A brief and unsupervised online intervention improves performance on a validated test of hazard perception skill used for driver licensing

Mark S. Horswill[a,*], Andrew Hill[a,b], David Rodwell[c], Grégoire S. Larue[c], Lyndel Bates[d,e], Barry Watson[c]

[a] School of Psychology, The University of Queensland, St Lucia, Brisbane QLD 4072, Australia

[b] Minerals Industry Safety and Health Centre, Sustainable Minerals Institute, The University of Queensland, St Lucia, Brisbane QLD 4072, Australia

[c] Queensland University of Technology (QUT), Centre for Accident Research and Road Safety – Queensland (CARRS-Q), Brisbane, Australia

[d] Griffith Criminology Institute, Griffith University, Brisbane, Australia

[e] School of Criminology and Criminal Justice, Griffith University, Brisbane, Australia

*Corresponding author:

Prof Mark Horswill

School of Psychology

The University of Queensland

St Lucia QLD 4072

Australia

E-mail: m.horswill@psy.uq.edu.au

Telephone: +65 7 334 69520; Fax: +65 7 336 54466

Abstract

Drivers who have higher levels of hazard perception skill also tend to have fewer crashes. Training designed to improve this skill has therefore been proposed as a strategy for reducing crash risk. To date, however, hazard perception training has only been evaluated in supervised settings. This means that improvements in hazard perception skill resulting from such training may not generalize to unsupervised situations, which may limit opportunities for large scale roll-out via automated delivery methods. In the present study, we investigated whether a brief video-based training intervention could improve hazard perception skill when drivers completed it online without supervision. The training involved drivers watching videos of traffic scenes, while generating a commentary of what they were searching for, monitoring, and anticipating in each scene. Drivers then compared their own commentary to a pre-recorded commentary generated by an expert driver, hence allowing for performance feedback without an instructor present. A convenience sample of 93 drivers (who did not receive any performance-related incentives) participated in a randomized control study. The training was found to significantly improve response times to hazards in stimuli from the official hazard perception test used for driver licensing in Queensland, Australia, which is known to predict crash involvement. That is, the training was effective in improving hazard perception skill (*Cohen's d* = .50), even though participants were aware that no one was monitoring the extent to which they engaged in the intervention. Given that the training could, in principle, be deployed at scale with minimal resources (e.g. via any online platform that allows video streaming), the intervention may represent a practical and effective opportunity to improve road safety.

Keywords: hazard perception, situation awareness, crash risk, anticipation, driving

1. Introduction

Drivers who are better able to anticipate potentially dangerous situations on the road ahead have fewer crashes (Horswill, Hill, & Wetton, 2015). This skill is referred to as *hazard perception*, and it is known to remain underdeveloped in many drivers (Horswill, 2016a). Consequently, one strategy that has been proposed to reduce crash risk is to improve drivers' hazard perception skill through training. Indeed, there is already a substantial body of research evidence which suggests that hazard perception is relatively easy to improve in both novice and experienced drivers, given an appropriate intervention (Horswill, 2016b).

Video-based commentary drive exercises are among the training techniques that have been found to be effective in improving hazard perception skill (Ābele, Haustein, Martinussen, & Møller, 2019; Cantwell, Isler, & Starkey, 2013; Castro et al., 2016; Crundall, Andrews, van Loon, & Chapman, 2010; Isler, Starkey, & Sheppard, 2011; McKenna, Horswill, & Alexander, 2006; Wetton, Hill, & Horswill, 2013). Typically, the trainee is shown a video clip of traffic filmed from the driver's perspective and is asked to provide a running commentary (Poulsen, Horswill, Wetton, Hill, & Lim, 2010). This involves identifying cues that might flag a dangerous situation, and predicting what might reasonably be expected to happen next. The trainee then watches the same clip again while listening to an overdubbed expert driver commentary. This allows them to receive performance feedback by comparing the expert commentary with their own. The trainee repeats the same process with several different clips, with the goal of trying to generate responses that are similar in quality to the expert commentaries.

Training interventions of this type have been shown to have a positive effect on hazard perception skill for drivers of all ages and experience levels (Horswill, Falconer, Pachana, Wetton, & Hill, 2015; Horswill, Taylor, Newnam, Wetton, & Hill, 2013; Wetton et al., 2013), suggesting that there may be value in making them available at scale, so that as many drivers as possible can benefit. However, one potential obstacle to the mass roll out of such an intervention is that all the evidence to date has been collected under supervised conditions. That is, there was always someone (typically a researcher) who was physically present to supervise the trainee and confirm that they complied with the exercise protocol (e.g. Horswill et al., 2013). It may be that this direct supervision is critical to the success of the intervention and, if so, would substantially increase the costs of mass roll out.

The aim of the present study is to determine whether video-based commentary drive training remains effective in the absence of supervision. If the training works without a supervisor, then the costs associated with mass roll-out may be dramatically reduced. For example, it would be possible for the stimuli to be hosted on an online video streaming platform, allowing trainees to access them from their own networked devices at any time.

In the current study, we will assess hazard perception skill with official stimuli from a hazard perception test used for driver licensing in Queensland, Australia (Horswill, Hill, et al., 2015; Wetton, Hill, & Horswill, 2011). This approach contrasts with previous hazard perception training studies, which have typically employed tests that were developed for research purposes and which therefore received less rigorous validation efforts. Indeed, the Queensland test underwent an extensive development and validation process, and has been shown to distinguish between experienced and novice drivers (n = 150; Wetton et al., 2011), and to predict crash risk both retrospectively (n = 33,105) and prospectively (n = 5,862) among young novices (Horswill et al., 2015). Furthermore, if we discover that the intervention directly affects a licensure test, then this in itself might act as an incentive for drivers to complete the training voluntarily. That is, we will be able to promote the intervention to drivers on provisional licenses, on the basis that it can increase their chances of passing the specific hazard perception test that stands between them and an unrestricted driver's licence.

2. Method

2.1. Participants

124 drivers, recruited from two University research participation pools, completed the study. Of these, 31 were excluded from the analysis because they: (1) had incomplete or potentially inaccurate data due to technical problems, such as issues associated with running the experiment software (which was necessary to run the hazard perception tests) on their own devices (24 drivers); (2) reported more than one attempt at the same hazard perception test (1 driver); (3) reported being interrupted during at least one of the hazard perception tests (5 drivers, one of whom also had technical problems); and/or (4) reported off-screen distractions during at least one of the tests (2 drivers). Drivers in the final two categories were excluded because interruptions or distractions that occurred during the tests would compromise the evaluation. However, we did not exclude people who reported not completing the training component properly, as this might reflect the realities of real-world deployment (though we did examine the training effect both with and without these individuals). The number of excluded participants did not differ significantly between the two experimental conditions, X^2 (1, N = 124) = .12, p = .734. The characteristics of the final sample of 93 drivers are described in Table 1.

Participants received course credit for participating. This study was approved by the Queensland University of Technology and the University of Queensland Human Research Ethics Committees.

Table 1

Participant Characteristics (*N* = 93)

Variable	**Overall**	**Trained group**	**Untrained group**	**Test of group difference**
Age (years)[1]				*Mann-Whitney U = 925, p*
Median	20	19.5	20	*= .293*
Mean	21.4	22.6	20.2	
SD	5.51	7.08	2.87	
Range	17 – 47	18 – 47	17 – 33	
Sex				X^2 (1, *N* = 93) = 1.12, *p* =
Female	66.7%	66.0%	67.4%	.570
Male	32.3%	34.0%	30.4%	
Non-binary	1.1%	0%	2.2%	
Licence Type[2]				X^2 (1, *N* = 93) = 3.65, *p* =
Learner	20.4%	23.4%	17.4%	.601
Provisional[3]	32.3%	31.9%	32.6%	
Open/unrestricted	37.6%	38.3%	37.0%	
Years since passing on-road driving test[4]				*Mann-Whitney U = 1029, p*
Median	3.17	3.33	2.96	*= .689*
Mean	4.22	5.30	3.11	
SD	5.26	6.99	2.05	
Range	0 – 28.3	0.08 – 28.3	0 – 11.7	
Distance driven per year (km)[5]				*Mann-Whitney U = 964, p*
Median	7515	10121	7284	*= .831*
Mean	9314	10017	8595	
SD	8175	9396	6738	
Range	0 – 48000	0 – 48000	0 – 25212	

[1]One participant declined to report their age. [2]Nine participants did not specify their licence type. [3]Can drive unsupervised with restrictions. [4]Does not include learners. [5]Three participants did not report distance driven per year.

2.2. Materials

2.2.4. Online Commentary Drive Training Intervention

The training intervention used in the present study comprised four commentary drive exercises (consistent with previous work; Wetton et al., 2013). This intervention has previously been found to improve hazard perception skill under supervised conditions, both on its own (Wetton et al., 2013) and as part of interventions that also included other components (Horswill, Falconer, et al., 2015; Horswill et al., 2013; Poulsen et al., 2010). After viewing an instruction video, the participant completed all four commentary drive exercises in turn. In each exercise, participants first viewed a video clip depicting part of a car journey (approximately 1 minute), filmed from the driver's perspective (see Figure 1). Their task at this stage, as outlined in the instruction video, was to generate a running commentary while watching the clip, explaining what they could see, noting any hidden elements in the scene, and indicating what might reasonably be expected to happen. They were instructed to speak out-loud unless this was not practical, in which case they could instead generate the commentary in their head. The instructions also advised participants to scan the road ahead and actively look for "clues" (i.e. cues) that might signal a traffic conflict (which was defined as an incident where the camera car would have to slow down or change course to avoid a collision). They were told that, in addition to the general behaviour of other road users, these cues might include specific features such as indicators, brake lights, road signs, traffic signals, and the layout of the road. Additionally, they were advised to monitor locations from which hidden road users might emerge, and that some cues might be in the far distance when they first appeared.

Fig. 1. Screenshot from the commentary drive training intervention

In the second part of the exercise, a version of the same clip overdubbed with a pre-recorded expert driver commentary was played. This was to allow participants the opportunity to compare their own commentary with that of an expert, obtain feedback on their performance and gain an insight into expert-level performance. Trainees were advised to try to make their commentaries more like that of the expert in subsequent exercises. Each expert commentary was scripted by merging details from commentaries generated in real-time by three driving examiners, and each script was performed in synchronisation with the relevant video by a voice artist in the role of "expert driver" (see Poulsen et al., 2010, for further details). In the final part of each exercise, participants were shown examples of real crashes that illustrated some of the hazards noted in the expert commentary, with narration to emphasize the link. The aim of this additional component (which was not used in previous commentary drive exercise studies) was to demonstrate the validity of the expert commentary to trainees.

2.2.1. Hazard Perception Test

Participants' hazard perception skill was measured using video clips drawn from the item pool for the Queensland Government's hazard perception test, which is used for driver licensing in Queensland, Australia (Horswill, Hill, et al., 2015; Wetton et al., 2011). Specifically, novice drivers need to pass this test to progress from a highly restrictive "Provisional 1" licence to a less restrictive "Provisional 2" licence, as they progress towards an unrestricted "Open" licence. The stimuli were presented using custom software designed to emulate the licencing test platform. Two alternate versions of the test were employed in the current study (Version 1 and Version 2) in order to test for changes associated with the training. Each was comprised of 15 items, as in the official licensure test. For the purposes of the present study, the 30 traffic conflicts were selected (using data from Wetton et al., 2011) to make the two versions as similar as possible in terms of: (1) the mean clip response times, (2) the mean standard deviations of clip response times, (3) item content, and (4) the magnitude of novice/experienced driver differences.

Before the first test, participants were shown an instruction video explaining the procedure and how to respond. The instructions explained that they would be shown a series of driver-perspective video clips of traffic footage, which would include other road users. For each clip, their task was to use their computer mouse to click on any other road user that was likely to become involved in a traffic conflict with the camera car, as soon as they predicted that a traffic conflict was likely to happen. Before the second test, participants viewed a text slide that summarized these instructions as a reminder.

The tests were scored by standardizing response times to each traffic conflict using a previously-collected standardisation sample (Wetton et al., 2011). If participants did not respond to a particular traffic conflict, then they received the maximum response time for that item (according to the standardisation sample). The standard scores for each conflict were then averaged for the whole of each test, and converted back into an overall response time in seconds for ease of interpretation. This process was designed to maximize the equivalence of test scores across the two different test versions.

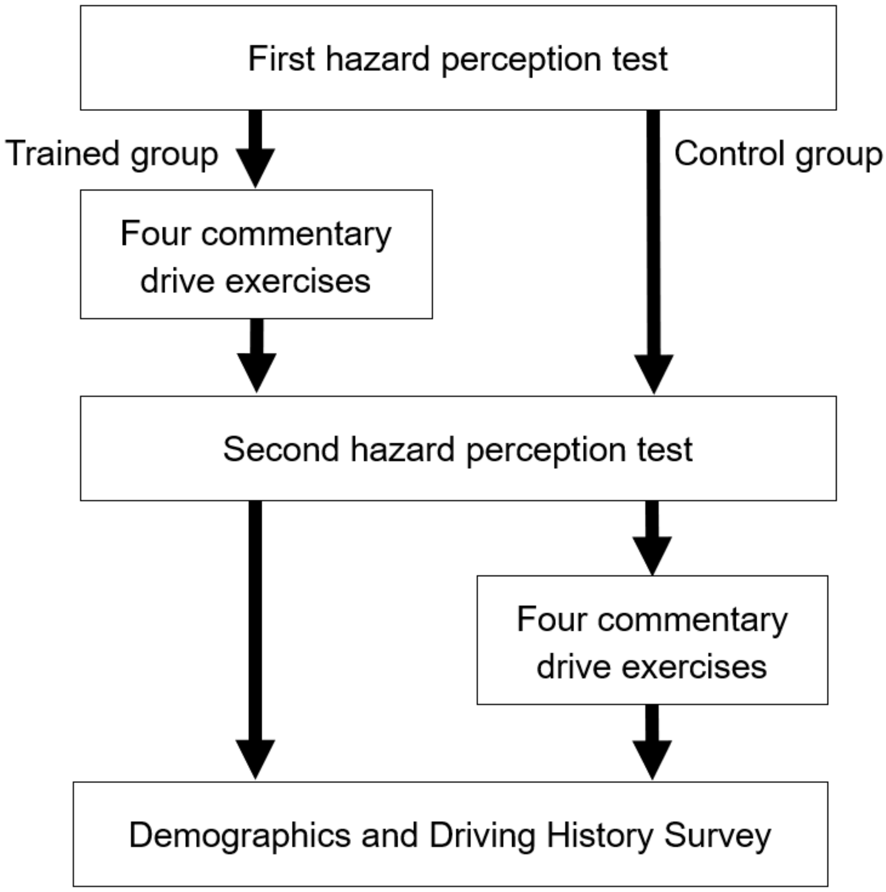
2.2.3. Questionnaires

Participants completed a survey that included questions about demographics and driving history (see Table 1). After each hazard perception test, they were asked to report if there were any problems with the test, and whether they were distracted during the test. After each commentary drive exercise, participants were also asked to report whether they engaged in that exercise as intended (i.e. whether they generated their own commentary and then listened to the expert's commentary).

2.3. Procedure

All participants were asked to access the study materials on a PC (Windows 7 or later) due to the compatibility limitations of the hazard perception test. They were also asked to ensure they would not be disturbed for the duration of the study (up to an hour). All participants completed two hazard perception tests and the commentary drive training intervention (see Figure 2). Each participant was randomly assigned to one of two training conditions, either completing the four commentary drive exercises between the two hazard perception tests ("trained group") or after finishing both tests ("control group"). They were also randomly allocated to one of two test version orders (Version 1 first, or Version 2 first). There was no significant association between test version order and training group, X^2 (1, N = 93) = .26, p = .61. All participants completed the demographics and driving history survey after the video-based tasks.

Fig. 2. Study design



2.4 Statistical Analysis

Statistical analysis was conducted using Jamovi (version 1.6). In order to assess the training effect, each participant's score on the second hazard perception test was subtracted from their score on the first test, yielding a measure of change in response time. This change score was found to have significant skew (*Shapiro-Wilk* $W = 0.76$, $p < .001$) that could not be corrected through transformation. As a result, we tested the training effect by conducting a non-parametric Mann-Whitney U test on the pre-post change scores between the trained and control groups.

3. Results

The trained group improved their hazard perception test scores to a significantly greater extent than the control group, *Mann-Whitney U* = 739, $p = .008$, *rank biserial correlation* = 0.32 (see Figure 3; Table 2). Note that this finding was robust to the violation of normality: An independent-samples t-test yielded a similarly significant result, $t(91) = -2.40$, $p = .018$, *Cohen's d* = -0.50, and the pattern of results also remained unchanged when an ANCOVA was used instead (with the baseline test as a covariate, the post-training test as the dependent variable, and group as the independent variable; noting that violation of the normality assumption also applied to this analysis).

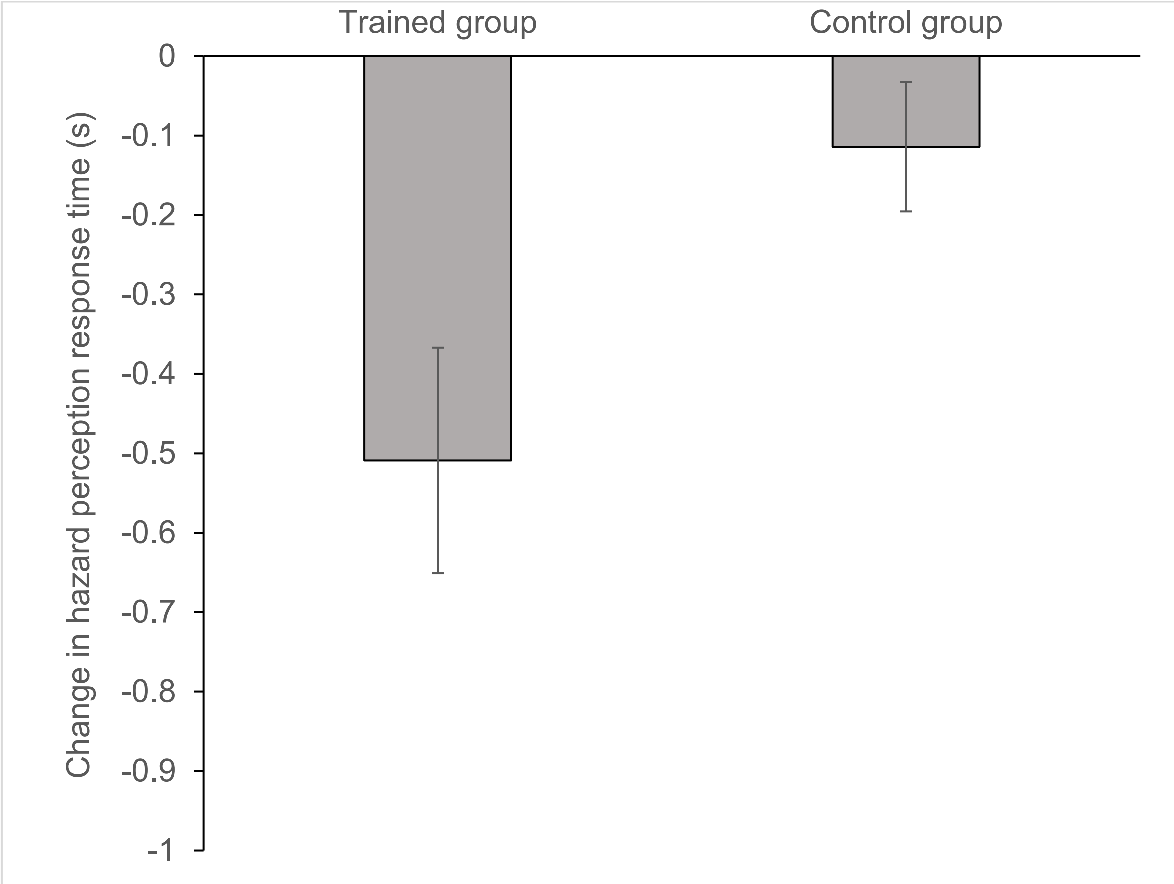


Fig. 3. Change in hazard perception test response times by group (error bars are SE of the mean). A negative value indicates that response times were faster in the second test.

Table 2

Median, means, and SD of hazard perception test scores (in seconds) by group

	Trained group	**Untrained group**
First hazard perception test	*Median* = 6.40	*Median* = 6.58
	M = 6.80	*M* = 6.60
	SD = 1.18	*SD* = 0.65
	Range: 5.56 - 11.30	*Range: 5.34 - 8.62*
Second hazard perception test	*Median* = 6.14	*Median* = 6.39
	M = 6.29	*M* = 6.49
	SD = 0.84	*SD* = 0.68
	Range: 5.06 - 9.82	*Range: 5.55 - 8.66*
Change in hazard perception test score	*Median* = -0.33	*Median* = -0.12
(second test minus first test)	*M* = -0.51	*M* = -0.11
	SD = 0.97	*SD* = 0.55
	Range: -5.13 - 1.30	*Range: -2.17 - 1.15*

Ten participants reported that they did not complete all four commentary drive exercises as intended (though all completed some elements of the training). The training effect remained significant when these individuals were excluded, *Mann-Whitney U* = 622, *p* = .029, *rank biserial correlation* = .28.

4. Discussion

This study found that a brief video-based commentary drive training intervention can improve scores on a validated hazard perception test used for driver licensing. No prior published study has evaluated this type of intervention without participants being supervised during the training. This has important implications for the mass deployment of such interventions, as costs would be significantly reduced if a program administrator were not required, making the training more financially viable. Unsupervised training would also be more convenient for trainees participating in the intervention, as it could be completed at any time without the need for prior arrangements to be made, as well as being more accessible for individuals living in remote and rural settings.

Despite the success of the intervention, it is worth noting that the magnitude of the training effect (Cohen's $d = 0.50$, i.e. half a standard deviation) would be classified as medium according to Cohen (1992). This is smaller than the effects achieved in studies involving supervised sessions with similar interventions. For instance, in the study by Wetton et al. (2013), supervised commentary drive training administered to young novice drivers had an effect on hazard perception test scores that Cohen would consider "large" (specifically, Cohen's $d = 1.09$). This suggests that, while unsupervised training appears to be effective, it is nonetheless likely to be less effective than if the training was supervised. However, it should also be acknowledged that the smaller effect size observed in the present study could equally be attributable to study elements not related to the training – particularly the hazard perception tests, which were also unsupervised. For instance, participants might have been less attentive to the traffic scenes in the tests than that they would have been if they were being monitored (or less than they would be during real driving), leading to additional measurement error. Another issue is that the study relied on self-reports to determine whether people engaged in the training, and also to indicate issues with the tests. This raises the possibility that participants might have under-reported non-compliance.

One limitation of the present study is that we did not evaluate whether the training effects transferred either to real driving performance or to crash involvement, both of which would be valuable avenues for future studies to explore. Another limitation is that we did not have a sufficient sample size (given the medium effect size found) to examine training effects separately for different driver groups (e.g. learners, provisional drivers, and unrestricted licence holders). Even though we have previously found significant training effects across the lifespan for this type of intervention (Horswill, Falconer, et al., 2015; Horswill et al., 2013; Wetton et al., 2013), it could be that removing the supervision component reduces the training effect more for some groups than for others.

A key advantage of the training intervention used in this study is its potential scalability. The technical requirements for the training amounted to no more than the ability to present a series of video clips. (Although additional software was needed to run the hazard perception tests that were required for research purposes, these are not part of the training.) This means any medium that allows video clips to be presented could, in principle, be used as a vehicle for the training. For example, the videos could be hosted on a streaming service, such as YouTube or Vimeo, and potentially further disseminated through social media platforms, such as Facebook.

If the training intervention were to be made widely available, one issue that would need to be addressed is how to persuade drivers to complete the exercises in the absence of the course credit incentive used in the current study. One option might be to leverage the finding that the training improved responses to stimuli taken from a formal driver licensing test. That is, drivers required to take such a test to obtain an unrestricted licence could be made aware that the intervention is known to improve test scores and hence is likely to increase their chances of passing. Other options for increasing the uptake of the training might include capitalising on social influence (e.g. engaging the parent supervisors of learner drivers to persuade their children to complete the exercises), offering financial rewards (e.g. tying the training to reductions in car insurance premiums), or linking the intervention to other driver training that is already being undertaken (e.g. when a driver takes traditional driving lessons, their instructor encourages them to complete the online training exercises as well). If effective strategies for engagement are implemented, then the current results indicate that the type of scalable intervention evaluated in this study could represent a cost-effective means of improving the road safety of drivers.

Acknowledgements

We thank Sharlene Makin and Glen Jocumsen of Roadcraft Driver Education (roadcraft.org.au) for their invaluable help. The software for the study was created by Paul Jackson. We thank Alana Hawkins for her help in recruiting some of the participants. We also acknowledge Mark Wetton for help with filming and curating the video clips used, and Francine Smith for constructing the alternate forms of the hazard perception test. We thank Queensland Government for funding the production of the commentary drives and permitting use of the licensure hazard perception test clips. We thank John Burrill and his team of driving examiners who provided the expert commentaries.

Funding

This research was funded by the Australian Research Council Linkage Project scheme (grant LP140100409).

Declarations of interest

None.

References

Ābele, L., Haustein, S., Martinussen, L. M., & Møller, M. (2019). Improving drivers' hazard perception in pedestrian-related situations based on a short simulator-based intervention. *Transportation Research Part F: Traffic Psychology and Behaviour, 62*, 1-10. doi:10.1016/j.trf.2018.12.013

Cantwell, S., Isler, R., & Starkey, N. (2013). *The effects of road commentary training on novice drivers' visual search behaviour: A preliminary investigation.* Paper presented at the Australasian Road Safety Research Policing Education Conference, Brisbane, Australia.

Castro, C., Ventsislavova, P., Pena-Suarez, E., Gugliotta, A., Garcia-Fernandez, P., Eisman, E., & Crundall, D. (2016). Proactive Listening to a Training Commentary improves hazard prediction. *Safety Science, 82*, 144-154. doi:10.1016/j.ssci.2015.09.018

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159.

Crundall, D., Andrews, B., van Loon, E., & Chapman, P. (2010). Commentary training improves responsiveness to hazards in a driving simulator. *Accident Analysis and Prevention, 42*(6), 2117-2124. doi:10.1016/j.aap.2010.07.001

Horswill, M. S. (2016a). Hazard Perception in Driving. *Current Directions in Psychological Science, 25*(6), 425-430. doi:10.1177/0963721416663186

Horswill, M. S. (2016b). Improving fitness to drive: the case for hazard perception training. *Australian Psychologist, 51*, 173-181. doi:10.1111/ap.12132

Horswill, M. S., Falconer, E. K., Pachana, N. A., Wetton, M., & 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, M. S., Hill, A., & Wetton, M. (2015). Can a video-based hazard perception test used for driver licensing predict crash involvement? *Accident Analysis & Prevention, 82*, 213-219. doi:10.1016/j.aap.2015.05.019

Horswill, M. S., Taylor, K., Newnam, S., Wetton, M., & Hill, A. (2013). Even highly experienced drivers benefit from a brief hazard perception training intervention. *Accident Analysis & Prevention, 52*, 100-110. doi:10.1016/j.aap.2012.12.014

Isler, R. B., Starkey, N. J., & Sheppard, P. (2011). Effects of higher-order driving skill training on young, inexperienced drivers' on-road driving performance. *Accident Analysis and Prevention, 43*(5), 1818-1827. doi:10.1016/j.aap.2011.04.017

McKenna, F. P., Horswill, M. S., & Alexander, J. L. (2006). Does anticipation training affect drivers' risk taking? *Journal of Experimental Psychology: Applied, 12*(1), 1-10. doi:10.1037/1076-898X.12.1.1

Poulsen, A. A., Horswill, M. S., Wetton, M. A., Hill, A., & Lim, S. M. (2010). A brief office-based hazard perception intervention for drivers with ADHD symptoms. *Australian and New Zealand Journal of Psychiatry, 44*(6), 528-534. doi:10.3109/00048671003596048

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

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