The hazard prediction test: a comparison of free-response and multiple-choice formats

Petya Ventsislavova and David Crundall

Nottingham Trent University

Author Note

Petya Ventsislavova

Department of Psychology, School of Social Sciences, Nottingham Trent University, UK

David Crundall

Department of Psychology, School of Social Sciences, Nottingham Trent University, UK

Correspondence concerning this article should be directed to Petya Ventsislavova (petya.petrova@ntu.ac.uk) or David Crundall (david.crundall@ntu.ac.uk).

Abstract

Hazard perception skill is often related to lower crash risk, and the hazard perception test has been widely employed to measure this ability in drivers. An increasingly popular test-variant is the hazard prediction test: driving videos are occluded immediately prior to a hazard and participants are asked to predict how the situation will develop. Early versions of this test asked participants to provide a free-response answer which was subsequently coded. Later versions, however, have used a multiple-choice format where participants are provided with four options presented on screen. While the benefits of a multiple-choice format are obvious in terms of providing immediate feedback without relying on subjective coding, it is unclear whether this change in format affects the discriminative validity of the test. For the current study, a free-response test and a multiple-choice test were created using the same video clips. The free-response test (experiment 1) was found to successfully discriminate between novice and experienced drivers, with the latter predicting more hazards correctly. The answers provided by participants in Experiment 1 were then used to generate the options for a multiple-choice test (experiment 2). This second test was also found to discriminate between novice and experienced drivers, and a comparison between the two tests failed to reveal an advantage for one over the other. Despite this, correlations between prediction accuracy and both years of post-license driving, and annual mileage, were only significant for the multiple-choice test. The results suggest that the multiple-choice format is not only time- and cost-efficient, but is ostensibly as good as the free-response test in discriminating between driver groups.

Keywords: Hazard perception; hazard prediction; multiple-choice format; driving

General introduction

Hazard perception refers to the skill of detecting on-road dangers in sufficient time to avoid a collision. Following 50 years of research in this field, there is now a general consensus that the hazard perception (HP) skill is related to crash risk (Horswill, 2016). The hazard perception test (HPT) has been widely employed to investigate this skill within the field of traffic and transport psychology, and traditionally involves the presentation of video clips from a driver's perspective, filmed from a moving vehicle. Hazards appear during the clips (e.g. a pedestrian may step into the road, a car may emerge from a side street, etc.), and participants are required to press a button as soon as they spot the danger (McKenna & Crick, 1991, Scialfa et al., 2011, Wetton et al., 2011). It has been documented over many studies that novice drivers and crash-involved drivers are slower, and less likely, to detect hazards in these clips than safer or more experienced drivers (e.g. Cheng at al., 2011, Horswill et al., 2010, McKenna and Horswill, 1999, Rosenbloom et al., 2011). Since the first known hazard perception test (Spicer, 1964, cited in Pelz & Krupat, 1974) many different versions have been developed for research purposes. In addition to the traditional speeded, push-button responses required by the typical hazard perception test, some tests have used levers or sliding scales for participants to indicate a level of hazardousness (Watts and Quimby, 1979, Crundall et al., 2003). Other test variants have required drivers to locate the hazard via a mouse click or a touch screen response (Wetton et al., 2010, 2011). An increasingly popular version requires drivers to predict imminent hazards following sudden occlusion of the video clips (often termed the 'What Happens Next?' test, or the hazard prediction test), measuring HP skill in terms of prediction accuracy (Castro et al., 2014; Crundall, 2016; Jackson et al., 2009).

Tests also differ in the medium chosen to present hazards to participants. The majority of research-based tests employ video clips of driving situations, filmed from a camera mounted on a moving vehicle to record the driver's view of the road. Other variants range from the use of computer-generated imagery to create clips (as introduced in the official UK hazard perception test in 2015), to the use of static road images (as used in the official hazard test in the Netherlands). Even those tests that adopt the more traditional video-based approach can differ in the way they present these clips: some are presented across multiple screens (e.g. Shahar et al., 2010), and have mirror information available to participants (e.g. Ventsislavova et al., submitted), while others present a single forward view (Horswill, Hill & Wetton, 2015). Other differences across tests include variations in instructions given to participants (e.g. Farrand & McKenna, 2001), the method of analysis (for example, the contentious issue of dealing with missing values; Parmet et al., 2014), the nature of the hazards (Crundall, 2016) and even the definition of what constitutes a hazard (e.g. Pradhan and Crundall, 2017; Crundall et al., 2012).

Given that the primary aim of any HP test is to discriminate between safe and unsafe drivers (often on the basis of surrogate measures such as previous crash history, or driving experience), then policy makers might be unconcerned about the particular design of a specific test, providing it successfully separates these driver groups. Certainly, at its best, hazard perception research has demonstrated both retrospective and prospective sensitivity to crash-likelihood, and even offers great hope that HP training may produce on-road improvements in driver safety (Chapman, Underwood and Roberts, 2002; Pradhan et al., 2009, Horswill, 2016; Thomas, Rilea, Blomberg, Peck, and Korbelak, 2016). However, the research field is littered with examples of failed attempts to discriminate between crash-involved, inexperienced drivers and their safer, more-experienced counterparts (e.g. Crundall et al., 1999, Lim et al., 2013; Sagberg, and Bjørnskau, 2006; Underwood, Ngai and Underwood, 2013, Yeung and Wong, 2015). Given the huge variety of test designs across research groups it has proved difficult to understand why some tests are successful and others are not.
For this reason, we argue that test design should be developed through empirical research, with each facet compared and analysed to assess whether it contributes to the validity of the test. This process has already begun in some research groups. For instance, Scialfa, Borkenhagen, Lyon, and Deschênes, (2013) correlated response time performance to hazards presented in static images and dynamic clips (though their results were inconclusive in identifying which was the better test).

Other researchers have noticed that their stimuli can produce different effects in participants depending on the nature of the hazards. Zimasa, Jamson and Henson (2017) found that their clips containing vulnerable road users elicited quicker response times than those involving cars. This distinction appears particularly relevant when comparing typical drivers to individuals on the autistic spectrum, with the latter showing reduced sensitivity for such 'social' hazards (Bishop-Johnson, Biasini, and Stavrinos, 2017; Sheppard, Ropar, Underwood, and van Loon, 2010).
The underlying structure of hazards has also been explored. Crundall (2016) compared 'environmental prediction' and 'behavioural prediction' hazards, and found the former to better discriminate between novice and experienced drivers. While 'behavioural prediction' hazards can be predicted by the actions of the soon-to-be hazard (e.g. the erratic driving of the car ahead), 'environment prediction' refers to hazards that appear out of obscuration (e.g. an oncoming car from around a blind bend, a pedestrian from behind a parked truck). In these scenarios, the

environment is the only clue to the possible upcoming hazard (i.e. the blind bend, the parked truck). These types of hazards have been a particular focus of the Risk Awareness and Perceptual Training programme designed by Fisher et al. (Fisher et al., 2010; Pradhan et al., 2009), where drivers are trained to spot occluded objects that may hide hazards. Similar levels of detailed analysis are also being given to the design of HP training programmes in Australia (Horswill, Garth, Hill and Watson, 2017; Wetton, Hill, & Horswill, 2013).

Despite these studies, more research is needed to provide the basic blueprint for a valid hazard perception test. While much of a test's validity is likely to lie with the content (i.e. the particular hazards that form the stimuli), there are many finer points of test design that may provide significant discriminative gains. These may include the method of presentation, the instructions given, and the required responses, to name but a few.

The hazard prediction test

The hazard prediction test differs from the traditional hazard perception test in that it shuns response times in favour of accuracy for predicting what happens next following an occlusion that occurs just as the hazard begins to develop. It is argued that this test format removes many of the potential problems associated with recording response times to hazards (Crundall, 2016). For instance, response-time measures require a scoring window to be defined. If a response is made between the onset and offset of a hazard, then the response is considered to be correct. However, there are no clear guidelines on how to define onsets and offsets, and there is always the possibility that excellent drivers will spot very subtle cues to upcoming hazards, and respond just before the scoring window (which would be counted as a miss). Even if drivers do press within the scoring window, we do not know if they are responding to the actual hazard, or to some other less hazardous aspect of the scene (see Crundall, 2016, for an argument as to why localised hazard responses are not a suitable solution for a lack of accuracy in the traditional test). Finally, 'hazard perception' is confounded by post-perceptual processes, such as criterion bias: expert drivers may delay or refrain from responding to hazards because they believe the unfolding event falls within the boundaries of their driving skill (Pradhan and Crundall, 2017). The hazard prediction test (or 'What happens next?' test) mitigates these confounds by removing reliance on response times, replacing them with the accuracy of drivers to predict what happens next following occlusion of the developing hazard.

A number of studies have demonstrated the ability of the hazard prediction test to successfully discriminate between safer, experienced drivers, and less-safe, inexperienced drivers (Jackson et al., 2009; Crundall 2016; Castro et al., 2014, 2016, Ventsislavova et al, 2016, Gugliotta et al, 2017; Lim et al., 2014). Several of these studies have also developed this occlusion-based methodology through a number of targeted experiments focusing on design elements. For instance, Jackson et al. (2009) demonstrated that an occlusion is necessary to discriminate driver groups, rather than just pausing on the final frame. A freeze-frame provides an unrealistic amount of time for novice drivers to identify clues to the impending hazard, whereas an occlusion ensures that the driver must be looking in the right place at the right time. As safer drivers are more likely to prioritise those areas of the scene that may develop into hazards, the occlusion is therefore more likely to identify the safest drivers (Crundall and Kroll, 2018).

Crundall (2016) addressed a number of methodological questions, including the impact of clip length on predictive accuracy. He found that longer clips resulted in lower prediction accuracy, especially for novice drivers, suggesting that novices suffer a greater vigilance decrement over time. In a separate experiment, Crundall manipulated the occlusion point. The results demonstrated a decline in prediction accuracy as the occlusion point became more temporally distant from the hazard. The novice/experienced driver distinction remained however and did not interact with the occlusion point. Thus it seems that hazards can be extrapolated from relatively early information (in this case, over a second prior to hazard onset), though at a reduced level of accuracy. Participants' confidence ratings in their predictions fall to baseline levels however at the most distal occlusion points.
These initial studies suggest that the hazard prediction test can provide a robust and simpler alternative to the more traditional hazard perception test. However, one of the problems with the version of the test used by many researchers (Castro et al., 2014, 2016; Crundall 2016; Jackson et al., 2009) is that participants give free-response answers which must be hand coded. This introduces the potential for rater error, and renders the test impractical for use on a wide scale. An alternative is to provide the participants with multiple options to choose from following occlusion, instead of inviting a verbal or typed response. This approach simplifies the test further and allows for automatic and unambiguous coding. This variant of the hazard prediction test was first employed by Lim et al (2014), and was followed by Ventsislavova et al. (2016), but there has never been a direct comparison between a multiple-choice test variant and hand-coded free responses. This paper aims to make that comparison, in order to assess the impact of using the (more practical) multiple-choice method, compared to free responses.

To this end, a new hazard prediction test was developed. This new test uses full HD video from four cameras attached to a moving vehicle. These four video streams are then synchronised and combined with a graphic overlay of a car interior, to provide the viewer with access to mirror information. We believe that providing mirror information may evoke more realistic responses to hazards, even when the hazard never appears in the mirrors (Alberti et al., 2014). Novice and experienced drivers were then tested on this new prediction task in order to assess whether it could discriminate between these groups. This first study (Experiment 1) required free responses to the question 'What happens next?', which is more typical of studies in this area. The second study (Experiment 2) presented the same videos, but participants were provided with 4 options to choose from at the end of each clip. These options were developed from the free-response answers provided in the first study. Finally, a matched sub-group of participants from study 1 and 2 were compared in terms of their prediction accuracy with the only difference being the response mode: free-response or selection of an answer from four options. We hypothesised that both test variants would discriminate between drivers to a certain degree, but we made no prediction about which test would be the better discriminator in the final analysis.

Experiment 1: A free-response hazard prediction test

Introduction

With both the improvement of camera technology and our evolving approach to the presentation of hazards to participants, it is inevitable that new tests will be developed and will require validating. The current stimuli differed from those used in Crundall (2016) both in terms of HD quality and the amount of information provided to participants. For instance, the current test includes mirror information that provides a more realistic level of perceptual demand on the driver, and also allows for hazards to approach from behind (see Figure 1).

To validate this new version of the test we recruited experienced and inexperienced drivers to take part. In addition to the basic validation of this test, we took the opportunity to revisit one of the questions raised by Crundall (2016): when should one occlude the clip to maximise discriminability? As noted above, Crundall found that occlusions that were proximal to the hazard resulted in the most accurate predictions. While the most distal occlusions (an average of 1200 ms prior to hazard onset) significantly reduced accuracy, participants were still able to predict the hazard on 57% of trials (with a free response). Surprisingly, the change in occlusion point did not affect the discriminability of the driver groups. One might suppose that experienced drivers would be relatively better than novices at more distal occlusions, as they may make better use of weaker hazard evidence (Pradhan and Crundall, 2017). However, as occlusions become extremely separated from the hazard we might further predict that all drivers would reach the same nadir, in the absence of even subtle cues to the nature of the upcoming hazard. Neither of these eventualities occurred in the study of Crundall (2016), possibly due to the particular occlusion points that were chosen and the nature of the individual hazards in those clips. Given the need to validate the new free-response version of our prediction test on experienced and inexperienced drivers, we also had the opportunity to investigate whether a novel set of stimuli produced different findings when the occlusion points were varied across three levels: proximal (temporally closest to the hazard), intermediate, and distal occlusion points.

Method

Participants

Sixty-one participants took part in this experiment. The sample was divided into experienced and inexperienced drivers. Experienced drivers (N = 30, mean age = 21.5 years) had a minimum of 1 year of driving experience since passing their driving test and inexperienced drivers (N = 31; mean age = 19.5 years) had less than 1 year of post-test driving experience, and included 14 learner drivers. Mean mileages for experienced and inexperienced drivers were 4553.3 miles and 290.9 miles in the year prior to the study, respectively. All participants had normal or corrected-to-normal vision and were University students and staff.

Design

A 2 x 3 between-group factorial design was employed, where the independent variables were the driving experience of participants (experienced vs. inexperienced) and the occlusion points (proximal vs. intermediate vs. distal). Five cells contained 10 participants, with the distal/inexperienced group containing 11 participants. The proximal occlusion point was the closest to the onset of the hazard, and would typically show several frames of the hazard following onset. This meant that participants who were looking in the right place at the right time, would have their predictions confirmed. For instance, if the participant believed the pedestrian ahead may step into the road, a fixation on this pedestrian immediately prior to occlusion (proximal condition only) would confirm this, as the pedestrian would be seen to move a leg in the direction of the road. Intermediate occlusion points were an average of 618 ms earlier than the proximal occlusion points, and would contain evidence of precursors, but without any confirmation of the hazard triggering. In this case, the pedestrian might be seen walking towards the edge of the pavement, perhaps turning her head in the direction of the film car. Finally, the distal occlusion points were an average of 1222 ms earlier than the proximal points. Typically, these would still contain precursors to the hazard, but the evidence of hazardousness would be weak. For instance, the pedestrian may be visible on the pavement, but still be walking parallel to the road with no obvious intention to cross. See Figure 1 for an alternative example of these three occlusion points.

The dependent variable was the percentage of the accuracy with which participants correctly identified the hazards across a total of 15 clips. Following occlusion, participants were presented with a free-text entry box to provide their answer. These typed responses were subsequently coded with one point awarded for a correct answer. Participants were instructed to consider the identity of the hazardous object, it's location within the scene, and how the event would unfold (e.g. "the pedestrian on the left is about to step into the road in front of me"). Where participants failed to report the three suggested items in their answer, but it was still unambiguously correct, they were still awarded the point. For instance, the response "The pedestrian is about to step into the road" could still be awarded the point if this was the only visible pedestrian, even though the response failed to locate the hazardous object as being on the left-side pavement. Two independent evaluators scored each answer. Cohen's kappa was 0.88 for the distal occlusion point, 0.88 for the intermediate occlusion point, and 0.80 for the proximal occlusion point. Where raters disagreed in their initial rating, a final score was agreed via discussion.

Figure 1 *Three frames illustrating the three occlusion points for one clip. While driving along a road, the car in front pulls over. This requires the film car to overtake the stopping vehicle. However, a safe driver should be aware that there is a vehicle visible in the right side mirror that might pose a problem (top panel; distal occlusion). As the clip progresses, the car in the side mirror disappears, as it moves into the blind spot ready to overtake. Thus, the evidence that this vehicle is going to pose a problem increases considerably (middle panel; intermediate*

occlusion). Finally, the overtaking car becomes visible in the forward view confirming that the car from behind is indeed overtaking and poses a hazard for the film car's intended action (bottom panel; proximal occlusion).

Materials

In order to create the hazard prediction test, new video stimuli were recorded on roads in and around Nottingham (UK). Four HD mini cameras were attached to the car. The first camera was attached on the inside of the front window, with the second camera attached on the inside of the rear window to obtain the footage for the rear-view mirror. The two remaining cameras were attached externally to the driver and passenger side windows to capture side-mirror footage. All clips were recorded in the daytime during fine weather conditions. We extracted a selection of hazard perception clips from the footage, each containing a naturally occurring hazard as judged by two traffic and transport psychologists (see Table 1 for a description of the hazards). Only fully developed hazards were selected (i.e. we did not select a clip with a pedestrian who *looked* like she might step into the road; we only selected the clip if she *did* step into the road, and at a close enough distance to be hazardous, as assessed by our judges). The duration of each clip varied between 31s and 64s prior to editing the occlusions.

Following selection, the four video streams were synchronised, with the three rear-facing video streams edited into mirror placeholders contained in a graphic overlay of the inside of a car. The A-pillars and roof were designed to be semi-transparent, allowing the forward view to be seen through these sections of the overlay and the dashboard and mirror placeholders were fully opaque. The reason behind the semi-transparent design is that real-world obscuration by A-pillars is offset somewhat by stereopsis and small head movements. A fully opaque overlay could therefore over-exaggerate the obscuration that a car frame causes.

The clips were edited to occlude (i.e. cut to an immediate black screen) at three separate points: proximal, intermediate and distal from the fully developed hazard. A proximal cut would allow several frames of the actual hazard to be seen (providing the participant was already looking at the location of the imminent hazard), confirming the participant's prediction. Intermediate occlusions and distal occlusions occurred an average of 618 and 1222 ms earlier than the proximal occlusion, respectively.

Clips were displayed on a Lenovo (ThinkPad) computer running E-Prime 2.0 Software (Psychology Software Tools, 2012), with a resolution of 1920x1080 and screen size of 34.5cm x 19.5cm. Participants responded using the keyboards and the mouse.

Procedure

Once participants gave consent, they were seated 60 cm from the screen and asked to answer demographic questions, including age, sex, year of obtaining driving license, driving collisions in the past 24 months, and miles driven in the past 12 months. Participants were then given both verbal and on-screen instructions for the test. They were told that they would watch 15 short video clips from a driver's perspective. They were asked to watch each clip carefully, as at some point the clip would stop and be occluded by a black screen. Following this, they were asked to describe "What happens next?" (i.e. how the driving situation was going to develop), by typing their answer in a free-response box on the screen (with a 150 character limit). They were instructed to include in their answer any source of potential hazard, the location of that source, and how the situation was going to develop. Following each answer, participants were asked to rate how hazardous they thought the predicted event would be for them, using a 7-point Likert scale (where 1 is 'not hazardous at all', and 7 is 'extremely hazardous').

Participants then watched a practice trial where they had the opportunity to familiarise themselves with the task. This practice trial also included feedback: following occlusion and their answer, the clip was replayed without an occlusion allowing participants to see what the hazard actually was. This practice trial was the only time when they received feedback about their performance. Once participants felt comfortable with the instructions and the practice, they began the experiment. In total, the experiment took an average of 20 minutes.

Nº	Video clips	Duration (ms)	Last sketch prior the clip occlusion	Hazard
1	A car ahead overshoots a red traffic signal. As the cross-traffic begins to enter the junction, the reversing light of the car ahead illuminates, and this car then reverses towards you to get out of the junction. The clip occludes following initial illumination of the reversing light.	64000		Reversing car ahead
2	As you are driving, the car indicates and pulls over blocking your path. As you try to overtake the car and continue your path, a car from behind overtakes you (passing from the right mirror, into the blind spot, before emerging into your forward view). The clip occludes at the point where the overtaking car enters the blindspot.	34000		A car from behind overtakes you
3	You are driving along a narrow urban street and the head of a pedestrian is visible above the parked cars on the left. The clip occludes as the pedestrian moves between the parked cars to step into the road.	34000		The pedestrian from the left crosses the street
4	You come to a standstill due to congestion in an urban street with shops and parked vehicles. A pedestrian exits a shop and then steps out from between two parked cars on the left, just as the traffic moves and you begin to accelerate. The clip occludes when the pedestrian turns his head to look at you, while stepping forward.	49000		A pedestrian from the left crosses the street

5	A distracted pedestrian is walking towards the street. The pedestrian crosses the road from the right without looking. The clip occludes as the pedestrian begins to cross.	54000		A distracted pedestrian crosses the street from the right
6	A car behind you is visible in the rear-view mirror and left side mirror. As you approach traffic lights, this car undertakes you and pulls in front of you, forcing you to brake. The clip occludes when the car is just emerging from the blind spot.	43000		The car from the left undertakes the camera car
7	There are pedestrians on both side of a narrow street. A pedestrian with a child's pushchair enters the road from the right without looking. Her entrance is partially obscured by pedestrians standing on the right. The clip occludes when the pushchair is partially visible among the pedestrians.	32000		A pedestrian with a child's pushchair enters the road from the right without looking
8	You are driving along a narrow urban road. A pedestrian approaches from the right with shopping bags. He is temporally obscured by a large plant pot, before he steps out into the road. The clips occludes just as he begins to step out from behind the plant pot.	46000		A pedestrian from the right crosses the street without looking for upcoming cars
9	A bus in a bus lane signals to pull away from a bus stop. Due to parked vehicles ahead in the bus lane, it pulls out into your lane forcing you to stop. The clip occludes following one flash of the indicator from the bus.	31000		The bus pulls out into your lane
10	An oncoming car cuts across your lane to park in a layby on your left. The clip occludes just before the car starts to cross your lane, following one flash of its indicator.	42000		The car in front cuts across your lane
11	A car behind you is visible in the rear-view mirror and right-side mirror. The car from behind overtakes your vehicle on a blind bend. The appearance of an oncoming car in the opposite lane forces the overtaking vehicle to pull back into your lane immediately in front of you. The clip occludes when the overtaking car begins to emerge from your blind spot, and the oncoming car becomes obvious.	34000		The car behind overtakes on a blind bend
12	While travelling at speed along a country road, a blind bend ahead reveals a queue of standing traffic, forcing you to slow and stop. The clip occludes immediately after passing the blind bend, when the brake lights of the cars ahead are partially visible	70000		Queue of standing traffic
13	While your car is slowing due to congestion ahead, a pedestrian looks over his shoulder before deciding to run in front of your vehicle forcing you to slow more abruptly than otherwise required. The clip occludes at the moment when the pedestrian looks directly at you, as he steps forward.	35000		A pedestrian crosses the street from the left
14	Driving through a school area, two pedestrians are visible on the right. You can see that the woman is looking at the car. The clip occludes at the moment she gesticulates with her arm, thinking that you have given way for her and her child to cross.	42000		Pedestrians from the right cross the street after indicating to the camera car
15	You are driving along a narrow road with parked cars on both sides, which are obscuring the view ahead. Suddenly an oncoming van appears ahead. The clip occludes at the moment the white van is partially visible.	46000		A van appears suddenly behind parked cars

Table 1: A description of the *a priori* hazards selected within each clip

Results

All participants were given a score out of 15 clips for the accuracy of their predictions. The scores were based on the interpretation of the typed free responses by two expert raters. Cohen's Kappa for each condition was acceptable (=>.8), with disagreements reconciled through discussion. Participants' scores were converted to percentages and subjected to a 2 x 3 Analysis of Variance (ANOVA; experienced vs. inexperienced drivers x 3 levels of occlusion point). A main effect was found for driving experience with experienced drivers being more accurate when predicting hazards than inexperienced (54.9% vs. 45.6%; $F(1,55) = 7.13$, $MSe = 161.2$, $p < .01$, $\eta_p^2 = .11$). A main effect was also found across all levels of occlusion point, with the distal, intermediate and proximal conditions producing average accuracy rates of 28.6%, 50.7%, and.72.3%, respectively ($F(2,55) = 60.6, MSe = 161.2, p < .001, \eta_p^2 =.69$). Planned repeated contrasts were conducted comparing distal to intermediate, and intermediate to proximal occlusion points. Both were significant (*ps* < .001). Despite a visible trend to suggest that the distal condition is the weakest discriminator of the driver groups (see Figure 2), no significant interaction was found between experienced group and occlusion point ($F(2,55) =.94, MSe = 151.2, p = .39, \eta_p^2 =.03$).

Ratings of hazardousness were also subjected to a 2x3 ANOVA but no differences were found across experience or occlusion points. The average hazard rating of experienced and inexperienced drivers for the distal clips was 3.46 and 3.40, respectively. For the intermediate clips, experienced drivers gave a rating of 3.76 while inexperienced drivers rated them at 3.72. Finally, for the proximal clips, the ratings for experienced and inexperienced drivers were 3.25 and 4.04.

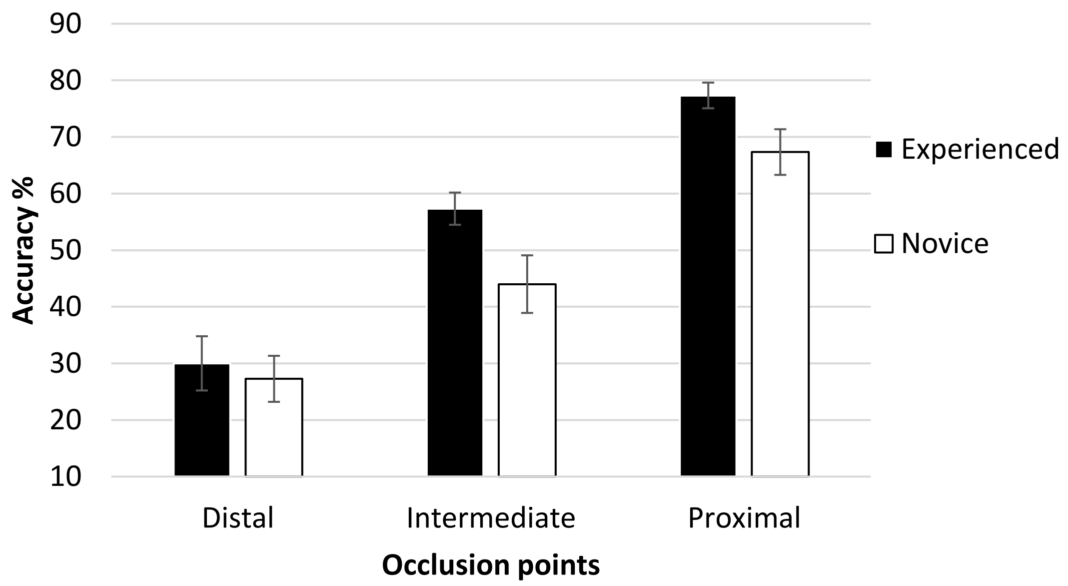

Figure 2: Percentage of correct responses across all 3 occlusion points and experienced groups (with standard error bars added)

Discussion

The primary finding of the current experiment is that this new version of the hazard prediction test, using a new set of clips from those used by Crundall (2016), has successfully discriminated between experienced and inexperienced drivers. With appropriate occlusion points, the clips successfully elicit more accurate responses from experienced drivers compared to inexperienced drivers.
The occlusion-point factor ostensibly produced identical findings to those of Crundall (2016), with declining accuracy as the temporal separation between occlusion and hazard increases. Crundall (2016) failed to find an interaction, and the current interaction between experience and occlusion point also failed to reach significance. However, a closer look at the data is suggestive that the ability of clips to discriminate between the driver groups at the most distal level of occlusion is noticeably degraded. While the data do not support any strong conclusions, there is a clear suggestion that cutting the clip too early reduces the link between precursors and hazards to such an extent that all drivers approach a floor. Note however that the overall accuracy to the distal condition in the current study is far below the 57% score in the comparable condition of Crundall (2016). This suggests that the distal occlusions in the current study are cut more severely in relation to the actual hazards. This cannot be due to absolute differences in distal occlusion points across the Crundall study (2016) and the current one (1200 ms vs. 1222 ms), so must therefore be due to differences in unfolding nature of the hazard precursors. It is likely that the lag between first evidence of a hazard, and the actual hazard beginning to develop, was greater in the Crundall (2016) clips, than in the current selection.

In regards to hazardousness ratings, no differences were noted across either factors, mirroring the results of Jackson et al. (2009), who also failed to find significant differences regarding the hazardousness of the future situation. A possible explanation for this is that drivers are basing their ratings more on the nature of the scene that they *do* see (for more than 30 seconds prior to occlusion), rather than the predicted hazard itself which predominantly remains unseen.

In conclusion, the current study has validated the new set of prediction clips by demonstrating an experiential difference in predictive accuracy. Furthermore, we have demonstrated that the choice of occlusion point has limited influence upon the discriminability of the test, within reasonable boundaries. If set too early however, none of the participants are likely to score highly due to the removal of all precursors to the upcoming hazard.

An additional advantage of the current study however is that the free responses provided by participants offer a range of incorrect predictions that can be developed into multiple-choice distracter answers for Experiment 2. The reduced accuracy of the intermediate and distal conditions was very useful in this regard, as the greater number of incorrect responses was concomitant with a greater variety of response, providing us with sufficient plausible answers to provide three distracter options for each clip in the following study.

Experiment 2: A multiple-choice hazard prediction test

Introduction

There have been several previous attempts to create a multiple-choice version of the hazard prediction test. Lim et al. (2014) used a multiple-choice hazard prediction test with Malaysian and UK drivers. Their results showed that experienced drivers were able to predict more hazards in comparison to the novices (though only for the Malaysian clips). While this result was in contrast to a previously unsuccessful study using response-time hazard perception clips (Lim et al, 2013), they did not directly compare the two datasets across the studies. More recently, Ventsislavova et al. (2016) created a multiple-choice Spanish prediction test in order to test the differences between experienced and novice drivers, asking both “What was the hazard” and “What happens next?”. They found that experienced drivers were more accurate in identifying and predicting the hazard than novices.

Two recent studies have also compared multiple-choice prediction tests directly to response-time tests (Crundall and Kroll, 2018; Malone et al., 2016). Malone et al (2016) described their two test variants as differing on ecological validity, with the response-time test representing high ecological validity, while the multiple-choice version represented low ecological validity. Even though they expected that the high ecological validity test would discriminate better between experienced groups, the multiple-choice version of the hazard perception test also yielded a significant difference between experienced and novice drivers. It should be noted however that Malone et al. used hazards that fully materialised. As participants saw the whole hazard, this test did not isolate the act of prediction, and is therefore less applicable to the current discussion. Crundall and Kroll (2018) compared a response-time HP test and a multiple-choice prediction test across groups of fire appliance drivers (with clips filmed from fire appliances on blue-light training runs). They found the multiple-choice test to be more sensitive to group differences.

Multiple-choice formats have also been employed in the training of hazard perception (Cockerton and Isler, 2003; Isler and Cockerton, 2003, Petzoldt et al., 2013). Petzoldt et al. (2013) compared a paper-based intervention with computer-based training using a multiple-choice format, with the latter being more beneficial for participants to spot earlier critical cues and scan relevant areas in the visual field.

One concern that has been raised regarding the development of multiple-choice tests relates to the plausibility of the distracters (Andrich and Marais, 2014). Poor or implausible distracter options can be easily rejected allowing the participant to choose the correct response by default. For this reason, we followed the guidelines of Worthen et al, (1999), creating the distracter items using the incorrect answers that were provided by the participants in Experiment 1. In order to test this new format, we divided our sample into four groups: learners, novices, moderately experienced drivers and highly experienced drivers. We predicted that there would be clear differences in prediction accuracy across our driving groups.

Method

Participants

Fifty-one participants were split into four different experienced groups: 12 learner drivers (L), 10 novice drivers (N), 15 drivers that have driven for less than 6 years (E<6y) and 14 drivers that have driven more than 6 years (E>6y) (see Table 2 for demographics). All participants had normal or corrected-to-normal vision and were identified as University students and staff.

Participants	Mean Age	Mean driving experience since test (in years)	Mean mileage in the previous 12 months
Learner drivers	18.3	0	9
Novice drivers	18.4	11 months	555
E<6 years	21.4	3 years	4933.3
E>6 years	35.9	16 years	7857.1

Table 2: Mean age, mileage and driving experience for the different driving groups

Design, Materials & Procedure

The proximal occlusion clips from Experiment 1 were edited to include a multiple-choice question, with the most frequent incorrect answers from the previous experiment chosen as the three distracters for each clip. Both the correct answer and all distracter options were written in short, simple sentences, following guidelines provided by Haladyna et al. (1999). All apparatus remained the same as that used in Experiment 1.

A 1 x 4 between-subjects design was employed, where the independent variable was the level of experience of the participants and the dependent variable was the percentage accuracy of hazards predicted. The procedure was identical to that used in Experiment 1 with the only difference being that this time participants were asked to predict the driving situation by choosing one of 4 options after each of 15 clips (See Figure 3). Participants were aware that only one option would be correct and, if they had been looking in the correct place at the time of occlusion, that they would have seen very brief confirmation that their prediction was correct. Following their selection of an option, participants were asked to rate how hazardous the predicted situation would have been for them, by selecting a point on a 1 to 7 Likert Scale (where 1 is 'not at all hazardous' and 7 is 'extremely hazardous'). Prior to the actual experiment, all participants saw a practice trial. The experiment took an average of 20 minutes to complete.

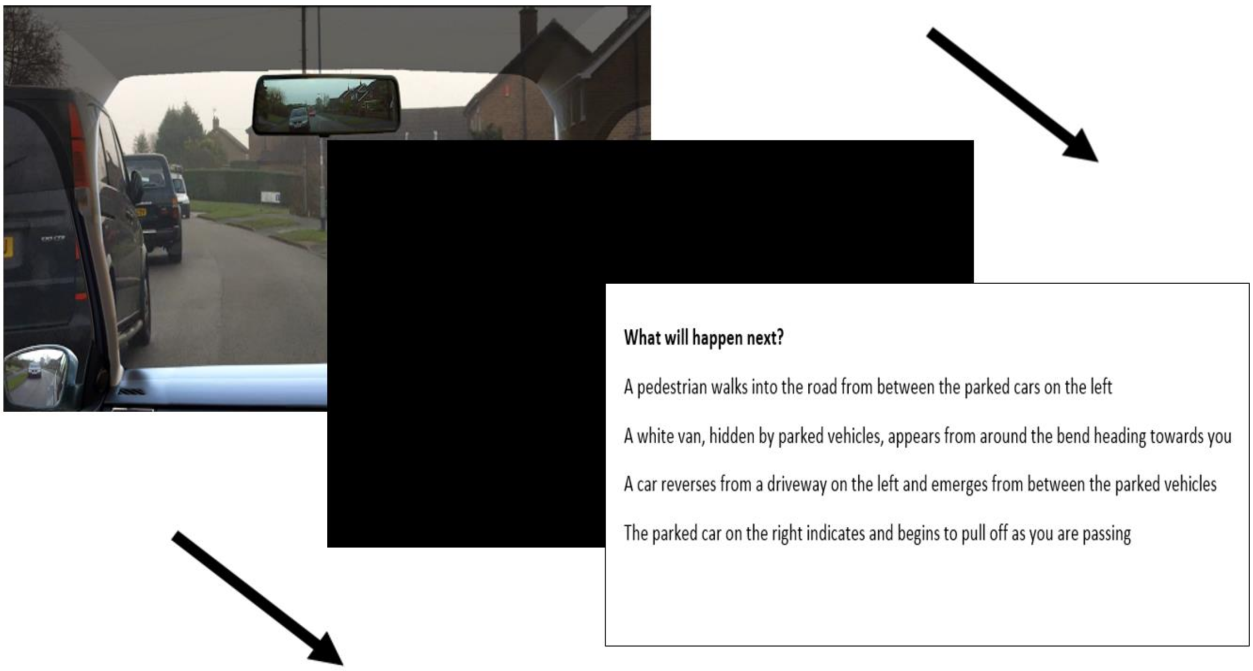

Figure 3: Three panels represent the end of a clip: the first panel is the final frame (where the impending hazard is an oncoming white van, hidden by parked vehicles). This is followed by an occlusion panel, before four multiple-choice options are displayed on screen.

Results

Accuracy was analysed using a 1 x 4 ANOVA, where the between-group factor was driver experience. A main effect was found (See Figure 4), with performance improving in line with increasing experience across the four groups (71.7%, 79.3%, 88.9% and 89.5% for learners, novices, moderately experienced and highly experienced drivers, respectively; $F(3,47) = 6.9$, $MSe = 134.2$, $p < .001$, $\eta_p^2 = .31$). Planned contrasts were undertaken to unpack this main effect. The first contrast pooled the learners and novices and compared them to the pooled experienced driver groups (mirroring the analysis in experiment 1 which just compared inexperienced drivers to experienced drivers). The difference was found to be significant with the experienced drivers out-performing the learner/novice group ($t(47) = 4.2$, $p < 0.001$). Repeated contrasts (learners vs. novices, novices vs. moderately-experienced drivers, and moderately-experienced vs. highly-experienced drivers) were also conducted to identify whether the test was sensitive to more subtle changes in experience, but the only significant comparison was found between novices and moderately-experienced drivers ($t(47) = 2.0$, $p < 0.05$). This suggests a clear experiential boundary in the ability of the test to discriminate between groups at the transition between novice and moderately-experienced drivers.

Ratings were also analysed and a main effect was found for driving experience across the four groups ($F(3,47) = 2.9$, $MSe = .85$, $p < .05$, $\eta_p^2 = .16$). While learners gave the highest risk ratings for the clips (4.5) and the highly-experienced group gave the lowest ratings (3.7), the experiential effect was not sufficient for any of the planned contrasts (L/N vs. all experienced drivers; L vs. N; N vs. E<6yr; E<6yr vs. E>6yrs) to reach significance (See Figure 4).

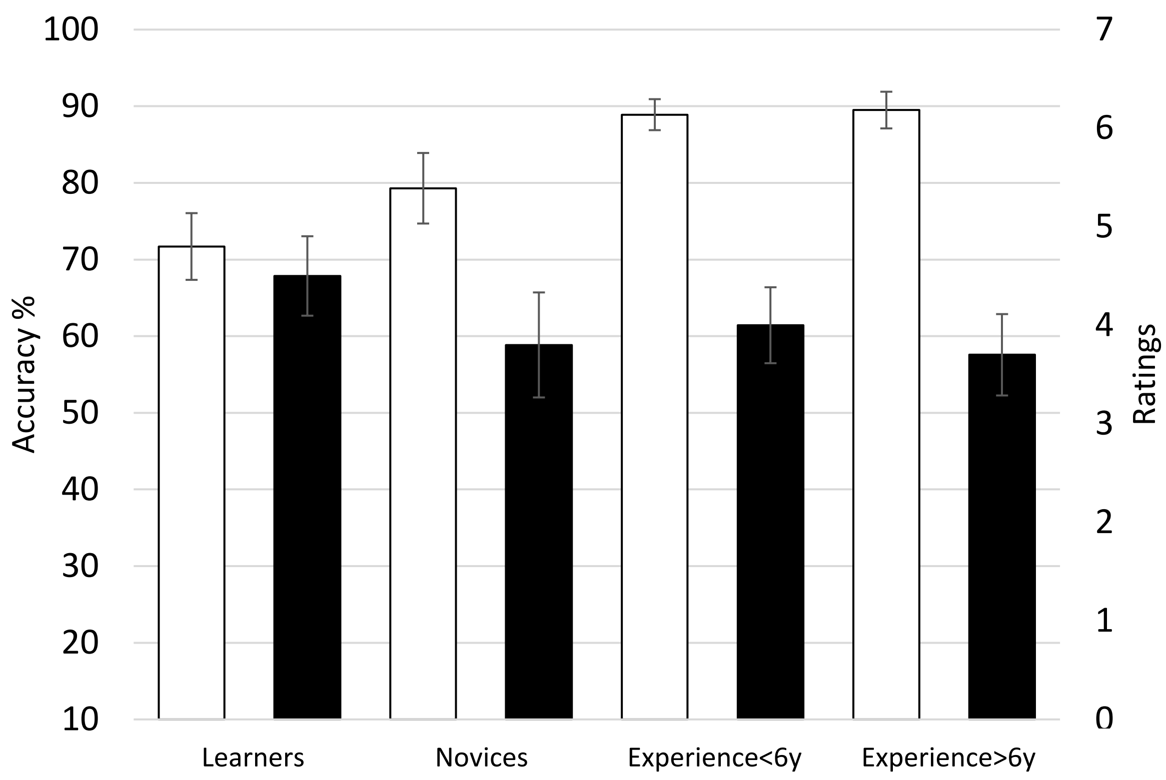

Figure 4: Percentage accuracy (white bars) and ratings (from 1-7; black bars) across all experienced groups: learners, novices, experienced drivers with less than 6 years of driving experience, and experienced drivers with more than 6 years of driving experience. Standard error bars are added.

Comparison of the two tests

A third additional analysis was conducted in order to directly compare the two tests and investigate whether there were differences between the test types in terms of accuracy and discrimination. For that purpose, all the participants that took part in the proximal occlusion condition of Experiment 1 (N=20; 10 novices and 10 experienced) were matched to participants with the most similar driving experience (in terms of passing the driving test and annual mileage) from Experiment 2 (N=20; 10 novices and 10 experienced). The average difference between matched pairs in terms of mileage was 76 miles for the novices and 400 miles for the experienced drivers, while the average difference in years' experience since passing the driving test was of 0.4 years for the novice drivers and 0.2 for the experienced.

A 2x2 between-groups ANOVA was conducted on the two data sets. A marginal difference was found between the tests ($F(1,36) = 3.5$, $MSe = 152.8$, $p=.07$, $\eta_p^2 = .09$), with performance on the multiple-choice test being ostensibly higher than that on the free-response version. The main effect of driving experience remained with experienced drivers outperforming novices (82.3 vs. 69.7; $F(1,36) = 10.5$, $MSe = 152.8$, $p<.01$, $\eta_p^2 = .23$), although there was no interaction between the test type and the experienced groups (See Figure 5). No differences were found for the ratings. The hazards were not rated as more hazardous in terms of test type. These results suggest that the tests do not differ regarding their ability to discriminate on the basis of driving experience, and that the type of question does not influence perceived hazardousness.

Pearson correlations were conducted to assess the relationship between the accuracy of responses, annual mileage and years of driving (post-license) for both types of test. Significant correlations were found for accuracy and years of driving ($r(18)=0.51$, $p <.05$), and accuracy and miles driven ($r(18)=0.56$, $p <.05$), but only with the data from the multiple-choice test. Neither of these correlations was significant for the free-response data from Experiment 1 (See Table 3).

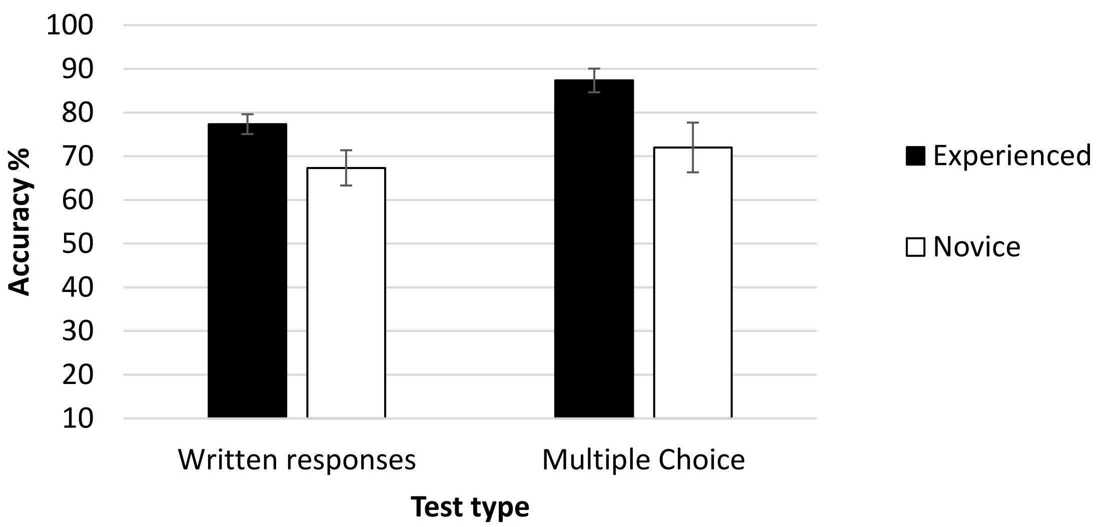

Figure 5: Percentage accuracy of correct responses across experience groups and test type (with standard error bars added).

	Multiple-Choice test			Free-Response test		
	Accuracy	Years	Miles	Accuracy	Years	Miles
Accuracy	–			–		
Years	.512 ($p < .05$)	–		.382 ($p = .096$)	–	
Miles	.562 ($p < .01$)	.697 ($p < .01$)	–	.395 ($p =.085$)	.531 ($p < .05$)	–

Table 3: Pearson correlations for accuracy vs. years of driving and accuracy vs. miles for both types of tests (multiple-choice and free-response type)

Discussion

The multiple-choice hazard prediction test was able to discriminate between the driving groups on the basis of experience. Experienced drivers (E<6y and E>6y) outperformed the learner and novice drivers in predicting hazards. Planned contrasts revealed that the threshold for performance differences lies between the novice and moderately-experienced group, suggestive of a step change in the ability to predict hazards following 1-2 years of post-license driving experience.

This is in keeping with the literature that consistently reports novice drivers, especially within the first 12 months of post-license driving, to be overrepresented in crashes. This suggests that safety-relevant skills are still developing in this first year of independent driving (Foss et al., 2011, McCartt et al., 2003; Williams and Tefft, 2014; Pradham and Crundall, 2017).

When overall accuracy rates for the two tests are examined, the multiple-choice test is ostensibly easier (though this effect did not reach conventional levels of statistical significance; $p = 0.07$). Such an effect can be understood in terms of the opportunity for participants to guess, or infer, the correct answer out of the multiple-choice options. Arguably, an act of inference may still tap into hazard prediction skill (as participants may choose the most likely response on the basis of the evidence they had gathered up to the point of occlusion). Pure guesses are potentially more problematic, but as the ostensible increase in accuracy did not negatively impact on the ability of the test to discriminate between our driver groups, we argue that it is of little consequence.

The ability of the multiple-choice test to discriminate between the driver groups suggests that it might have similar validity to that of the free-response test. Considering the correlation analyses, one might be tempted to argue that the multiple-choice test is even more sensitive to driving experience than the free response test, with years of driving (post-license) and mileage only correlating with the former but not the latter.

On the basis of the current results, it appears that the multiple-choice test is no worse than the free response test at discriminating between our driver groups. Given the positive correlations with measures of experience, and the pragmatic advantages that a multiple-choice format offers (objective marking, immediate feedback, quick administration etc.), we argue that providing participants with four options to choose between should be the preferred format.

General discussion

The current study compared two different versions of the hazard prediction test: free response and multiple-choice formats. The results suggest that both versions are able to discriminate between experienced and novice drivers, however significant relationships between both driving experience and mileage with prediction accuracy were only found for the multiple-choice format.

While several studies have reported finding significant differences in accuracy between novice and experienced drivers when using multiple-choice formats (Lim et al., 2014, Ventsislavova et al., 2016), others have failed to find these differences (Malone and Brünken, 2013). However, when Malone and Brünken (2016) compared the traditional reaction time paradigm with the multiple-choice format, they found that experienced drivers outperformed novices in both test types. Following these results, they concluded that the ability of a test to discriminate between driver groups is more likely to be due to the scenario type (i.e. the particular videos used) rather than the assessment methodology employed.

On the basis of the current results, and those of other researchers reported in the literature, it appears that the multiple-choice format is no less discriminative than other test formats, yet pragmatically it offers a range of other advantages. For instance, the multiple-choice format allows quick administration and objective scoring of the answers. In contrast, a free response prediction test can take much longer to administer, and, without the development of natural language processing algorithms, must be scored offline by hand. Not only is this time consuming and risks errors during subjective coding, but it also means that participants cannot receive an immediate score for their performance. For research purposes, participants do not need to be given a score at the end of the test (though they typically prefer this). However, if the prediction format was to be used in driver training or licensing procedures, then the immediate feedback that can be provided with a multiple-choice format is a necessity.

There are other potential concerns however regarding the use of multiple-choice formats**.** Malone and Brünken (2016) argued that providing multiple options for drivers to choose between might simply tap into memory processes rather than driving skill. This was particularly pertinent for their study which only presented the options after a hazard clip had been viewed in its entirety. In this case, drivers might have to choose an option that refers to a hazard that they had seen tens of seconds ago. While such a short temporal gap might not be considered to place a huge strain on memory, the point remains valid: any gap between the target event and the probe question might produce results confounded by memory processes. However, in the current study, the clip occludes just as the hazard begins to develop, and the options are immediately provided. The safest drivers will be aware of this hazard at the point of occlusion, and thus this answer should be readily available in working memory.

One further concern with the multiple-choice format is the generation of incorrect options that provide plausible alternatives to the correct answer. If the distracter options are too unlikely, then the viewer may be able to guess the correct answer by rejecting the implausible options. For this reason, our distracters were developed using the responses given by participants in Experiment 1. We recommend this as a method for generating plausible and realistic distracters. While this process is time consuming, it helps ensure that the correct answer does not stand out.

In addition to supporting the development of a multiple-choice format, the current studies have provided additional insight into the act of hazard prediction. First, the fact that we have two further experiments that demonstrate the effectiveness of the hazard prediction test serves to bolster the standing of this methodology, adding to the growing literature that has successfully employed the occlusion format (Castro et al., 2014; Crundall, 2016; Jackson et al., 2009; Lehtonen et al., 2017; Lim et al., 2014; Ventsislavova et al., 2016). This is especially important given the more equivocal nature of the findings in the field when using more traditional response-time hazard perception tests (Wetton et al., 2010; Horswill, Hill & Wetton, 2015).

Specifically, the results from Experiment 1 again demonstrate the robustness of the discriminative validity across varying occlusion points (see also Crundall, 2016). Admittedly, the gap between experienced and novice driver performance appears at its weakest with the distal occlusion, though this is understandable: if the clip is cut before there is any evidence to point towards the impending hazard, then neither group should be able to provide a correct answer. Conversely, if the clip is cut *after* the hazard has fully appeared, one might expect all driver groups to get the answer correct (though, interestingly, Malone and Brünken, 2016, still found experiential differences using probes that were presented after the hazard had fully developed).

Experiment 2 also provided new insight via the comparison of four experiential groups. The experiential differences lay between novices and the moderately-experienced driver group. The fact that novices were not significantly better than learners, could suggest that hazard prediction is a skill that is still developing during the first year of independent driving. This lack of improvement in novice drivers over learners is surprising given that all of our novices trained for, and passed, the official UK hazard perception test within (at most) 12 months prior to taking part in the study. This raises the possibility that the introduction of the UK hazard perception test in 2002 into the licensing procedure has not reached its full potential for improving the hazard detection skills of our new drivers. Certainly, the young driver problem has not been eradicated in the UK, with one recent report from the Department of Transport noting that young car drivers are involved in 18% of road collisions, despite only accounting for 5% of the total miles driven (DfT, 2015). It is possible that the focus on *responding* to hazards in the UK test, rather than *predicting* them, may have missed an opportunity to ensure learners are trained in the most important aspects of hazard avoidance (Pradhan and Crundall, 2017).

It should be noted however that any test is only as good as the stimuli that comprise it, and vagaries in stimuli may always prevent wider generalisation of effects. In this particular study, for example, 7 out of 15 clips contained pedestrian hazards. This may be considered to place too much emphasis on this particular type of hazard. In defence of the clip selection, we draw attention to the fact that all stimuli were collected naturalistically, across a variety of roadways (rural, dual carriageway, urban and suburban). Furthermore, other studies have demonstrated the discriminative validity of the hazard prediction paradigm using different clips and different hazards (e.g. Lim et al., 2014, Castro et al., 2014), and the current results are consistent with the emerging pattern in the literature. Nonetheless, the need to assess the hazard prediction test across a wider range of hazards should be addressed in future research, and ideally with larger samples.

In conclusion, the results suggest that a multiple-choice format of the hazard prediction test can provide an efficient and effective tool for discriminating between driver groups on the basis of driving experience. It is particularly sensitive to the boundary between novice and moderately-experienced drivers, which accords with the literature that identifies the first 12 months post-license as being particularly problematic for young drivers. With the growing evidence-base for hazard prediction tests as a valid method of discriminating driver groups, this study has demonstrated how such a test could be designed for widespread automated deployment, thus providing a valid alternative to the current UK hazard perception test that still relies on confounded response-time measures.

References

Alberti, C. F., Shahar, A., & Crundall, D. (2014). Are experienced drivers more likely than novice drivers to benefit from driving simulations with a wide field of view? *Transportation Research Part F: Traffic Psychology and Behaviour, 27*, 124–132.

Andrich, D., & Marais, I. (2014). Person proficiency estimates in the dichotomous Rasch model when random guessing is removed from difficulty estimates of multiple choice items. *Applied Psychological Measurement, 36*, 432-449.

Bishop, H. J., Biasini, F. J. & Stavrinos, D. (2017). Social and Non-social Hazard Response in Drivers with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders, 47*(4), 905-917.

Castro, C., Padilla J.L., Roca J., Benítez I., García-Fernández P., Estévez B., López-Ramón M.F. & Crundall, D. (2014). Development and validation of the Spanish HP Test. *Traffic Injury Prevention*, *15*(8), 817-2.

Castro, C., Ventsislavova, P., Peña-Suárez, E., Gugliotta, A., García-Fernández, P., Eisman, E., & Crundall, D. (2016). Proactive listening to a training commentary improves hazard prediction. *Safety Science, 82*, 144–154.

Chapman, P., Underwood, G. & Roberts, K., (2002). Visual search patterns in trained and untrained novice drivers. *Transportation Research Part F, 5*, 157–167.

Cheng, A.S.K., Ng, T.C.K. & Lee, H.C. (2011). A comparison of the hazard perception ability of accident-involved and accident-free motorcycle riders. *Accident Analysis and Prevention, 43*(4), 1464-1471.

Cockerton, C. & Isler, R.B. (2003). Developing real-life driving simulations for novice driver education. In: Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications, Chesapeake, VA

Crundall, D. (2016). Hazard Prediction discriminates between novice and experienced drivers. *Accident, Analysis and Prevention, 86*, 47-58.

Crundall, D., Chapman, P., Phelps, N. & Underwood, G. (2003). Eye movements and HP in Police Pursuit and Emergency Response Driving. *Journal of Experimental Psychology: Applied, 9*(3), 163-174.

Crundall, D., Crundall, E., Clarke, D. & Shahar, A. (2012). Why do car drivers fail to give way to motorcycles at T-junctions? *Accident Analysis & Prevention, 44*, 88-96.

Crundall, D. & Kroll, V. R. (2018). Prediction and perception of hazards in professional drivers: Does hazard perception skill differ between safe and less-safe fire-appliance drivers? *Accident Analysis and Prevention.*

Crundall, D., Underwood, G., Chapman, P., (1999). Driving experience and the functional field of view*. Perception, 28*, 1075–1087.

DfT (2015). Facts on young car drivers. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/448039/young-car-drivers-2013-data.pdf (last accessed, 21/12/17).

Farrand, P., McKenna, F.P. (2001). Risk perception in novice drivers: the relationship between questionnaire measures and response latency. *Transportation Research Part F: Traffic Psychology & Behaviour 4*(3), 201–212.

Fisher, D. L., Thomas, F. D., Pradhan, A. K., Pollatsek, A., Blomberg, R. D. & Reagan, I. (2010). Development and evaluation of a PC-based attention maintenance training program (Report No. DOT HS 811 252). Washington, DC: National Highway Traffic Safety Administration.

Foss, R. D., Martell, C. A., Goodwin, A. H., & O'Brien, N. P. (2011). Measuring changes in teenage driver crash characteristics during the early months of driving. AAA Foundation for Traffic Safety.

Gugliotta, A., Ventsislavova, P., Garcia-Fernandez, P., Peña-Suarez, E., Eisman, E., Crundall, D. & Castro, C. (2017). Are Situation Awareness and Decision-Making in driving totally conscious processes? Results of a Hazard Prediction task. *Transportation research Part F: Traffic Psychology and Behaviour, 44*, 168-179.

Haladyna, T. M. (1999). Developing and validating multiple choice test items. 2nd Ed.Mahwah, NJ: Lawrence Erlbaum Associates.

Horswill, M. S. (2016). Hazard perception in driving. *Current Directions in Psychological Science, 25*(6) 425-430.

Horswill, M., Anstey, K. J., Wood, J. & Hatherly, C. (2010). The crash involvement of older drivers is associated with their hazard perception latencies. *Journal of the International Neuropsycholocial Society, 16*(5), 939–944.

Horswill, M. S., Gartha, M., Hill, A. & Watson, M. O. (2017). The effect of performance feedback on drivers' hazard perception ability and self-ratings. *Accident Analysis and Prevention, 101*, 135-142.

Horswill, M.S., Hill, A., Wetton, M. (2015). Can a video-based hazard perception test used for driver licensing predict crash involvement? *Accident Analysis and Prevention, 82*(23), 213-219.

Isler, R.B. & Cockerton, C. (2003). A Computer-based interactive multimedia training CD-RO M for novice drivers in New Zealand. In: Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications, Chesapeake, VA

Jackson, L., Chapman, P., Crundall, D., (2009). What happens next? Predicting other road users' behavior as a function of driving experience and processing time. *Ergonomics 52*(2), 154–164.

Lehtonen, E., Sahlberg, H., Rovamo, E., Summala, H. (2017). Learning game for training child bicyclists' situation awareness. *Accident Analysis and Prevention, 105*, 72-83

Lim, P.C., Sheppard, E., Crundall, D. (2013). Cross-cultural effects on drivers' hazard perception. *Transportation Research, Part F, 21*, 194-206.

Lim, P. C., Sheppard, E. & Crundall, D. (2014). A predictive HP paradigm differentiates driving experience cross-culturally. *Transportation Research Part F, 26,* 210-217.

Malone, S. & Brünken, R. (2013). Assessment of driving expertise using multiple choice questions including static vs. animated presentation of driving scenarios. *Accident Analysis and Prevention, 51*, 112-119.

Malone, S. & Brünken, R. (2016). The role of ecological validity in hazard perception assessment. *Transportation research Part F: Traffic Psychology and Behaviour, 40,* 91-103.

McCartt, A. T., Shabanova, V. I., & Leaf, W. A. (2003). Driving experience, crashes and traffic citations of teenage beginning drivers. *Accident Analysis and Prevention, 35*(3), 311-320.

McKenna, F. P., & Crick, J. L. (1991). Hazard perception in drivers: A methodology for testing and training. Crowthorne, UK: Transport and Road Research Laboratory.

McKenna, F. P. & Horswill, M. S. (1999). Hazard perception and its relevance for driver licensing. *Journal of the International Association of Traffic and Safety Sciences, 23*(1), 36–41.

Parmet, Y., and Borowsky, A., (2014). Survival analysis: A fresh approach for analyzing response times in driving-related hazard perception paradigms. *Transportation Research, Part F, 25*, 98-107.

Pelz, D.C. & Krupat, E. (1974). Caution profile y driving record of undergraduate males. *Accident Analysis and Prevention 6*, 45–58.

Petzoldt, T., Weiß, T., Franke, T., Krems, J.F. & Bannert, M., (2013). Can driver education be improved by computer based training of cognitive skills? *Accident Analysis and Prevention, 50*, 1185–1192.

Pradhan, A. K., & Crundall, D. (2017). Hazard avoidance in young novice drivers: Definitions and a framework. In D. L. Fisher et al. (Eds.), Handbook of teen and novice drivers: Research, practice, policy and directions. Boca Raton, FL: CRC Press.

Pradhan, A. K., Pollatsek, A., Knodler, M. & Fisher, D.L (2009). Can younger drivers be trained to scan for information that will reduce their risk in roadway traffic scenarios that are hard to identify as hazardous? *Ergonomics, 52*(6), 657-673.

Rosenbloom, T., Perlman, A., Pereg, A. (2011). Hazard perception of motorcyclists and car drivers. *Accident Analysis and Prevention, 43*(3), 601-604.

Sagberg, F., & Bjørnskau, T. (2006). Hazard perception and driving experience among novice drivers. *Accident Analysis and Prevention, 38*(2), 407–414.

Scialfa, C. T., Borkenhagen, D., Lyon, J. & Deschênes, M. (2013). A comparison of static and dynamic hazard perception tests. *Accident Analysis and Prevention, 51*, 268–273.

Scialfa, C. T., Deschênes, M. C., Ference, J., Boone, J., Horswill, M. S. & Wetton, M. A. (2011). A hazard perception test for novice drivers. *Accident Analysis and Prevention, 43*(1), 204–208.

Shahar, A., Alberti, C.F., Clarke, D., Crundall, D. (2010). Hazard perception as a function of target location and the field of view. *Accident Analysis and Prevention, 42*(6), 1577-1584.

Sheppard, E., Ropar, D., Underwood, G. & van Loon, E. (2010). Brief report: Driving hazard perception in autism. *Journal of Autism and Developmental Disorders, 40*, 504-508.

Spicer, (1964) citado en Pelz, D.C. & Krupat, E. (1974). Caution profile and driving record of undergraduate males. *Accident, Analysis and Prevention, 6,* 45-58.

Thomas, F. D., Rilea, S. L., Blomberg, R. D., Peck, R. C. & Korbelak, K. T. (2016). Evaluation of the safety benefits of the risk awareness and perception training program for novice teen drivers (Report No. DOT HS 812 235). Washington, DC: National Highway Traffic Safety Administration.

Underwood, G., Ngai, A. & Underwood, J. (2013). Driving experience and situation awareness in HP. *Safety Science, 56,* 29-35.

Ventsislavova, P., Gugliotta, A., Peña-Suarez, E., García-Fernandez, P., Eisman, E., Crundall, D. & Castro, C. (2016). What happens when drivers face hazards on the road? *Accident Analysis and Prevention, 91,* 43–54.

Watts, G.R. and Quimby, A.R. (1979). Design and validation of a driving simulator for use in perceptual studies. Report 907. Crowthorne, UK: Transport Research Laboratory.

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.

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.

Wetton, M. A., Horswill, M. S., Hatherly, C., Wood, J. M., Pachana, N. A. & Anstey, K. J. (2010). The development and validation of two complementary measures of drivers' hazard perception ability. *Accident Analysis and Prevention, 42*(4), 1232–1239.

Williams, A. F., & Tefft, B. C. (2014). Characteristics of teens-with-teens fatal crashes in the United States, 2005–2010. *Journal of Safety Research, 48*, 37-42.

Worthen, B. R., White, K. R., Fan, X., & Sudweeks, R. R. (1999). Measurement and assessment in schools. New York: Longman**.**

Yeung, J.S., and Wong, Y.D. (2015). Effects of driver age and experience in abrupt-onset hazards. *Accident Analysis and Prevention, 78*, 110-117.

Zimasa, T., Jamson, S. & Henson, B. (2017). Are happy drivers safer drivers? Evidence from hazard response times and eye tracking data. *Transportation Research Part F: Traffic Psychology and Behaviour, 46*, 14-23.