asian Transport Studies,* 3(2):220-233, doi:10.11175/eastsats.3.220.

Crundall D (2016) 'Hazard prediction discriminates between novice and experienced drivers,' *Accident Analysis and Prevention*, 86:47-58, doi:10.1016/j.aap.2015.10.006.

Crundall D, van Loon E, Baguley T and Kroll V (2021) 'A novel driving assessment combining hazard perception, hazard prediction and theory questions,' *Accident Analysis and Prevention*, 149, Article 105847, doi:10.1016/j.aap.2020.105847.

Dingus TA, Guo F, Antin JF, Perez M, Buchanan-King M and Hankey J (2016) 'Driver crash risk factors and prevalence evaluation using naturalistic driving data,' *Proceedings of the National Academy of Sciences*, 113(10), 2636-2641, doi:10.1073/pnas.1513271113.

Dreßler A, Bredow B, Rößger L, Rüdel M and Sturzbecher D (2017) 'Development and trial of an item set for testing traffic perception and hazard avoidance' In Di Bucchianica G, Landry S, Vallicelli A and Stanton NA (Eds.) *Advances in Human Aspects of Transportation,* 84:1003-1013, doi:10.1007/978-3-319-41682-3_82.

Egea-Caparrós DA, García-Sevilla J, Pedraja MJ, Romero-Medina A, Marco-Cramer M and Pineda-Egea L (2016) 'Late detection of hazards in traffic: A matter of response bias?' *Accident Analysis & Prevention*, 94:188-197, doi:10.1016/j.aap.2016.06.002.

Ehsani JP, Seymour KE, Chirles T and Kinnear N (2020) 'Developing and testing a Hazard Prediction Task for novice drivers: A novel application of naturalistic driving videos,' *Journal of Safety Research*, 73:303-309, doi:10.1016/j.jsr.2020.03.010.

Endriulaitienė A, Šeibokaitė L, Markšaitytė R, Slavinskienė J, Crundall D and Ventsislavova P (2022) 'Correlations among self-report, static image, and video-based hazard perception assessments: The validity of a new Lithuanian hazard prediction test,' *Accident Analysis & Prevention*, 173: Article 106716, doi:10.1016/j.aap.2022.106716.

Farrand P and McKenna FP (2001) 'Risk perception in novice drivers: The relationship between questionnaire measures and response latency,' *Transportation Research Part F: Traffic Psychology and Behaviour,* 4(3):201-212, doi:10.1016/S1369-8478(01)00024-9

Fisher DL, Young J, Zhang L, Knodler M and Samuel S (2017) 'Accelerating teen driver learning: Anywhere, anytime training,' AAA Foundation for Traffic Safety, accessed 3 February 2025.

Fisher DL (2008) 'Evaluation of PC-based novice driver risk awareness,' National Highway Traffic Safety Administration (NHTSA), accessed 3 February 2025.

Folli M and Bennett JM (2023) 'Factors that predict hazard perception in older adult road-users: A systematic review,' *Transportation Research Part F: Traffic Psychology and Behaviour*, 99:19-35, doi:10.1016/j.trf.2023.10.008.

Glassman J, Yahoodik S, Samuel S, Young J, Knodler MK, Zhang T, Zafian T, Fisher DL and Yamani Y (2022) 'Booster dose of attention training program for young novice drivers: A longitudinal driving simulator evaluation study,' *Human Factors*, 66(3):933-953, doi:10.1177/00187208221109993.

Glen FC, Smith ND, Jones L and Crabb DP (2016) '"I didn't see that coming": simulated visual fields and driving hazard perception test performance,' *Clinical and Experimental Optometry,* 99(5):469-475, doi:10.1111/cxo.12435.

Goodge T, Kroll V, Vernon M, Ventsislavova P and Crundall D (2021) 'A comparison of cybersickness symptoms across 360-degree hazard perception and hazard prediction tests for drivers,' *Applied Ergonomics*, 97: Article 103549, doi:10.1016/j.apergo.2021.103549.

Grayson GB, Maycock G, Groeger JA, Hammond SM and Field DT (2003) *Risk, Hazard Perception and Perceived Control*, Transport Research Laboratory.

Gugliotta A, Ventsislavova P, Garia-Fernandez P, Pena-Suarez E, Eisma E, Crundall D and Castro C (2017) 'Are situation awareness and decision-making in driving totally conscious processes? Results of a hazard prediction task,' *Transportation Research Part F*, 44: 168-179, doi:10.1016/j.trf.2016.11.005.

Gusenbauer M, and Haddaway N R (2020) 'Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources,' *Research Synthesis Methods*, 11(2):181-217, doi:10.1002/jrsm.1378

Horswill MS, Anstey KJ, Hatherly C, Wood JM and Pachana NA (2011) 'Older drivers' insight into their hazard perception ability,' *Accident Analysis & Prevention,* 43(6):2121-2127, doi:10.1016/j.aap.2011.05.035.

Horswill MS, Hill A, Bemi-Morrison N and Watson MO (2021) 'Learner drivers (and their parent-supervisors) benefit from an online hazard perception course incorporating evidence-based training strategies and extensive crash footage,' *Accident Analysis & Prevention,* 161: Article 106340, doi:10.1016/j.aap.2021.106340.

Horswill MS, Hill A, Buckley L, Kieseker G and Elrose F (2022) 'An online hazard perception training course reduces heavy braking, speeding, and over-revving rates during everyday driving,' *Transportation Research Part F: Psychology and Behaviour*, 87:54-68, doi:10.1016/j.trf.2022.03.018.

Horswill MS, Falconer EK, Pachana NA, Wetton M and Hill A (2015) 'The Longer-Term Effects of a Brief Hazard Perception Training Intervention in Older Drivers,' *Psychology and Aging, 30*(1):62-67, doi:10.1037/a0038671.

Horswill MS, Hill A and Jackson T (2020) ‘Scores on a new hazard prediction test are associated with both driver experience and crash involvement,’ *Transportation Research Part F: Traffic Psychology and Behaviour*, 71:98-109, doi:10.1016/j.trf.2020.03.016.

Horswill MS, Hill A, Rodwell D, Larue GS, Bates L and Watson B (2021) ‘A brief and unsupervised online intervention improves performance on a validated test of hazard perception skill used for driver licensing,’ *Transportation Research Part F: Traffic Psychology and Behaviour*, 78:130-136, doi:10.1016/j.trf.2021.02.003.

Horswill MS, Hill A, Silapurem L and Watson MO (2021) ‘A thousand years of crash experience in three hours: An online hazard perception training course for drivers,’ *Accident Analysis & Prevention,* 152: Article 105969, doi:10.1016/j.aap.2020.105969.

Horswill MS, Wetton MA and Hill A (2011) ‘The development and validation of a hazard perception test for use in driver licensing,’ *Accident Analysis & Prevention,* 43(5):1759-1770, doi:10.1016/j.aap.2011.04.007.

Isler RB and Starkey N (4–6 October 2012) ‘Driver education and training as evidence-based road safety Interventions,’ [conference presentation] *Australasian Road Safety Research, Policing and Education Conference*, Wellington, New Zealand.

Jackson L, Chapman P and 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, doi:10.1080/00140130802030714.

Jones SJ (2017) ‘Girls crash too: trends in single vehicle crash rates in young and adult, male and female drivers,’ *Injury Prevention,* 23: 186-189, doi:10.1136/injuryprev-2016-042027.

Katrahmani A, Ahmadi N and Romoser M (2017) ‘Using Situation Awareness as a measure of driver Hazard Perception ability’*, Driving Assessment Conference Archive 9(2017)*, 256-262. doi: https://doi.org/10.17077/drivingassessment.1644.

Kloeden CN (9-12 November 2008) ‘The crash experience of newly licensed young drivers in South Australia’ [conference presentation], *Australasian Road Safety Research, Policing and Education Conference*, Adelaide, South Australia.

Krishnan A, Samuel S, Yamani Y, Romoser MRE and Fisher DL (2019) ‘Effectiveness of a strategic hazard anticipation training intervention in high risk scenarios,’ *Transportation Research Part F: Traffic Psychology and Behaviour,* 67:43-56, doi:10.1016/j.trf.2019.10.004.

Kroll V, Mackenzie AK, Goodge T, Hill R, Davies R and Crundall D (2020) ‘Creating a hazard-based training and assessment tool for emergency response drivers,’ *Accident Analysis & Prevention*, 144: Article 105607, doi:10.1016/j.aap.2020.105607.

Lyon J, Borkenhagen D, Scialfa C, Deschenes M and Horswil MS (2011) ‘Developing a North American static Hazard Perception Test’, *Driving Assessment Conference Archive 6(2011)*, 629-634. doi: https://doi.org/10.17077/drivingassessment.1456.

Macpherson AK, Zagorski B, Saskin R, Howard AW, Harris MA, Namin S and Rothman L (2024) ‘Comparison of the number of pedestrian and cyclist injuries captured in police data compared with health service utilisation data in Toronto, Canada 2016-2021,’ *Injury Prevention*, 30(2):161-166, doi:10.1136/ip-2023-044974.

Malone S and Brünken R (2016) ‘The role of ecological validity in hazard perception assessment,’ *Transportation Research Part F: Traffic Psychology and Behaviour*, 40:91-103, doi:10.1016/j.trf.2016.04.008.

Manley H, Paisarnsrisomsuk N, Hill A and Horswill MS (2020) 'The development and validation of a hazard perception test for Thai drivers,' *Transportation Research Part F: Traffic Psychology and Behaviour*, 71:229-237, doi:10.1016/j.trf.2020.04.011.

Meir A, Borowsky A and Oron-Gilad T (2014) 'Formation and evaluation of Act and Anticipate Hazard Perception Training (AAHPT) intervention for young novice drivers,' *Traffic Injury Prevention*, 15(2):172-180, doi:10.1080/15389588.2013.802775.

Moran C, Bennett JM and Prabhakharan P (2019) 'Road user Hazard Perception Tests: A systematic review of current methodologies,' *Accident Analysis & Prevention*, 129: 309-333, doi:10.1016/j.aap.2019.05.021.

Muela I, Chica AB, Garcia-Fernandez P and Castro C (2021) 'Visual attention in realistic driving situations: Attentional capture and hazard prediction,' *Applied Ergonomics*, 90: Article 103235, doi:10.1016/j.apergo.2020.103235.

Muttart JW, Dinakar S, Fisher DL, Garrison TM and Samuel S (2019) 'Evaluation of a training intervention to improve novice drivers' hazard mitigation when approaching left turn scenarios,' *Transportation Research Record*, 2673(12):474-484, doi:10.1177/0361198119860484.

Petzoldt T, Weiß T, Franke T, Krems JF and Bannert M (2013) 'Can driver education be improved by computer based training of cognitive skills?' *Accident Analysis & Prevention*, 50:1185-1192, doi:10.1016/j.aap.2012.09.016.

Plumert JM, Reyes M, O'Neal EE, Vecera S, Allen S and McGehee D (2021) *Extended evaluation of training programs to accelerate hazard anticipation skills in novice teens drivers*, Safety Research Using Simulation (SAFER-SIM) University Transportation Center, Iowa City, IA, USA.

Prabhakharan P, Bennett JM, Hurden A and Crundall D (2024) 'The efficacy of hazard perception training and education: A systematic review and meta-analysis,' *Accident Analysis & Prevention,* 202: Article 107554, doi:10.1016/j.aap/2024/107554.

Pradhan A, Pollatsek A and Fisher DL (2007) 'Comparison of trained and untrained novice drivers' gaze behaviour in risky and non-risky scenarios' *Driving Assessment Conference Archive* 4(2007), 328-334. doi: https://doi.org/10.17077/drivingassessment.1257.

Preece MHW, Horswill MS and Geffen GM (2011) 'Assessment of drivers' ability to anticipate traffic hazards after Traumatic Brain Injury,' *Journal of Neurology, Neurosurgery and Psychiatry*, 82(4):447-451.

Roberts SC, Zhang F, Fisher D and Vaca FE (2021) 'The effect of hazard awareness training on teen drivers of varying socioeconomic status,' *Traffic Injury Prevention,* 22(6):455-459, doi:10.1080/15389588.2021.1940984.

Rosenbloom T, Perlman A and Pereg A (2011) 'Hazard perception of motorcyclists and car drivers,' *Accident Analysis & Prevention*, 43:601-604, doi:10.1016/j.aap.2010.08.005.

Sagberg F and Bjørnskau T (2006) 'Hazard perception and driving experience among novice drivers,' *Accident Analysis & Prevention,* 38(2):407-414, doi:10.1016.j.aap.2005.10.014.

Sawada T, Uda H, Suzuk Ai, Tomori K, Ohno K, Iga H, Okita Y and Fujita Y (2021) 'The pilot study of the hazard perception test for evaluation of the driver's skill using virtual reality,' *Electronics,* 10(9): Article 1114, doi:10.3390/electronics10091114.

Scialfa CT, Borkenhagen D, Lyon J, Deschênes M, Horswill MS and Wetton M (2012) 'The effects of driving experience on responses to a static hazard perception test,' *Accident Analysis & Prevention*, 45:547-553, doi:10.1016/j.aap.2011.09.005.

Scialfa CT, Deschênes MC, Ference J, Boone J, Horswill MS and Wetton M (2011) 'A hazard perception test for novice drivers,' *Accident Analysis & Prevention,* 43(1):204-208, doi:10.1016/j.aap.2010.08.010.

Scialfa CT, Pereverseff RS and Borkenhagen D (2014) 'Short-term reliability of a brief hazard perception test,' *Accident Analysis & Prevention*, 73:41-46, doi:10.1016/j.aap.2014.08.007.

Šeibokaitė L, Endriulaitienė A, Markšaitytė R and Slavinskienė J (2022) 'Improvement of hazard prediction accuracy after training: Moderation effect of driving self-efficacy and road safety attitudes of learner drivers,' *Safety Science*, 151: Article 105742, doi:10.1016/j.ssci.2022.105742.

Sumer N, Unal AB, Birdal A, Çınar P and Çevikoglu S (2007) 'Comparison of Self-Reported and Computer-Based Hazard Perception Skills Among Novice and Experienced Drivers,' *Driving Assessment Conference Archive 4(2007)*, 160-166, doi: https://doi.org/10.17077/drivingassessment.1231.

Sun L, Cheng L and Zhang Q (2021) 'The differences in hazard response time and driving styles of violation-involved and violation-free taxi drivers,' *Transportation Research Part F: Traffic Psychology and Behaviour*, 82:178-186, doi:10.1016/j.trf.2021.08.011.

Sun L, Hu W-J, Cheng L and Zhang C-L (2024) 'Effects of hazard type and confidence level on hazard perception in young male drivers: an ERP study,' *NeuroReport*, 35(5):299-305, doi:10.1097/WNR.0000000000002007.

Sun L and Hua L (2019) 'Effects of hazard types on drivers' risk rating and hazard response in a video-based hazard perception task,' *PLOS ONE* 14(3):Article e0214226, doi:10.1371/journal.pone.0215744.

Sun L, Zhang M, Qiu Y and Zhang C (2023) 'Effects of sleep deprivation and hazard types on the visual search patterns and hazard response times of taxi drivers,' *Behavioral Sciences*, 13(12): Article 1005, doi:10.3390/bs13121005.

Thomas FD, Korbelak KT, Divekar GU, Blomberg RD, Romoser MRE and Fisher DL (2017) 'Evaluation of an updated version of the Risk Awareness and Perception Training program for young drivers,' US Department of Transport National Highway Traffic Safety Administration, accessed 3 February 2025.

Thomas FD, Rilea SL, Blomberg RD, Peck RC and Korelak KT (2016) 'Evaluation of the safety benefits of the risk awareness and perception training program for novice Teen drivers' (Report No. DOT HS 812 235), National Highway Traffic Safety Administration, Washington DC, accessed 3 February 2025.

Underwood G, Ngai A and Underwood J (2013) 'Driving experience and situation awareness in hazard detection,' *Safety Science*, 56:29-35, doi:10.1016/j.ssci.2012.05.025.

van der Kint ST, van Schagen I, Vlakveld W, Mons C, de Zwart R and Hoekstra T (2024) 'A brief pc-based hazard prediction training program improves young novice drivers' hazard perception skills compared to a control group over time,' *Transportation Research Part F: Traffic Psychology and Behaviour*, 102:64-76, doi:10.1016/j.trf.2024.02.007.

Ventsislavova P and Crundall D (2018) 'The hazard prediction test: A comparison of free-response and multiple-choice formats,' *Safety Science*, 109:246-255, doi:10.1016/j.ssci.2018.06.004.

Ventsislavova P, Crundall F, Baguley T, Castro C, Gugliotta A, Garcia-Fernandez P, Zhang W, Ba Y and Li Q (2019) 'A comparison of hazard perception and hazard prediction tests across China, Spain and the UK,' *Accident Analysis and Prevention*, 122: 268-286, doi:10.1016/j.aap.2018.10.010.

Ventsislavova P, Gugliotta A, Peña-Suarez E, Garcia-Fernandez P, Eisman E, Crundall D and Castro C (2016) 'What happens when drivers face hazards on the road?' *Accident Analysis and Prevention,* 91:43-54, doi:10.1016/j.aap.2016.02.013.

Ventsislavova P, Rosenbloom T, Leunissen J, Spivak Y and Crundall D (2022) 'An online hazard prediction test demonstrates differences in the ability to identify hazardous situations between different driving groups,' *Ergonomics,* 65(8):1119-1137, doi:10.1080/00140139.2021.2016999.

Vlakveld WP (2014) ‘A comparative study of two desktop hazard perception tasks suitable for mass testing in which scores are not based on response latencies,’ *Transportation Research Part F: Traffic Psychology and Behaviour*, 22:218-231, doi:10.1016/j.trf.2013.12.013.

Vogel K, Kircher A, Alm H and Nilsson L (2003) ‘Traffic sense: which factors influence the skill to predict the development of traffic scenes?’ *Accident Analysis & Prevention*, 35(5):749-762, doi:10.1016/S0001-4575(02)00078-7.

Wells P, Tong S, Sexton B, Grayson G and Jones E (2008) ‘Cohort II: A study of learner and new drivers Volume 1 – Main Report,’ Department for Transport, London.

Wetton MA, Hill A and Horswill MS (2011) ‘The development and validation of a hazard perception test for use in driver licensing,’ *Accident Analysis & Prevention,* 43:1759-1770, doi:10.1016/j.aap.2011.04.007.

Wetton MA, Hill A and Horswill MS (2013) ‘Are what happens next exercises and self-generated commentaries useful additions to hazard perception training for novice drivers?’ *Accident Analysis & Prevention*, 54:57-66, doi:10.1016/j.aap.2013.02.013.

Wetton MA, Horswill MS, Hatherly C, Wood JM, Pachana NA and Anstey KJ (2010) ‘The development and validation of two complementary measures of drivers’ hazard perception ability,’ *Accident Analysis & Prevention,* 42(4):1232-1239, doi:10.1016/j.aap.2010.01.017.

Wu B, Sun L and Gu N (2021) ‘Development and validity of a hazard prediction test for Chinese drivers,’ *PLoS ONE,* 16(1): Article e0245843, doi:10.1371/journal.pone.0245843.

Yamani Y, Samuel S, Knodler MA and Fisher DL (2016) ‘Evaluation of the effectiveness of a multi-skill program for training younger drivers on higher cognitive skills,’ *Applied Ergonomics*, 52:135-141, doi:10.1016/j.apergo.2015.07.005.

Yang J, Liu C, Chu P, Wen X and Zhang Y (2021) ‘Exploration of the Relationships between Hazard Perception and Eye Movement for Young Drivers,’ *Journal of Advanced Transportation*, 2021: Article 664299, doi:10.1155/2021/6642999.

Zafian TM, Samuel S, Coppola J, O’Neill EG, Romoser MRE and Fisher DL (2016) ‘On-Road Effectiveness of a Tablet-Based Teen Driver Training Intervention’, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 60(1):1926-1930, doi:10.1177/1541931213601439.

Austroads

Level 9, 570 George Street
Sydney NSW 2000 Australia

Phone: +61 2 8265 3300

austroads@austroads.gov.au
www.austroads.gov.au