PUBLISHED PROJECT REPORT PPR2071

Validation of CGI hazard perception clips

S Helman, J Weekley, I Marino

Report details

Report prepared for:	Driver Vehicle and Standards Agency		
Project/customer reference:			
Copyright:	© TRL Limited		
Report date:			
Report status/version:	Issue 0.1		
Quality approval:			
Project Manager	Navya Keshettivar	**Technical Reviewer**	Jill Weekley

Disclaimer

Contents amendment record

This report has been amended and issued as follows:

Version	Date	Description	Editor	Technical Reviewer
0.1	24/07/25	Draft report for client comment	SH, IM, JW	JW
1.0	15/09/25	Final for publication	JW	JW

Executive Summary

In 2002, a video hazard perception test was introduced into the driver licencing system in the UK. The test requires learner drivers to detect developing hazards on the road ahead in video scenes filmed from a driver's perspective. The reason for the introduction of the test is that a large body of research undertaken by TRL and several UK universities throughout the 1980s and 1990s (see e.g. Maycock, Lockwood & Lester, 1991) had shown that new drivers become less likely to be collision-involved as they accumulate on road driving experience, with hazard perception being identified as one component of driving skill that can be trained and enhanced during this period. It was noted that if levels of hazard perception skill could be increased through requiring the passing of a test before licensure, it had the potential to improve safety for newly licensed drivers. Since this change, the skill of hazard perception as measured by official tests as part of licence acquisition has been shown to be related to safety outcomes (Wells et al., 2008; Boufous, Ivers, Senserrick & Stevenson, 2011).

Grayson and Sexton (2002) summarised the development of the original hazard perception test materials in Great Britain. These authors note four essential characteristics of hazard perception skill that make it suitable as part of the licensing process. These are: it can be measured objectively and reliably; it is related to driving experience (with more experienced drivers scoring higher on the skill because they spot hazards earlier); it can be trained; and it is related to collision involvement (with those having higher skill being less likely to be involved in collisions). It was based on these characteristics that the original test was introduced.

In 2015, the original video-based versions of the clips used in the test were replaced with very high-quality computer-generated imagery (CGI) versions. These modern clips are useful as they update the ‘look and feel’ of clips to appear more modern, and they permit the design of new hazard perception clips from scratch, without needing to engage in real world filming. However, one issue is that the CGI versions of the clips have never been validated against any of the criteria noted in the earlier work (although the original video clips on which the CGI versions had been based had been). Therefore, this study seeks to validate some CGI clips against the criterion of driving experience.

A sample of experienced drivers (those with 10 or more years of license holding) and learner/pre-learner drivers undertook a hazard perception test in which they viewed 16 CGI clips, containing a total of 18 hazards from the official UK testing materials. Hazard perception skill was measured by assigning a higher score for earlier detection of each hazard, based on the proportion of the hazard development remaining when detection was indicated. The experienced drivers scored significantly higher (a mean score of 55.6 out of 100) then was the case for learners/pre-learners (a mean score of 48.3 out of 100).

This work therefore demonstrates for the first time that the CGI clips used in the UK hazard perception test have concurrent validity based on their ability to discriminate between driver groups that differ in their amount of driving experience.

Two clips did not show the expected pattern of results, suggesting that there may be ways in which clips can behave unexpectedly, against the criterion of driving experience. It is

recommended that future clips designed as part of the test should be validated using this method. Suggestions for future work are also made, including how ongoing validation can be built into the licensing system, for example by measuring onward collision risk of samples of candidates and relating this back to their hazard perception test scores. Such work would ensure that the UK hazard perception test remains at the forefront of driver licensing and testing in this important area.

Table of Contents

1 Hazard perception test validation project

1.1 The introduction of hazard perception testing in the UK

The video hazard perception test was introduced into the driving theory test in the UK in November 2002. The test requires theory test candidates to respond to video scenes filmed from a driver's perspective, pressing a button as early as they can when they spot 'developing hazards' in the road scene depicted.

A 'developing hazard' is defined as an event that eventually requires the camera vehicle to take some action such as slowing or steering to avoid a collision. For example, something might move into the path of the camera vehicle; this could include a vehicle emerging from a side road, or some pedestrians stepping into the road to move past a blockage on a footpath.

Such hazards do not happen 'without warning'. Rather, they are preceded by clues that drivers with sufficient levels of skill and experience can use to predict each hazards' future occurrence. The reasoning behind the test is that it measures a skill that is important for safe and competent driving.

1.2 Newly qualified drivers, collision risk, and driving skill

The implementation of the video hazard perception test arose from a large body of research undertaken by TRL and UK universities from the 1980s to the late 1990s. This work had shown that newly qualified drivers were at a much higher risk of being involved in a collision than were more experienced drivers, and that new drivers became less likely to be collision-involved as they accumulated on-road driving experience (Maycock, Lockwood & Lester, 1991). Inexperienced drivers also performed poorly on tests of hazard perception, relative to experienced drivers.

There was therefore a growing understanding that improving hazard perception skill in newly qualified drivers could improve their safety. This is why the hazard perception test was introduced; it provided an objective measure of hazard perception skill and a stimulus for test candidates to engage in training to achieve the level of skill required to pass the test.

1.3 The development of the test since 2002

Grayson and Sexton (2002) summarised the development of the original hazard perception test materials. They noted that the essential characteristics of hazard perception skill that made it suitable as a part of the licensing process were:

1. It can be measured objectively and reliably
2. It is related to driving experience
3. It can be trained
4. It is related to collision involvement

The original test materials (those introduced into the Theory Test in November 2002) met these criteria both in laboratory studies (for a summary see Grayson & Sexton, 2002) and in large real-world samples (Wells, Tong, Sexton, Grayson and Jones, 2008).

In 2015, the Driver and Vehicle Standards Agency (DVSA) began a programme by which the original video-based versions of the hazard perception clips were replaced with very high-quality computer-generated imagery (CGI) versions. This was done for two reasons. First, the original clips were beginning to look dated, and the creation of new versions of the same clips with modern-looking ‘skins’ (newer cars for example) was one way to overcome this. Second, with an ever-changing road environment (for example new vehicle types such as e-bikes) it was desirable to have the ability to produce new clips from scratch, without necessarily needing to undertake a large programme of real-world filming, as was the case with the original clips.

1.4 This project – revisiting validation

1.4.1 The three ways to validate a hazard perception test

Of the four ‘essential characteristics’ of hazard perception listed by Grayson and Sexton (2002) three concern its ‘validity’. A test of any outcome is said to possess validity if it measures what it claims to measure (in this case the ability to detect developing hazards).

There are multiple ways to demonstrate validity; in this example it is demonstrated through the fact that people with more driving experience or specific training perform better on the test, and by the fact that people who perform better on the test have fewer collisions.

When demonstrating the validity of a test, one must trade off the effort and resources required against how well such tests are established and understood already. In the case of hazard perception there are many examples in the academic literature on its use in research; for example, see the systematic reviews of Cao et al. (2022) and Habibzadeh, Yarmohammadian and Sadeghi-Bazargani (2023). However, to the knowledge of the authors, only two real-world implementations of hazard perception tests have been independently evaluated and validated. These are the tests used in Great Britain and the test used in the Australian State of New South Wales. Both these tests have large-scale studies demonstrating the link between test performance and collision involvement (Wells et al., 2008; Boufous, Ivers, Senserrick & Stevenson, 2011). Such validity data can be thought of as the ‘gold standard’ for hazard perception since it relates performance on the test to the outcome or ‘end point’ (Greenhalgh, 1997) of interest to the authorities – in this case road collision involvement.

One problem with using a link to collision involvement to validate hazard perception tests is the need for very large samples of drivers. Even in high-risk groups such as new drivers, collisions are rare. This means that very large samples of participants (typically thousands) and long periods of time (typically at least 6 - 12 months) are required to collect enough data to be able to demonstrate an association. The demonstration of training effects is also a relatively complex undertaking, since it requires the involvement of some training intervention as well as the hazard perception test itself.

For these reasons, in the development of hazard perception tests the first approach taken to validation has typically been to rely on finding expected existing differences between groups. By giving a test to people with almost no driving experience (for example learner drivers) and those with a great deal of driving experience (for example those who have been driving for at least 10 years) we can quickly check that the test is behaving as we would expect, with more experienced drivers having higher performance.

1.4.2 *Validation in the age of widespread hazard perception testing, and CGI clips*

In the UK, the use of inexperienced and experienced drivers as comparison groups in validating hazard perception clips is potentially a different proposition in 2025 than it was in the late 1990s and early 2000s. The reason for this is that, since the introduction of the hazard perception test in 2002, even inexperienced drivers will have almost certainly undergone specific training to improve their hazard perception skill when they take their theory test. Nearly two decades ago, Wells et al. (2008) noted that 96% of the candidates for a UK driving licence used materials (typically multimedia products of some kind) to prepare for their hazard perception test. A simple web search for ‘hazard perception test UK’ at the time of writing (March 2025) provides countless options for hazard perception training products online. Very experienced drivers (those who passed their test before November 2002) on the other hand will not have *needed* to engage with any such training products and are therefore less likely to have done so. This means that inexperienced and experienced drivers are now different in more ways, relevant to hazard perception skill, than was the case before the introduction of the test in 2002. We might term this a ‘saturation of training’ effect.

There is empirical data that hints at such an issue, albeit in motorcyclists. Crundall, van Loon, Stedmon and Crundall (2013) showed that experienced motorcyclists responded *later* (rather than *earlier*) in a hazard perception test than did novice riders, unless they had received advanced riding training; one proposed reason for this was the one noted above.

As noted previously, one obvious advantage to DVSA using CGI in developing hazard perception clips is that it permits rapid prototyping of clips to future-proof the test. A potential issue with CGI clips is that, especially as they become less like the original video clips they replaced, there may be elements of their design that particularly affect how well they reflect real-world ‘developing hazards’. Decisions in the design process for example may result in clues present in the clips being too obvious, or not obvious enough, to permit prediction of the hazard in real time even by someone with the right level of skill. Therefore, this project also sought to answer the question:

1. Can a sample of CGI clips be shown to discriminate between experienced and inexperienced drivers in the ways expected?

This was addressed by looking at differences in hazard perception performance in drivers with differing levels of driving experience (using licence holding as a proxy for this) when responding to 16 CGI clips (containing 18 hazards). In addition, a small number of original video clips (matched to their CGI equivalents) were compared for their ability to discriminate the different groups.

2. Are there any differences in hazard perception score between those experienced drivers who have and have not engaged in hazard perception training?

This was done to check whether in this or future work on hazard perception, there might be a need to account for training experience (and its recency) in understanding group differences in hazard perception skill, in line with Crundall et al. (2013). This was addressed through examining the hazard perception performance of people who have never needed to do any hazard perception training as part of their licensing (those people who had passed their driving test before 2003 and had never, therefore, done a hazard perception test) and comparing this with the performance of people who have (those who passed since 2003).

2 Method

2.1 Participants

2.1.1 Recruitment

Participants were recruited in several ways to obtain a sample of both inexperienced (learner) drivers and experienced drivers (held a full licence for at least 10 years).

For experienced participants, the inclusion criteria were that they should hold a full UK car driving licence and have passed their driving test between 1990 and 2015. Initial recruitment was aimed at those passing between 2003 and 2015 who would have taken a theory test which included the hazard perception element; during recruitment a decision was taken to extend this interval to obtain a larger sample.

TRL's participant database was used for experienced participants only; this database contains the details (demographics, driver licensing status and other details) of around 1,000 people living in the Berkshire and Thames Valley area, who have previously indicated that they are happy to take part in research projects. Those people in the database who matched the inclusion criteria for the study were emailed in March 2025, and invited to sign up for the study using the scheduling tool Calendly. Other experienced drivers were recruited through delivering leaflets to houses surrounding the TRL Head Office in Crowthorne. TRL staff who met the criteria were also allowed to take part, although members of the project team and those who had a direct understanding of the hazard perception test (having worked on other projects related to driver licensing for example) were excluded from participation. Finally for experienced drivers, all these groups were asked to spread the information about the trial throughout their friendship groups, although avoiding the use of public social media. All these participants were tested at the TRL Head Office in Crowthorne.

For learner participants, the inclusion criteria were that they should be over 16 either just about to start learning to drive or recently started learning to drive, but not yet taken the practical driving test. The focus was on recruiting learners as early in the learning to drive process as possible, ideally before taking the theory test, so for some of the recruitment mechanisms this was also specified.

Learner drivers were initially recruited through word of mouth as described above, and through contact with several schools and colleges, and university departments. Schools and colleges in the local area were invited to put up posters advertising the study, and any interested participants could sign up through the Calendly link. Finally, emails were sent by DVSA to recent theory test bookers at several test centres local to TRL's Crowthorne Head Office, and the Smart Mobility Living Lab offices run by TRL in Woolwich, London.

Each participant was given a £30 Amazon voucher for taking part.

2.1.2 *Sample*

188 participants were tested, with 105 of these being in the 'experienced' group for analysis and 73 in the 'learners' group. (10 of the learner group were excluded from the analysis as they had a full driving licence from another country.)

Table 1 shows the gender distribution for the participants by group.

Table 1: Participant gender distribution for each group

Gender	Learners	Experienced
Male	53%	47%
Female	44%	52%
Other	3%	0%
Prefer not to say	0%	1%

Figure 1 shows the age distribution for the participants for each group. As would be expected the learner group is dominated by the 16-24 age group, whereas the experienced group consists mainly of those in the 35-54 age groups. The mean (and standard deviation) for age was 22.9 (12.4) for learners and 42.0 (7.5).

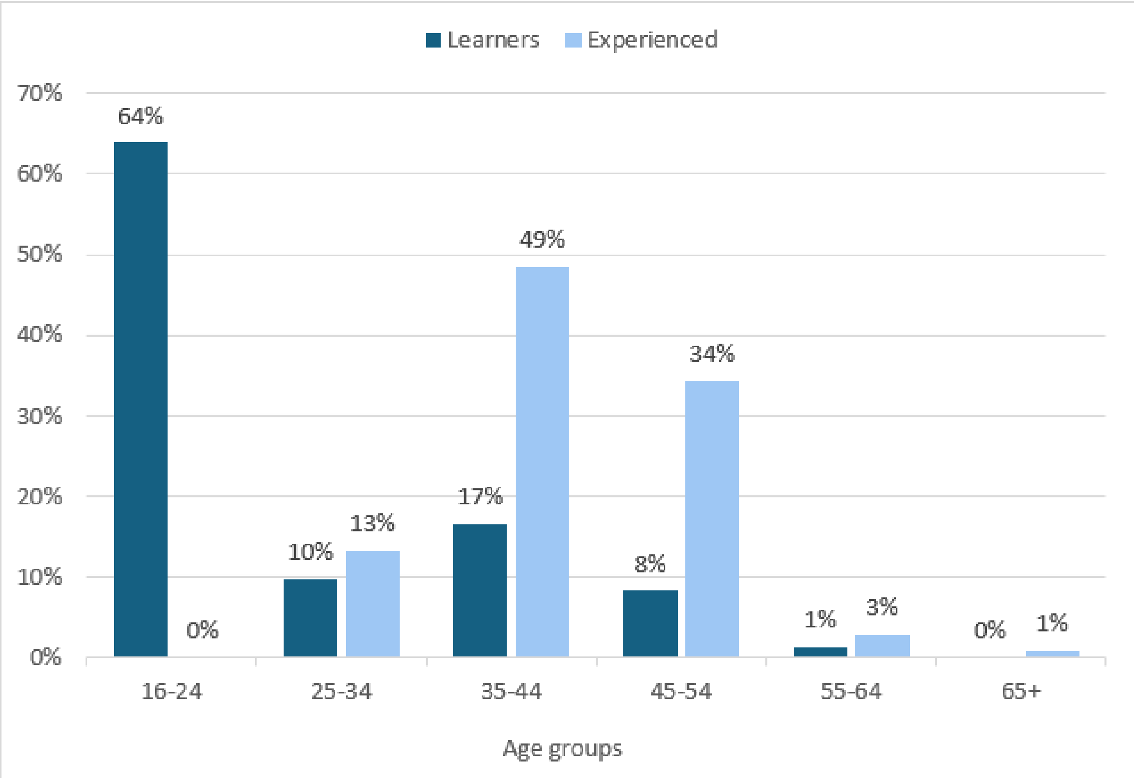


Figure 1: Age distribution of participants for each group

Experienced drivers were asked about the number of years they have held their licence; the mean (and standard deviation) was 21.6 (7.0). Note that this variable was self-reported by participants, no validation of exact passing date was undertaken.

The theory test was introduced in July 1996 and the hazard perception element was introduced in November 2002. Therefore, experienced drivers who have held their licence for more than 23 years are unlikely to have been exposed to hazard perception training materials (and those who have held their licence for more than 29 years will not have taken a theory test at all). Note however that there is usually a long delay between theory test and practical test and hence this differentiation should be seen as indicative only. Figure 2 shows the distribution of experienced participants based on these three groups.

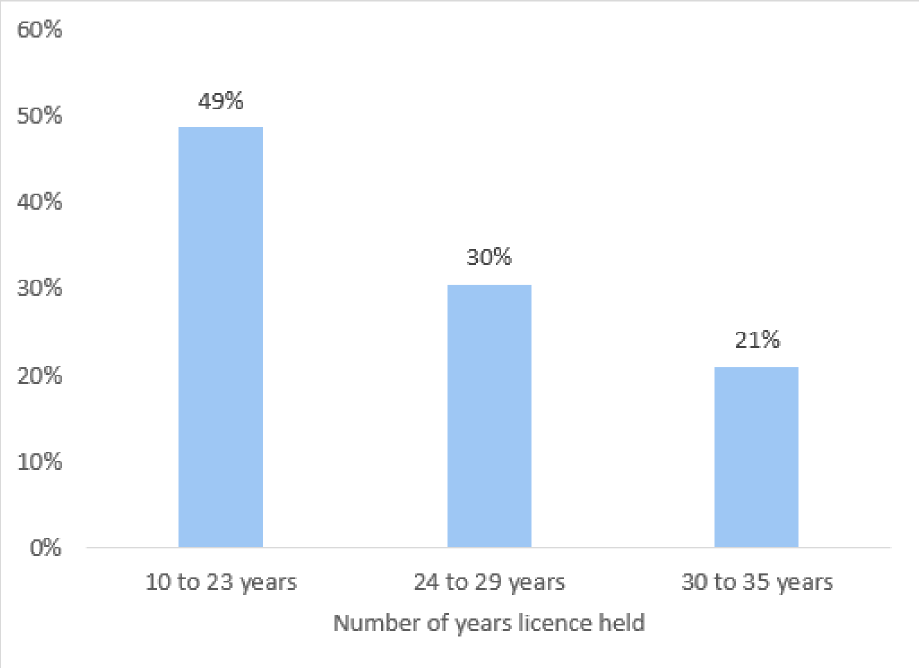


Figure 2: Distribution of experienced participants by number of years they have held their licence

Experienced participants were also asked to estimate their annual mileage, to ensure that they are regular drivers. The distribution of experienced drivers by annual mileage is shown in Table 2.

Table 2: Distribution of experiences participants by estimated annual mileage

Range	Percentage
<1000	3%
1000-9999	57%
10000-19999	32%
20000-29999	5%
>=30000	3%

The learner participants were also asked additional questions regarding their experience, both about their driving experience generally and about their experience with hazard perception preparation and testing.

Figure 3 shows the distribution of learner participants by the level of driving experience that they reported. As intended, nearly all the sample (92%) had limited on-road experience (fewer than 20 hours), and more than half (59%) had no on-road experience at all.

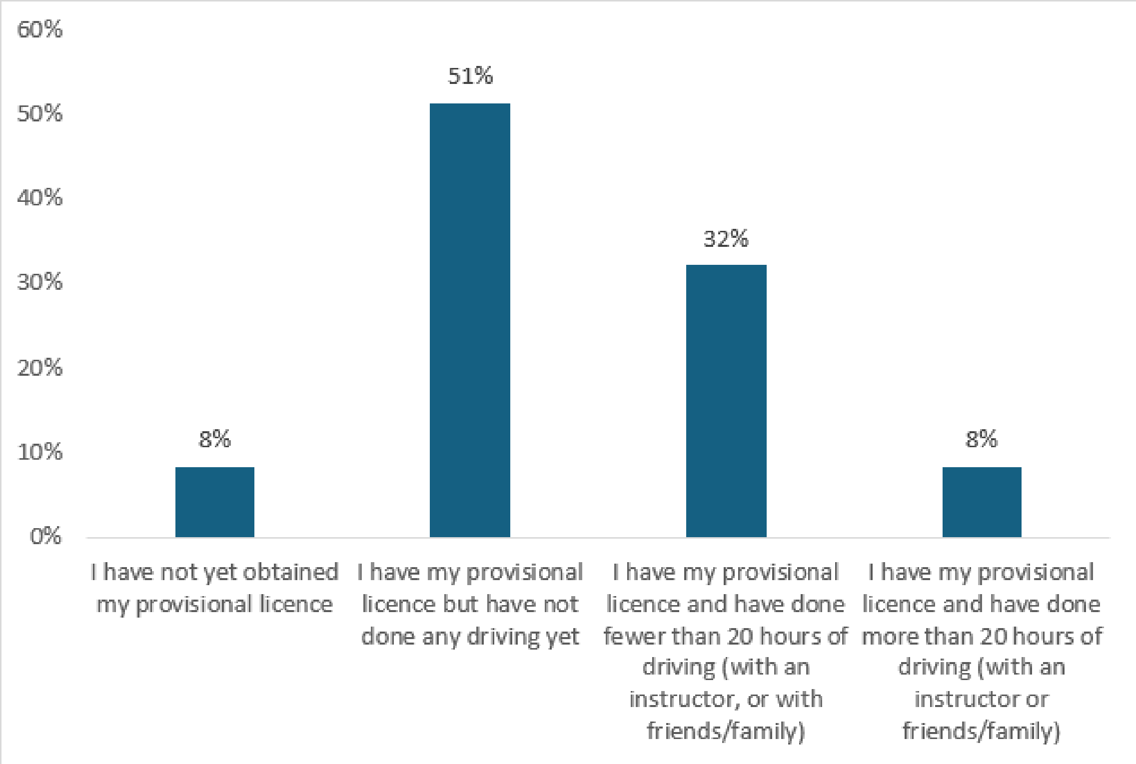


Figure 3: Distribution of learner participants by level of driving experience.

Figure 4 shows the percentage of learner participants by the level of experience they have had with hazard perception preparation or testing. Note that participants were free to select more than one answer, for example ‘I have prepared for my hazard perception test – less than 10 hours’ and ‘I have taken my theory test and failed so far’, and therefore the percentages do not total 100%. Again, as intended, more than half of the learner participants (67%) had done fewer than 10 hours’ preparation for the hazard perception test.

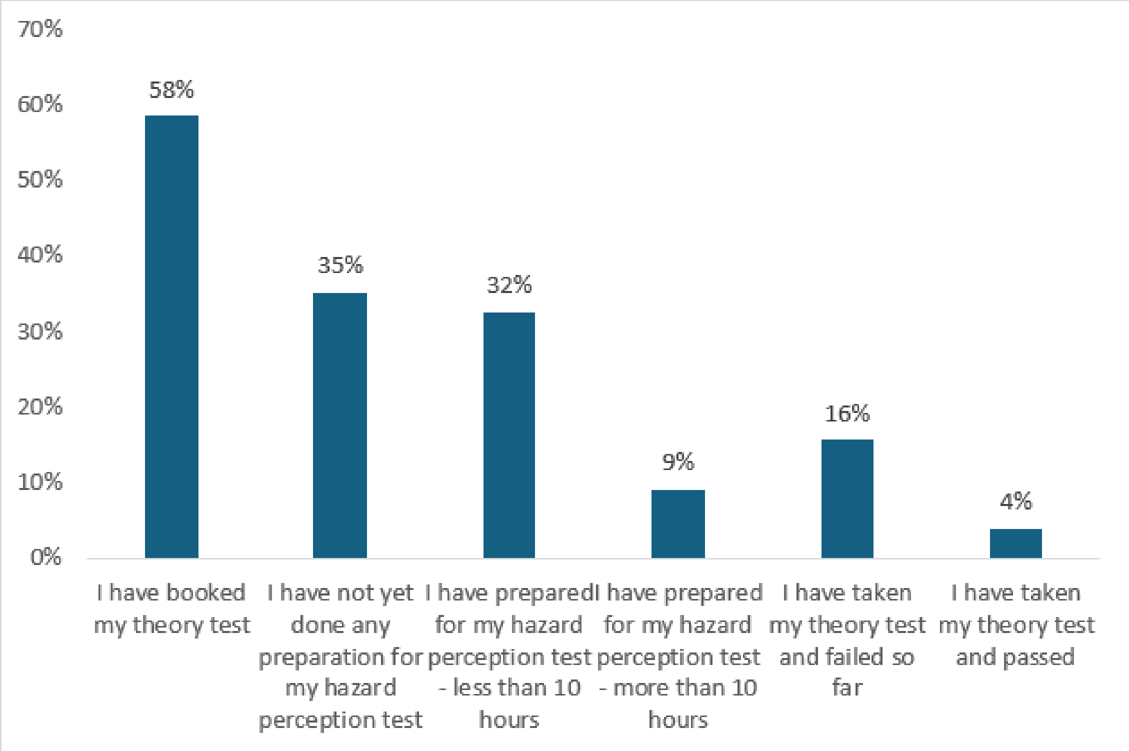


Figure 4: Percentage of learner participants by level of hazard perception preparation or testing experience

2.2 Design

The dependent variable in the study was hazard perception skill. This was measured by asking participants to undertake an experiment as similar as possible to the real hazard perception test delivered as part of the theory test. In the experiment participants were asked to view short video clips from a car-driver's perspective, and respond, via a button press, when they identified a 'developing hazard'. The timing of these responses then produced a score for each clip, as in the real hazard perception test.

Before discussing how the experiment was implemented, it is necessary to understand the way in which hazard perception clips work in relation to their 'scoring windows'.

In the UK hazard perception test, each developing hazard is assigned a 'scoring window' during which button presses are counted as indicating awareness of the hazard in question. Figure 5 shows the very first point and the very last point of this window for an example clip where the hazard is an oncoming emergency vehicle.

Figure 5: Example scoring windows on a hazard perception clip

The reasoning behind such an approach is that there is a point at which a hazard becomes relatively predictable with a sufficient level of skill; in the top image, the blue lights and narrow bridge are just visible. By the frame of the bottom image the hazard is obvious; the emergency vehicle is at the bridge and causing a conflict with the usual priority on the narrow bridge.

The consequence of the scoring window approach is that any button presses before it opens (the top image in this case) or after it closes (the bottom image in this case) would not count towards someone’s score on the test. This approach encourages respondents to focus on spotting hazards as they develop and press the button as early as possible as anticipatory clues becomes available. It also means that drivers cannot simply press the button at the very beginning of a clip (before the hazard elements are even visible) to 'cheat' a high score. Algorithms in the real test also check for patterns of frequent button presses and can disqualify respondents if they do this.

In the real test, candidates score from 1 to 5 points, depending on how early they press the button within the scoring window; the window is divided into 5 smaller windows, with button presses in the first of these scoring highest. In the current study, a slightly different approach was used:

1. We instead operationalised all scoring windows as being of length 100, so it reflected a percentage score
2. We treated the first button press in the window as the respondent's score
3. The score was calculated as being the percentage of the frames remaining in the scoring window at the point of the first button press.

There were 19 clips in all (see section 2.3 for full details), as follows:

1. 16 clips with 18 hazards (two clips had two hazards each, with the other 14 having just one hazard each)
2. Three clips with one hazard each, available as each of two versions – CGI, and original video

The design for the first research question was between-groups. The independent variable (driving experience) had two levels (learner, experienced). The dependent variable was mean hazard perception score on the 16 CGI-only clips (which contained a total of 18 hazards) as described above. A sub-set analysis was also undertaken on the mean hazard perception score on the three clips that had both CGI and video versions; approximately half of each group saw CGI versions of these three clips, and approximately half saw video versions, so these clips were analysed in a between-groups design with again driving experience having two levels, and clip type having two also (CGI, video).

Before analysing the score itself, it is useful to first examine the response rates from the two groups to determine the percentage of participants who successfully identified the hazard within the predefined scoring window. This initial comparison provides context for interpreting the main outcome measure and helps assess whether both groups engaged with the task as expected.

Building on this, the research question – "Can a sample of CGI clips be used to discriminate between experienced and inexperienced drivers in the ways expected?" – is translated into a statistical test comparing group means. Specifically, the test evaluates whether we can reject the null hypothesis that the means of the two groups are equal, in favour of the alternative hypothesis that experienced drivers have a higher mean score. Formally, the test assesses the following:

$$H_0: \text{Mean}(\text{Score}_{\text{Experienced}}) = \text{Mean}(\text{Score}_{\text{Learners}})$$

vs

$$H_1: \text{Mean}(\text{Score}_{\text{Experienced}}) > \text{Mean}(\text{Score}_{\text{Learners}})$$

If the variables are normally distributed, or if the sample size is sufficiently large ($n > 30$), a t-test can be performed to address the research question. When we assume unequal variances between the groups, the Welch's t-test is more appropriate.

Even if the assumptions underlying the parametric test are only mildly violated, it is advisable to perform a non-parametric test in addition, to assess whether the findings are robust across different statistical approaches.

A similar analysis was conducted to explore differences in performance between CGI and video clips. It is important to note that only three clips were used and only approximately half of each group viewed one or the other. This limited stimulus set and sample size should be considered when interpreting the results, as it may constrain the generalisability of any observed differences.

For this analysis a two-way ANOVA was performed to assess whether video type (CGI or Video), driving experience level (Experienced or Learners), and the interaction between these two factors had a statistically significant effect on the scores. This approach allowed for a more nuanced understanding of how these variables independently and jointly influenced performance

The second research question was addressed through a planned analyses on sub-groups in the main sample, again in a between-groups design, using the 16 CGI-only clips. Hazard perception skill in experienced drivers who passed their theory test in 2003 or later (i.e. after the implementation of the hazard perception test) was compared with the skill of those who passed before 2003.

The second research question – “Are there any issues with using pre-existing differences in driving experience as a validation measure of hazard perception clips for car drivers in 2025?” – was therefore formally addressed by comparing the hazard perception performance score of individuals who would have been unlikely to have undergone formal hazard perception training (those who passed their theory test before 2003) with those who likely would have (those who passed since 2003).

In this case, a two-tailed t-test was used to assess whether there is a statistically significant difference in mean scores between the two groups. Unlike the previous analysis, equal variances were assumed, allowing for the use of a standard independent samples t-test. This approach tests the following hypotheses:

$$H_0: \text{Mean}(\text{Score}_{pre2003}) = \text{Mean}(\text{Score}_{post2003})$$

vs

$$H_1: \text{Mean}(\text{Score}_{pre2003}) \neq \text{Mean}(\text{Score}_{post2003})$$

This analysis helps evaluate whether the likely presence or absence of formal hazard perception training is reflected in performance, and whether such differences can be reliably used to validate the clips.

2.3 Apparatus and materials

The study was run using the online experiment generation software Gorilla (https://gorilla.sc/). However, due to the slightly sensitive nature of the materials (see below in procedure) although the stimuli were hosted on Gorilla, the testing was done on laptop computers at specific sites, to control the way in which participants interacted with the stimuli. Dell Latitude laptops (model 5480/5490) running Windows 10 were used, all with Intel Core i5 processors and either 8GB or 16GB RAM installed. The laptops were mains powered during the tests, with brightness set to maximum for ease of viewing and consistency.

The stimuli were 19 CGI clips from the existing bank of materials used in the hazard perception part of the theory test (two of which had two hazards each). Three of these also had the original video versions from which they had been copied. To overcome issues with loading and lag, all clips were compressed using a software tool called HandBrake (https://handbrake.fr/). Compressed videos were compared against their original counterparts with no noticeable loss in video quality. The frame rate of the video clips was set to 23.976 frames per second (fps), which can be perceived smoothly by the human eye. Gorilla is coded to count frames and ensure that stimuli is shown for as accurate a time as possible (Timing in Gorilla) with timestamps recorded in the experiment data for clip start, response(s), and clip end.

Sixteen CGI clips (containing 18 total hazards) were used for the main analyses, and the three CGI/Video pairs (each with one hazard) were used for the analysis looking at how the different formats behaved.

2.4 Procedure

After arriving for the study, participants were first asked to complete a consent form, and for those participants with driving licences these were offered and checked. Participants were then invited into the testing room. Two rooms were used at TRL's Head Office with up to eight people being tested at once. Both rooms had office lighting on the full working setting, and in the case of one room with windows, blinds were set to downward position. Another similar room was used at the SMLL site for some of the learner participants, again with full lighting.

On entering the room, participants were shown the instructions in Appendix A (these are for experienced drivers – the learner ones had 'learner drivers' in place of 'experienced drivers' and minor differences). These instructions outlined what the participant would be expected to do, namely spend around 20 minutes viewing video clips on a laptop computer, responding via a button press to each 'developing hazard' identified. They then signed the consent form, and the experimenter read out the further instructions in Appendix B. This included watching the official DVSA guide to the Hazard Perception Test 2025 (seen here on the DVSA YouTube channel: Hazard perception test 2025: official DVSA guide). This was intended to help participants understand how a 'developing hazard' is defined. These instructions also outlined the differences between the real hazard perception test and the current experiment. Participants were asked to put their phones face down behind the laptop screen during testing, and not to take any photos of any clips. After the further instructions and any clarification questions, participants (either alone or in groups) took the test in silence.

At the end of the test, participants were presented with a small number of survey questions collecting demographic information and information about their licence status, driving experience and hazard perception experience (depending on group). The questions that were asked are included in Appendix C. At the end of the session, participants were reminded not to discuss the specifics of what they did in the session to help make sure any future participants did not arrive for a later session with prior expectations. Participants were then given their £30 voucher and left.

3 Results

3.1 Research question 1: Using CGI clips to discriminate between experienced and inexperienced drivers

3.1.1 Main analysis

The first research question concerned the ability of the CGI clips in discriminating between those drivers with different levels of driving experience. This is one of the classic validation tests for hazard perception clips.

In the current study there were 16 CGI clips used for this analysis. These 16 clips had a total of 18 hazards, as two clips had two hazards each rather than the usual one. Across the 18 hazards, an average of 86% of the experienced drivers responded within the predefined windows, compared with an average of 84% for the learners.

For the analysis, the mean hazard perception score for the 18 hazards was calculated for each participant in each group. Figure 6 shows the mean of these scores for each group – 55.60 for the experienced drivers and 48.30 for the learner drivers. A two-sample t-test showed that the mean hazard perception score was significantly higher in the experienced group than in the learner group ($t(125) = 2.85$, $p = 0.003$). A nonparametric Mann-Whitney U test was also run as the learner group distribution deviated from normality. This test confirmed the difference in hazard perception scores between the groups ($U = 4613.5$, $p = 0.021$).

The findings therefore confirm that the CGI clips used in the current study behaved as expected; experienced drivers scored higher in terms of their hazard perception skill on these clips than was the case for learner drivers.

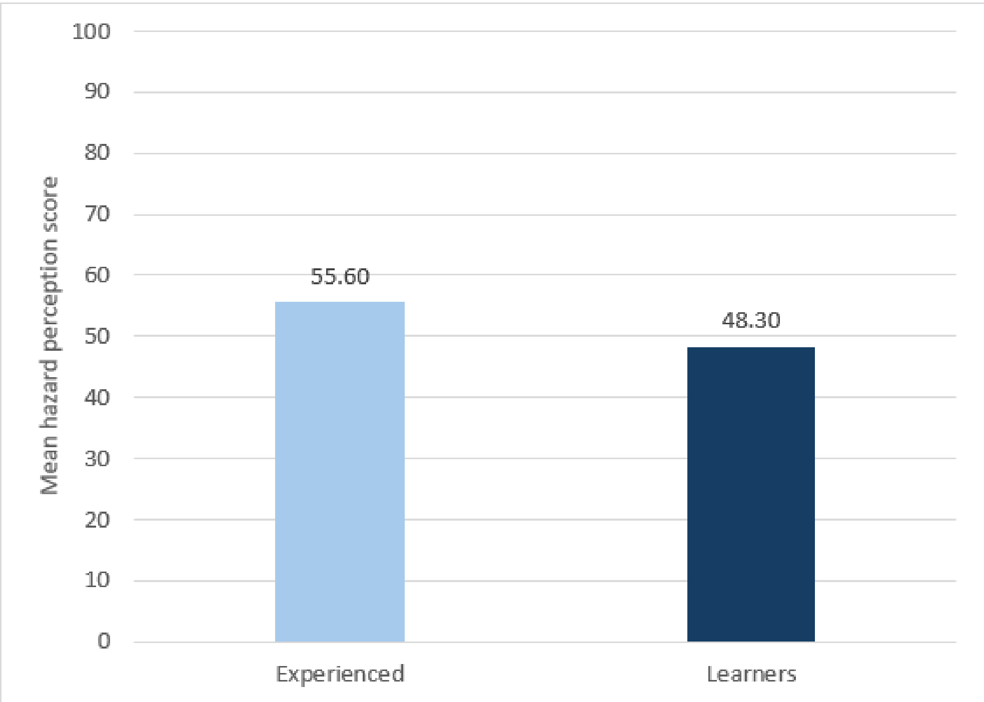


Figure 6: Mean hazard perception scores by experience level

3.1.2 *Scores for individual clips*

The analysis described in section 3.1.1 was undertaken on all 18 hazards in the 16 CGI clips combined. Table 3 shows the mean hazard perception score was higher in the experienced group (as expected) in 16 of the 18 hazards. This shows that while the CGI-clip-based test discriminated as expected between experienced drivers and learners, this was not the case for all clips individually, with two (CG0151_2 and CG0156 – shown with asterisks in the table) showing the opposite pattern.

Table 3: Mean hazard perception score by experienced and learner groups for the 18 individual hazards included in the 16 CGI clips

Hazard	Experienced	Learners
CG0109	**62.52**	54.48
CG0118	**57.05**	54.73
CG0128	**75.46**	64.53
CG0130	**53.81**	47.84
CG0142	**67.15**	56.56
CG0151_1	**62.06**	52.62
CG0151_2*	31.84	**44.57**
CG0156*	64.34	**69.85**
CG0159	**51.22**	42.92
CG0162	**49.74**	46.06
CG0163	**81.86**	77.83
CG0167	**42.09**	31.54
CG0168	**48.40**	37.35
CG0179	**57.19**	42.21
CG0229	**38.19**	27.42
CG0266	**57.88**	42.10
CG0466_1	**53.73**	38.69
CG0466_2	**46.35**	38.17

A detailed description of these two clips cannot be discussed in this report; however the first of the clips shows a scenario with a number of what might be termed unusual things in the scene. One speculation as to why learners score more highly on this hazard is that they have recognised the combination of elements from training materials, as being likely to indicate a hazard. The second clip shows a more traditional event that would be very common in driving situations, so it is not clear why learners are scoring higher (although scores are closer on this hazard than on the first).

3.1.3 *CGI-Video clip comparison*

Another useful question to answer is whether equivalent clips that differ only in their implementation as either video based, or CGI based, behave the same in terms of response characteristics and do both types of clips still discriminate between experienced and learner drivers, when the hazards in the clips are the same?

In the current study there were three CGI clips that had corresponding video versions (that is to say, the video clip from which the CGI clip was created). This made it possible to examine mean hazard perception score for these three clips by clip type as well as by driver type. Figure 7 shows mean hazard perception scores for these clip types separately for the experienced and learner groups.

A two-way analysis of variance was undertaken to test for the main effect of clip type the main effect of group, and the interaction between these two variables. This analysis showed that video clips had significantly higher mean hazard perception scores than CGI clips ($F(1,174) = 11.45$, $p < 0.001$). The main effect of group ($F(1,174) = 2.15$, $p = 0.14$) and the group by clip type interaction ($F(1,174) = 2.64$, $p < 0.11$) were both non-significant.

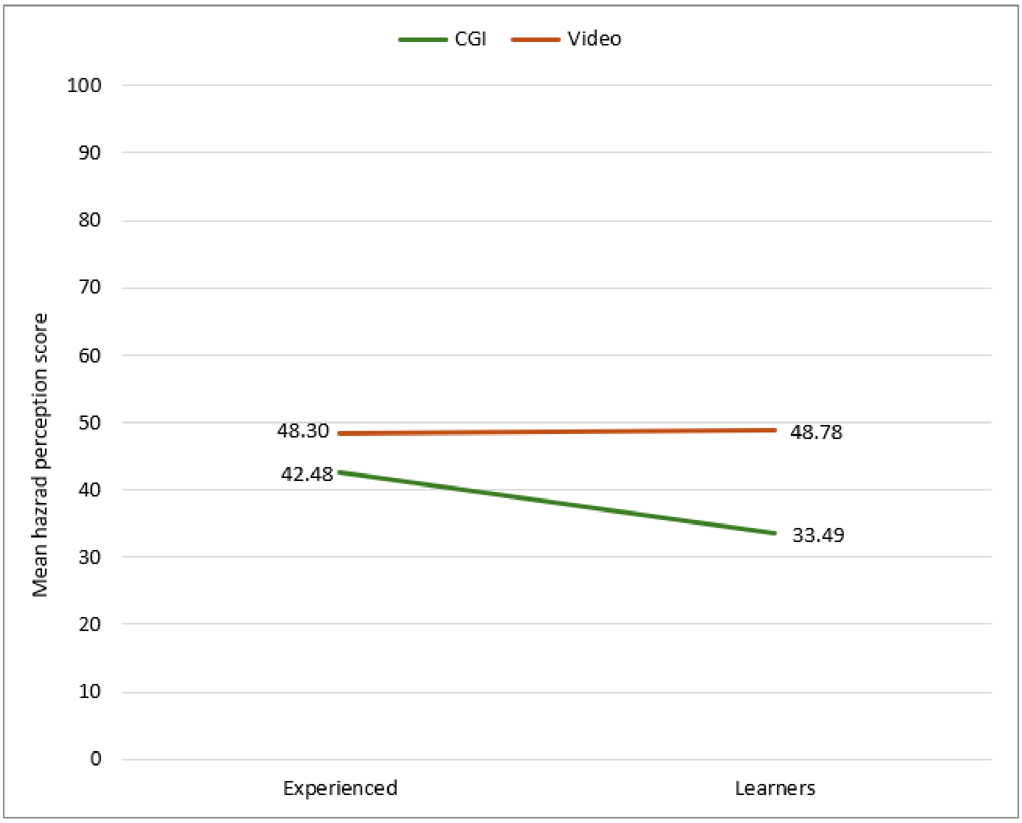


Figure 7: Mean hazard perception scores by experience level and clip type (three CGI clips and three video clips)

It should be noted that the available statistical power in this analysis was low compared with the main analysis reported above. This is because approximately half of each group were assigned to see the CGI versions of these clips, with remaining participants assigned to see the video versions. In addition, there were only three clips that possessed matched CGI and video versions, meaning that the estimate of hazard perception skill will not have been as precise as for the main analysis.

The only thing that can be reasonably reliably concluded from this analysis is that video clips appear to give rise to higher hazard perception scores than CGI clips (on the percentage-based metric used in this study). This would however need to be tested further with a larger sample of clips to be more certain.

3.2 Research question 2: Testing for differences between those with and without hazard perception test experience

The second overall research question concerned the differences between experienced drivers who had varying levels of engagement in hazard perception testing as part of their licence acquisition.

This analysis used just those participants in the experienced group. Mean hazard perception score was compared for those drivers who reported passing the practical driving test from 2003 onwards, and those who passed before this. The video hazard perception test was introduced into the driver licencing process in the UK in November 2002, meaning:

1. Those passing their practical test from 2003 onwards would have likely had some involvement in hazard perception testing as part of their original licencing and would therefore have been likely to have prepared using training materials of some kind
2. Most people passing before 2003 would have had no (or less) such hazard perception training in preparation for their test.

Note that the number of years licence held was estimated by participants to the nearest year and no validation of exact passing date was undertaken; in addition, there is usually a long period between taking the theory test and passing the practical test. Therefore, some participants who reported passing in 2002, 2003 and 2004 (eight participants) may have been assigned to the incorrect group in this analysis. Participants were asked if they remembered taking a hazard perception test – of these eight, five were in line with the assumptions above.

A two-sample t-test failed to find any evidence that the two groups of experienced drivers differed in terms of their hazard perception score ($t(103) = -1.30$, $p = 0.2$). As with the analysis reported in section 3.1.3, the statistical power present in this analysis was much lower than that present in the main analysis.

4 Discussion

The main purpose of this study was to check that a hazard perception test comprising CGI clips from the UK hazard perception test pool was able to demonstrate concurrent validity through discriminating between experienced and inexperienced drivers.

The test included 16 CGI clips, and 18 hazards. In keeping with previous findings with more traditional video clips, experienced drivers (licence held for 10+ years) returned significantly higher hazard perception scores than inexperienced drivers (learner drivers). This is the main finding of the study. The CGI clips on which the UK Hazard Perception test is based demonstrate one of the three validation characteristics identified by Grayson and Section (2002) as being required in a hazard perception test. This means they can discriminate between drivers based on their hazard perception skill – an essential feature of clips included in a test designed to ensure that candidates have a sufficient level of hazard perception skill to progress through the licensing system.

When looking at individual hazards, all but two showed the expected direction of effect (experienced > learner). The two clips that showed the opposite pattern were not obviously different in terms of their characteristics, with the possible exception than one of them had what might be described as unusual interacting elements. Further work may be of use to examine these and other individual clips (with larger samples) to understand whether this pattern of findings is persistent. Any such clips could be considered for replacement.

A secondary analysis of three CGI and video versions of the same hazards could not replicate the experience advantage seen in the main analysis, likely due to the low number of clips available for this analysis. Video versions were scored more highly than CGI clips, however, the low number of clips in this analysis makes it impossible to draw any firm conclusions. Given that CGI clips have been validated by the main analysis, we suggest that video is simply seen as one option to produce valid clips for a hazard perception test (video clips having been validated in many previous studies).

No evidence was found of any ‘saturation of training’ effect in line with that reported by Crundall et al. (2013) as a possible reason for their findings that experienced motorcyclists (without advanced training) were worse at hazard perception than those less experienced. This suggests that driving experience is still the critical factor in the wider driving population that determines hazard perception skill; this means that the ongoing validation of new hazard perception materials can still use driving experience as a grouping variable. However, the way this was tested in the current study had two serious limitations. First, the sample sizes were smaller than hoped. Second, the comparison groups will have differed in ways other than the one being tested (the likelihood of them being involved in specific hazard perception training in preparation for licence acquisition). Importantly the pre-2003 licence holders, while they would have been less likely to have been involved in hazard perception training, had also had their licence for longer (a proxy in this study for experience). Future work interested in such training saturation effects would need to isolate this variable more than was done in this exploratory analysis.

The main conclusion from the study is that the CGI clips used in the UK Hazard Perception Test possess validity in the sense that they discriminate based on driving experience. Future work should establish their validity as it relates to collision involvement, potentially through targeted surveys of candidates to ask about collision history and relate this to scores.

References

Boufous, S., Ivers, R, Senserrick, T & Stevenson, M. (2011). Attempts at the Practical On-Road Driving Test and the Hazard Perception Test and the Risk of Traffic Crashes in Young Drivers. *Traffic Injury Prevention*, 12(5), 475–482. https://doi.org/10.1080/15389588.2011.591856

Cao, S., Samuel, S., Murzello, Y., Ding, W., Zhang, X. & Niu, J. (2022). Hazard Perception in Driving: A Systematic Literature Review. Transportation Research Record, 2676(12), 666–690. https://doi.org/10.1177/03611981221096666

Crundall, D., van Loon, E., Stedmon, A. W., & Crundall, E. (2013). Motorcycling experience and hazard perception. *Accident Analysis & Prevention*, 50, 456–464. https://doi.org/10.1016/j.aap.2012.05.021

Habibzadeh, Y., Yarmohammadian, M. H., & Sadeghi-Bazargani, H. (2023). Driving Hazard Perception Components: A Systematic Review and Meta-Analysis. *Bulletin of Emergency & Trauma*, 11(1), 1–12. https://doi.org/10.30476/beat.2023.95410.1356

Grayson, G. B. & Sexton, B. F. (2002). *The development of hazard perception testing.* TRL Report (TRL558). Crowthorne: Transport Research Laboratory.

Greenhalgh, T. (1997). How to read a paper: Papers that report drug trials. *BMJ*, *315*(7106), 480–483. https://doi.org/10.1136/bmj.315.7106.480

Maycock, G., Lockwood, C. R. & Lester, J. F. (1991). *The accident liability of car drivers*. Research Report (RR315). Crowthorne: Transport Research Laboratory.

Wells, P., Tong, S., Sexton, B., Grayson, G. & Jones, E. (2008). *Cohort II: A Study of Learner and New Drivers Volume 1 – Main Report.* Road Research Report 81. London: Department for Transport.

Acknowledgements

This work could not have happened without the help from the Driver and Vehicle Standards Agency, who provided the clips for testing and helped with recruitment through the emails to candidates. Additional thanks go to the team of staff at TRL who helped with the extensive in-person testing sessions: David Jenkins, Chris Redfern, Arran Deans, Aisling Nuttall, Labib Salam, and Olivia Murtagh-Smith. In addition, Jack Hitchings set up the Gorilla Software and the video adjustments needed to make it work in the way needed. Themistoklis Pallikarides supported with analysis. Nottingham University and Royal Holloway University of London were also extremely helpful in arranging for potential testing sessions on their premises, although these were not, in the event, needed. Finally, thanks should go to the participants who provided data.

Appendix A Information for participants and consent forms

Thank you for your interest in this research.

What is the research about?

TRL has been commissioned to undertake this work by DVSA (Driver Vehicle and Standards Agency). The research is seeking feedback from experienced drivers on the types of materials that are used in driver training and testing for licence acquisition. (We are also seeking feedback from learners, but these will be recruited separately). The feedback will be used to improve future materials

What will happen now?

Firstly, the researcher will introduce the task and show you a short instructional video. You will then spend around 20 minutes viewing video clips from a car-driver's perspective on a laptop computer, responding to these using simple button presses. After the computer task, there will be a very short survey to complete, where you provide some further feedback on the clips, and provide information about your driving experience and demographic information. This will complete the study and you will receive a £30 Amazon voucher to reimburse you for your time and effort in taking part.

We do not expect any adverse effects from taking part in the study. A very small number of people may find that the clips cause some minor discomfort due to visual movement, but if this happens you would not be expected to complete the task and you will still receive your voucher.

Taking part in the research is completely voluntary, and you can choose to withdraw participation at any time.

The personal information you provided for booking and on this consent form will not be linked with the data you provide as part of the computer-based task or follow-up survey. No other personal data will be collected. All personal data will be deleted at the end of the study.

If you have further questions please ask the researcher.

Consent

	Yes or No
1. I confirm that I have read and understood the information above and have had the opportunity to ask questions	
2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving reason	
3. I understand that if I have any concerns after the study I can get in touch with a representative from TRL using the email address trials@trl.co.uk	

Participant name: ______________________________

Signature: ________________________

Date: _________________

Appendix B Further instructions

Note: instructions are in italics with prompts for the experimenter in bold and parentheses.

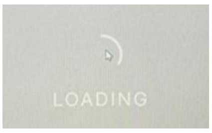
Please select a laptop to sit down in front of and make yourself comfortable. Please do not touch anything until you are instructed to do so. **[Allow this to happen]**

The nearest toilet is through the wooden door at the end of the room, and then through the next door on the right. In the event of a fire alarm sounding please follow me to the nearest fire exit – also point out where muster point is (outside back window of Pegasus)

The trial will last 20 to 30 minutes and will be followed by a short questionnaire.

Please read, complete and sign the consent form that is next to your computer.

Now, on the screen in front of you, you should see a 'Before starting' screen. Please carefully read the instructions on the screen and press the Next button.



[Allow this to happen - it will show for approx 45secs, so while this is happening, prepare to play the video and say the following immediately before playing]

- *Please let the session load, but don't press any further buttons for now.*
- *Today we are going to show you clips from the video hazard perception test and ask you to respond to them as if you were taking the hazard perception test for real.*
- *Before we show you the video clips on your individual computers, we are going to show you an instructional video on the screen that explains how the hazard perception test works, and how you are to respond to the video clips*
- *Note that because today is not a REAL hazard perception test, there are several minor differences between what you are going to do on the laptops, and what the video instructs you to do; we will explain these differences after the video has concluded*
- *Please pay full attention to the video, as it does explain how you should be looking out for hazards, and specifically what a 'developing hazard' is - these are the ones we are wanting you to spot and respond to in the clips we show you.*

[Play video]

[Immediately after the video, say the following:]

- *There are a number of things in the video that are slightly different to what you are going to do now. Specifically:*
  - *We are going to show you 19 clips; 17 of these have one developing hazard each, and 2 have two developing hazards each.*
  - *You will respond by pressing the* ***space bar*** *- not the mouse button or any other button.*
  - *There will be no yellow or red circles in the videos on the laptops.*

  - *You will not see any 'red flags' on screen, but each time you press the space bar, you will see a green tick on the screen to show you have pressed. Please make sure you are pressing the space bar fully; you'll know if you have as a green tick will show.*
  - *This is NOT a real hazard perception test, and therefore you will not see any warning message for pressing too often or 'in a pattern' since you are not being scored as per a real test. Please just try to follow the instructions in the video.*
  - *Please press the space bar to respond to 'developing hazards' as soon as you spot them; remember a developing hazard is something you think would require you, as the driver, to take some kind of action such as changing speed or direction.*
- *Because these clips are either in the test or might be in the future, we have asked everyone to place their mobile phone on silent, behind their laptop, face down. No use of mobile devices or the internet is permitted while viewing the videos, especially taking pictures or videos.*
- *Please try and keep noise to a minimum to avoid distracting other participants, and please do not communicate with other participants.*
- *Note you will each be seeing different things at different times, so do not worry if other participants seem to be pressing the space bar when you aren't, or if other people finish before you.*
- *After the video clips, there will be a short questionnaire for you to complete on the laptop.*
- *Once everyone has finished, we will issue you each with a paper code for an Amazon voucher and ask you to sign to say you have received it. We will then take you back to reception to return your visitor passes.*
- *Finally, if any friends or family are taking part in the trial after you, please don't discuss it with them before they take part.*
- *Questions?*
- *You may now commence the trial by following the on-screen instructions.*

Appendix C Demographic and driving status questions

C.1 Learners

Please input your age:

How would you describe your gender

- Male
- Female
- Other
- Prefer not to say

How would you describe your learning to drive so far in terms of actual driving?

- I have not yet obtained my provisional licence
- I have my provisional licence but have not done any driving yet
- I have my provisional licence and have done fewer than 20 hours of driving (with an instructor, or with friends/family)
- I have my provisional licence and have done more than 20 hours of driving (with an instructor or friends/family)

How would you describe your learning to drive so far in terms of the hazard perception test (part of the theory test)? Please select all that apply.

- I have not yet done any preparation for my hazard perception test
- I have prepared for my hazard perception test – less than 10 hours
- I have prepared for my hazard perception test – more than 10 hours
- I have booked my theory test
- I have taken my theory test and failed so far
- I have taken my theory test and passed

Do you have a full driving licence issued in another country?

- Yes
- No

[If yes to previous question] In years, roughly how long have you held your foreign driving licence for?

C.2 Experienced

Please input your age:

How would you describe your gender

- Male
- Female
- Other
- Prefer not to say

Do you currently hold a driver's licence?

- Yes
- No

How many years have you held your driver's licence? Round up or down to the nearest whole year.

What is your annual mileage? Please provide as accurate an estimate as you can.

When you did your driving theory test (if you have done one), did it include a hazard perception element similar to the task you have just completed, where you watch video clips?

- Yes
- No
- I can't remember
- I have not done a theory test

Validation of CGI hazard perception clips

The UK hazard perception test was introduced into the driver licensing process in 2002. Previous work has shown that the hazard perception test is a reliable way of measuring a skill that is critical for safe driving, with candidates who score higher having a better safety record.

In 2015, the original video hazard perception clips were changed to CGI versions, which permit updating of 'look and feel' such as more modern vehicles in clips that are beginning to look dated. This study provides the first attempt to validate that a set of the CGI clips used in the official test, by showing that they can discriminate between two groups of drivers with very different levels of experience (those with more than 10 years' driving, and those who have not yet received their licence) based on their hazard perception skill.

Experienced drivers (N=105) and learner/pre-learner drivers (N=73) sat a hazard perception test with 16 clips, containing a total of 18 hazards. Hazard perception skill was measured in a similar way to how it is done in the official test, with responses to hazards in each of their 'scoring windows' being used to indicate how early participants spotted the hazards (a score of 0-100 being used, with each participant receiving a mean score based on their responses to all hazards).

The experienced group scored significantly higher (a mean score of 55.6) on hazard perception skill than the learners and pre-learners (a mean score of 48.3) showing that CGI clips can be a valid measure of hazard perception skill on this criterion. Two clips did not show this pattern of findings, and it is noted that this is something worthy of further investigation.

Other titles from this subject area

TRL558	Grayson, G. B. & Sexton, B. F. (2002). The development of hazard perception testing. TRL Report (TRL558). Crowthorne: Transport Research Laboratory.
RSRP81	Wells, P., Tong, S., Sexton, B., Grayson, G. & Jones, E. (2008). Cohort II: A Study of Learner and New Drivers Volume 1 – Main Report. Road Research Report 81. London: Department for Transport
https://doi.org/10.1080/15389588.2011.591856	Boufous, S., Ivers, R, Senserrick, T & Stevenson, M. (2011). Attempts at the Practical On-Road Driving Test and the Hazard Perception Test and the Risk of Traffic Crashes in Young Drivers. Traffic Injury Prevention, 12(5), 475–482

TRL
Crowthorne House, Nine Mile Ride,
Wokingham, Berkshire, RG40 3GA,
United Kingdom
T: +44 (0) 1344 773131
F: +44 (0) 1344 770356
E: enquiries@trl.co.uk
W: www.trl.co.uk

ISSN 2514-9652

DOI 10.58446/wxvf1428

PPR2071