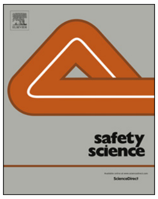
Safety Science 109 (2018) 246–255

Contents lists available at ScienceDirect

Safety Science

journal homepage: www.elsevier.com/locate/safety



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

Petya Ventsislavova*, David Crundall

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

ARTICLE INFO

Keywords:
Hazard perception
Hazard prediction
Multiple-choice format
Driving

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.

1. 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 and 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 et 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 and 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

---

* Corresponding author.
E-mail addresses: petya.petrova@ntu.ac.uk (P. Ventsislavova), david.crundall@ntu.ac.uk (D. Crundall).

https://doi.org/10.1016/j.ssci.2018.06.004
Received 19 January 2018; Received in revised form 5 June 2018; Accepted 11 June 2018

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 et al., 2015). Other differences across tests include variations in instructions given to participants (e.g. Farrand and McKenna, 2001), the method of analysis (for example, the contentious issue of dealing with missing values; Parmet and Borowsky, 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 et al., 2002; Pradhan et al., 2009, Horswill, 2016; Thomas et al., 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 et al., 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 et al. (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 et al. (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 et al., 2017; Sheppard et al., 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 et al., 2017; Wetton et al., 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.

2. 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