Correlations among self-report, static image, and video-based hazard perception assessments: The validity of a new Lithuanian hazard prediction test

Auksė Endriulaitienė[1], Laura Šeibokaitė[1], Rasa Markšaitytė[1], Justina Slavinskienė[1], David Crundall[2], Petya Ventsislavova[2]

[1]Vytautas Magnus University, Lithuania,
[2]Nottingham Trent University, United Kingdom

Corresponding author:
Laura Šeibokaitė, Vytautas Magnus University, Jonavos str. 66-330, Kaunas, Lithuania, LT-44191. laura.seibokaite@vdu.lt

Abstract

Scholars suggest that improving hazard perception (HP) skills among different road users can lead to safer driving. To this end, a small number of countries have developed a national hazard perception test for their licensing procedures. The purpose of the current research is to develop and validate a new video-based hazard prediction test for the Lithuanian driving context. Eighty-eight drivers participated in the study. Initially, 25 naturalistic clips were included and edited into a *hazard prediction* test following the “what happens next?” paradigm. The overall score of the test was higher for experienced drivers when compared with novices. On further inspection the overall group difference was driven by 12 clips. The Lithuanian Hazard Perception test HPT-LIT featuring static images (Tūskė et al., 2019) and Hazard Perception Questionnaire developed by White et al. (2011) were used to validate the new test.

This new version of the test (12 clips, LHP_{12}) showed acceptable psychometric properties. Drivers with less driving experience demonstrated poorer hazard prediction skills than experienced drivers. The results revealed a negative correlation between hazard prediction scores on LHP_{12} and crash involvement in the novice driver group, as well as a positive correlation between prediction scores and self-reported hazard prediction skills. It was concluded that the new Lithuanian hazard test shows potential for future research.

Keywords: Hazard prediction test, validity, novice drivers, experienced drivers, Lithuanian drivers.

Introduction

Road traffic crashes and fatalities remain a serious public health concern world-wide. Despite the efforts of different countries to implement preventive actions, the number of people killed or injured on the road remains high with 50 million injured per year (WHO, 2018). The high death and injury rates, and the associated economic, social, and psychological costs associated with them, require many governments to look for more efficient traffic safety policies and measurements (Ventsislavova et al., 2019). Scholars suggest that good hazard perception (HP) skill is a core cognitive component of safe driving, and assessment and training of this skill could improve traffic safety significantly.

Hazard perception is typically defined as the driver's "ability to detect dangerous traffic situations" in sufficient time to make a safe response (Horswill, & McKenna, 2004, p.156). Within this skill lie a range of subprocesses that draw on a range of cognitive abilities, such as visual scanning, processing, anticipation, and appraisal (Moran et al., 2020; Pradhan and Crundall, 2016). This skill is typically measured by presenting drivers with a series of video clips, filmed from the perspective of a driver in a moving vehicle, each containing a hazard that viewers must detect. A traditional hazard perception test (as exemplified by the UK test that forms part of their licensing procedure) measures the time taken to respond to such hazards as the primary indication of skill, with shorter response times reflecting safer drivers (Horswill & McKenna, 2004).

Several studies have found that better hazard perception skill relates to lower crash involvement (e.g. Horswill, et al., 2020, Horswill et al., 2015, Rosenbloom et al., 2011), though only a small number of countries have included HP assessment and training as part of their licensing procedure (Moran, et al., 2019, 2020) or introduced intervention programs in order to improve road safety (Assailly, 2017; Slavinskienė et al., 2019). For example, the UK and some states in Australia have implemented the HP test as part of their official driving tests with data suggesting that they have been successful in improving road safety (Wells et al., 2008; Horswill et al., 2015).

Hazard perception in an international context

Why have more countries not developed their own national hazard test? Despite the many studies that have shown the HP test to differentiate successfully between safe and less-safe drivers (e.g. Cheng at al., 2011; Horswill et al., 2015; Rosenbloom et al., 2011), there are other studies that have reported failed attempts to separate driver groups on the basis of risk or experience (e.g. Lim et al., 2013; Sagberg & Bjørnskau, 2006; Yeung & Wong, 2015). It is possible that such equivocal results sow seeds of doubt among policy makers in different countries, especially in countries where no local research has been undertaken that examines hazard perception performance in relation to a specific geographical and cultural context.

One reason for these conflicting results in the literature lies with crucial differences in the design of these various tests produced by different research groups. There is no accepted standard of how clips should be produced, or what measures should be recorded from participants (Ventsislavova et al, 2019). This inevitably leads to a wide variation in test types, and without empirical examination of every element of a hazard test, it is possible that some versions of hazard perception tests have missed a crucial element that renders them ineffective at separating driver groups based on risk or experience (Crundall et al., 2021, Ventsislavova & Crundall, 2018).

A lack of test consistency across the literature is unsurprising given the lack of research into the theoretical processes underpinning hazard perception skill, or even the lack of agreed terminology and definitions (Pradhan & Crundall, 2016). Though attempts have been made to link HP to various theories, they are often merely used as ways to describe the act of hazard perception rather than to motivate fundamental research in the area. Specifically, two theoretical frameworks are most frequently cited: Signal Detection Theory (SDT, Green, Swets, 1966), focusing on the drivers' ability to differentiate and react to hazardous and non-hazardous situations; and Endsley's model of situation awareness (Endsley, 1995; Moran et al., 2020), which likens hazard perception to the three stages of perception, comprehension, and prediction. A few researchers have suggested more specific frameworks to explain hazard avoidance, such as Pradhan and Crundall (2016). They emphasized a range of sub-processes including precursor prioritization, hazard processing, and appraisal, that interact with the distance at which one perceives hazardous cues. In addition to these limited attempts to link theoretical frameworks to hazard perception, recent research has also attempted to link domain-free cognitive skills to driving though usually without reference to an underlying theory (e.g. Moran et al., 2020, Mackenzie & Harris, 2017).

The lack of a standard hazard perception format was reflected in a recent systematic review by Moran et al. (2020). They identified four major types of test used in the wider research community – static images, video-based clips, driving simulators and naturalistic driving. Self-reported measures have also been discussed as a plausible measure of hazard perception skills (Scialfa et al., 2012; Scialfa et al., 2011). For instance, Abele et al. (2018) found that self-rated hazard perception skills were correlated with better performance in driving simulator hazard perception tasks. The validity and efficiency of static image tests has been supported by several studies, with experienced drivers identifying hazards within such images faster than novices (Scialfa et al., 2012, DiStasi et al., 2020). Performance on a static image test has also been related to crash involvement (Tūskė et al., 2019). The efficiency of static image tests is acknowledged by the fact that static HP test was included as a part of their official driving test in Netherlands (www.cbr.nl). However, the most common method of measuring HP skill is the video-based method (either using naturalistic clips recorded from real vehicles, or clips developed with computer-generated imagery). Non-staged, naturalistic hazardous video clips are cost-effective to produce in comparison to other methods (e.g. high-fidelity CGI or driving simulators), though do not afford control over how the hazard unfolds. Some studies have reported good psychometric qualities of video-based hazard perception tests in different countries (Lim et al., 2014; Ventsislavova et al., 2019), as well as more consistent results than other measurement strategies in terms of detecting more or less skilful drivers (Moran et al., 2019; Moran et al., 2020, Crundall et al., 2021). These results provide a steer for countries considering implementing their own hazard perception tests. The use of dynamic clips viewed on a computer monitor provides a level of realism beyond probing drivers' understanding of mere static images, while providing a more consistent and controlled assessment that can be obtained with simulators or on-road tests. As such, dynamic clips offer a middle ground, balancing the pragmatics of developing a large-scale test with the ability to generalize results back to on-road safety.

The type of participant response, however, must also be considered. The national UK hazard test requires participants to press a button whenever they see a hazard (e.g. Wallis &

Horswill, 2007). This response-time measure has a number of potential flaws however (Crundall, 2016). For instance, one cannot be certain that a button press is made in response to the appearance of a hazard just because the two are coincident in time. A distracted driver may have been responding to some other element in the scene which fortunately fell in a similar timeframe to the appearance of the hazard. Indeed, the requirement of a temporal scoring window creates many problems, such as excellent drivers making valid anticipatory responses just a fraction of a second before the scoring window opens. Such responses would be classed as misses even though they might reflect a very early (and therefore safe) identification of an upcoming hazard. Localised hazard perception responses have been used to avoid this problem (i.e. ‘click on the hazard when spotted’, e.g. Smith et al., 2009), though these are not without their problems (e.g. Ventsislavova et al., 2019).

Hazard prediction as an international alternative

One alternative to the traditional response time measure is to employ a *hazard prediction* methodology. This prediction paradigm asks participants to predict a hazardous situation following occlusion of the scene just as the hazard appears (e.g., Crundall, 2016). Following occlusion, participants are usually asked one or more probe questions about ‘what happens next?’ in the clip (i.e., if the occlusion had not happened, how would events have unfolded; e.g. Jackson et al., 2009). More recently, this procedure has been simplified to asking participants to choose between several on-screen options, each of which reflect a possible outcome for the occluded clip (Ventsislavova & Crundall, 2018). The rationale of this test is that participants must be looking in the right place at the right time to identify the clues to the imminent hazard. This visual interrogation of the scene, coupled with prioritization of precursors, leads drivers to inspect those elements of the scene that are most likely to result in a hazard. The clip is occluded a few frames after the hazard begins to unfold, providing the most accurate participants with confirmation that their prediction was correct.

It has been argued that the prediction paradigm is more suitable for international export as it removes many problems associated with response time measures. Criterion bias is a particular issue for countries where on-road hazards are more prevalent, where otherwise safe drivers may become so desensitized to hazards that they require more evidence that an unfolding event is more dangerous than everyday driving. This can result in slower responses in a traditional hazard perception test (e.g. Ventsislavova et al., 2019). The hazard prediction test however provides drivers with four options regarding ‘what happens next’ following occlusion of the clip. This gives a measure of accuracy as the primary variable, which is not confounded by the driver’s threshold for what constitutes a danger to them on the road.

The hazard prediction paradigm has been found to successfully differentiate between experienced and inexperienced drivers in different countries (Jackson et al., 2009; Crundall 2016; Castro et al., 2014; Lim et al., 2014). Three studies have directly compared the hazard prediction test to the hazard perception test. In two of these studies, the hazard prediction test out-performed the perception test (Crundall & Kroll, 2018; Ventsislavova et al., 2019) while the third study suggested that both formats could be successful under the right conditions, thought participants still favoured the prediction format (Crundall et al., 2021).

The Lithuanian context

Lithuania has only initiated non-systematic explorations of hazard perception so far (Tūskė et al., 2019; Slavinskienė et al., 2020). Despite the number of road traffic collisions (63 per million inhabitants per year, 175 deaths in 2020, the sixth place in Europe; ec.europa.eu, 2020; WHO, 2018), the HP test has not yet been considered as a possible candidate for training and testing due to the lack of consistent evidence-based policies and practices (Slavinskienė et al., 2020, Endriulaitienė et al., 2013). Consequently, the main objective of the current study is to develop and validate a new hazard prediction test adapted to the Lithuanian driving context. Despite the variety of available tests, scholars suggest the development of a specific version of the HP test that matches the infrastructure and traffic culture of each country (i.e. the driving clips must be specific to that geographic region and the hazards must by typical of local driving behaviours; Lim et al., 2014). The need to develop and validate clips specific to each country is evident due to the differences in nature of driving, legal and social regulations, as well as uneven frequency of typical hazards (Ventsislavova, 2019; Sun & Chang, 2018).

Lithuanian researchers have already attempted to validate a static image-based hazard perception test for scientific purposes (Tūskė et al., 2019), however there are at least two reasons to seek validation of a video-based hazard perception methodology. First, the majority of research in the field has employed a video-based methodology (Moran et al., 2019), and there is good reason to believe that dynamic stimuli are more likely to evoke a skill similar to that used on real roads. Second, if one accepts that the hazard prediction format captures a less confounded measure, then dynamic stimuli are necessary to provide a suitable context for a prediction to be made.

The most prominent validation criterion of available hazard perception and prediction tests is the capacity of the test to differentiate between groups of safe and less-safe drivers (often using the surrogate of driving experience; Horswill et al., 2020; Pammer et al., 2018, Tūskė et al., 2019). In the current research we hypothesized that the newly developed test will show good psychometric properties and will differentiate between novice and experienced driver groups. Our first hypothesis is that experienced drivers will outperform novices in prediction accuracy. Our second hypothesis is that the newly developed test will positively correlate with self-reported hazard perception skills and the scores of a previously-developed hazard test using static images (HP-LIT).

Method

Sample

Eighty-eight drivers participated in the study. They comprised two groups – 45 novices with driving experience less than 12 months and 43 experienced drivers who had more than 36 months of experience. The cut score of 12 months was selected based on previous suggestions in the hazard perception literature, and the typical step-change in crash risk that occurs over the first year of driving (Ventsislavova & Crundall, 2018). The mean age of novice drivers was 19.76 years, SD=3.51 (range 18-36 years), whereas in the experienced driver group the age range was 21-48 years, M=28. 28, SD=8.19. All participants held a provisional or full driving license and drove at least once per month. Forty-six percent reported everyday driving. More specific demographic details are provided in Table 1.

[Insert Table 1 here]

All respondents volunteered and received no incentive for participation in the study. The majority of the participants were Lithuanian university students. In order to increase the sample size, a snowball sampling procedure was used. Initial participants were invited via social media and were then they asked to nominate other eligible peers. This resulted in a convenience sample with the potential for bias due to the non-random selection procedure. In line with previous research (Sun & Chang, 2018) the self-reported crash rate was low. Only 6 drivers (two novices and four experienced) reported an at-fault collision in the past year, while 16 drivers reported that they received traffic fines in the same period. The study was conducted following the APA ethical standards and requirements for confidentiality and data protection. The study plan underwent ethical review in the process of application for funding, therefore institutional ethics approval was not required. Oral consent for participation was obtained from all participants. No potential harm was deemed likely to occur for participants.

Materials, instruments, and procedure

Filming

The video stimuli were filmed on Lithuanian roads in September 2019. Four GoPro HERO 4 mini cameras were attached to the car using suction mounts (one inside the front window, one inside the rear window, and two cameras were attached outside to the right and left mirrors). Go Pro Hero 4 Silver video cameras were used, recording in full high definition format (1080p, 16:9 ratio, medium-angle settings). The resolution of clips was 1920 x 1080. The filming was conducted in the daytime in different locations – cities, suburbs and rural areas. The driver of the film car was an experienced, native driver, with previous experience of conducting driving safety research. Video material from multiple cameras was synchronized with Adobe Premier CC and edited into a hazard perception video clip using a graphic overlay representing the interior of a car. The filming and editing protocol that was followed was identical to the one followed by Ventsislavova et al. (2019). A screenshot of the driving clip is presented in Figure 1, where the pedestrian is considered to be a hazard.

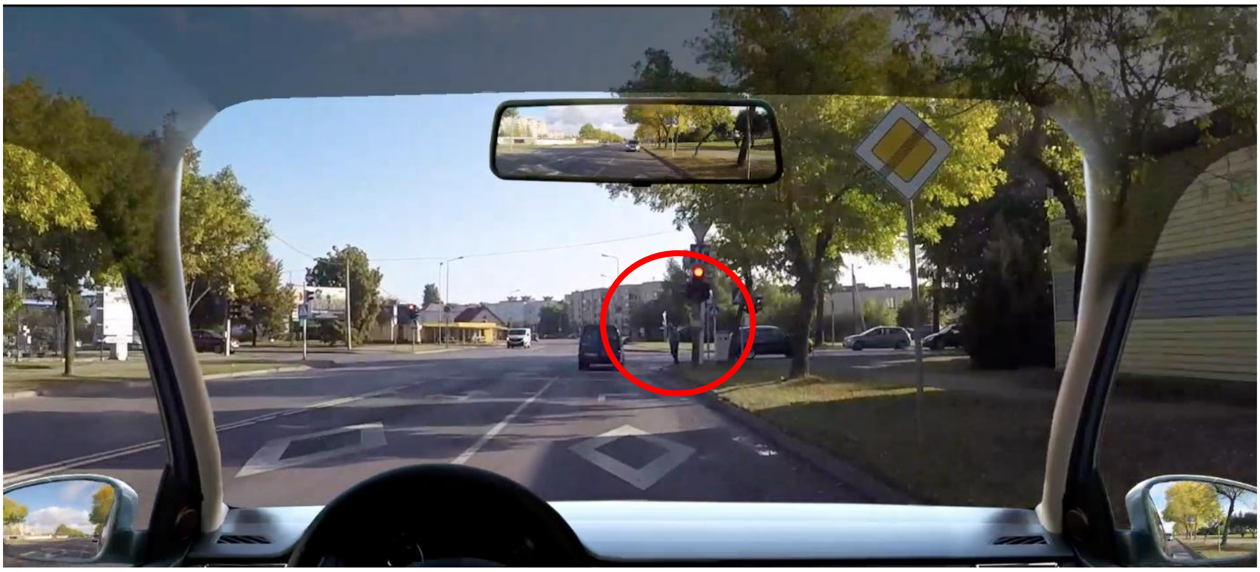

Figure 1. An example screenshot of a driving clip and a hazard. The circle is added to aid the reader in identifying the location of the hazard. Circles did not appear on the actual test clips.

Clip selection and editing

Initially, seventy hazardous clips were created and were evaluated by four traffic psychologists in order to select the final set of clips. The basic selection criteria for further use were as follows – the clip should contain only one hazard, and that hazard should have a precursor to allow accurate prediction of the hazardous situation. The precursors are defined as “the clues to an upcoming hazard” (Pradhan & Crundall, 2016, p. 64). Taking Fig. 1 as an example, the posture of the pedestrian – stood on the pavement yet leaning towards the road – serves as the precursor of this hazard. This indicates the intention of the pedestrian to cross the street. In addition, each clip should contain enough potential distractors to provide alternative options following occlusion. The number of potential distractors was similar, but not equal in each scenario. As the clips were created from field-based material (filmed in real driving situations), there was no possibility to constrain clips of in such a way to have an equal number of distracters. The final selection was comprised of 37 clips. These clips were edited and cut to occlude at a point immediately prior to the hazardous situation, such that none of the hazards fully materialised in the clips. Subsequently, two focus groups were organized in order to generate multiple choice options for each clip as potential answers to the question "What happens next?" One group consisted of two novices (with less than one year of driving experience) and two experienced drivers (with more than three years driving experience). The second group consisted of four traffic psychologists who developed the final list of options for each clip. Each clip contained four possible answer options (one correct and three incorrect, derived from the distracters identified by the focus groups) following examples from the literature (Kroll et al., 2020; Ventsislavova & Crundall, 2018).

Pilot study

All clips were subjected to a pilot study. Twenty-one novice drivers (with driving experience less than 12 months) and thirty-four experienced drivers (with driving experience more than 36 months) agreed to participate as volunteers. Those participants who agreed to participate in the study fully completed all tests and questionnaires (i.e., there were no partial attempts), resulting in all data being taken to the next stage. The mean age of novice drivers was 25.19 years, SD=7.89 (range 18-48 years), while the mean age of experienced drivers was 33.44 years, SD=8.98 (range 20-52 years). Forty-seven percent reported everyday driving. The procedure of testing was identical to that of the full study (see *Procedure*). The main purpose of the pilot study was to assess the difficulty of clips (i.e., whether there are any floor or ceiling effects) and to select those clips that successfully differentiate between the novice and experienced driver groups. Initial piloting reduced the set from 37 to 25 of the most promising clips. The occlusion points were re-edited for some of remaining clips.

Additional stimuli

The previously validated Lithuanian Hazard Perception test HPT-LIT, featuring static images (Tūskė et al., 2019), was administered immediately after the HP test. The HPT-LIT test consisted of 27 static images (24 hazardous and 3 non-hazardous images). Examples of the static scenarios in the HPT-LIT test include: “In an urban two-lane street you are approaching a pedestrian zebra crossing. A pedestrian (hazard) is about to cross the zebra crossing from your right. Ahead of you a trolleybus is driving in your lane. The lane next to you is full of vehicles moving the same direction as you.”/ “In the highway you are driving the first lane, in front of you both lanes are occupied with vehicles. In the distance of 25-45 m vehicle (hazard) intends to enter your lane from the right exit lane”. The descriptions were not given to the participants. The full list of scenarios is provided in Tūskė et al. (2019). Each image was displayed on the screen for three seconds and participants were asked to report, by simply typing yes or no, whether they had seen a hazard. A total number of correct answers was calculated for each participant, with higher scores indicating better hazard perception skills. The internal consistency of the HPT-LIT was .75 (Cronbach alpha).

Self-reported hazard perception skills were also measured. Participants were asked to evaluate their skills (e.g. “spotting hazards quickly” or “reacting to more than one potential hazard at a time”) with the Hazard Perception Questionnaire developed by White et al. (2011). They were asked to compare their skills to a typical Lithuanian driver of their age and rate it on a seven-point scale. The internal consistency of this six-item scale was .95. (Cronbach alpha).

Procedure of test development and validation study

Participants first completed a demographic questionnaire, followed by the newly developed Lithuanian hazard prediction test. Then, they were asked to complete the previously validated HPT-LIT featuring static images (Tūskė et al., 2019). Finally, drivers were asked to give a self-evaluation of their HP skill. All tests and questions were accessed online via a weblink provided to participants.

The clips of the Lithuanian HP test were presented on a computer screen and participants were instructed to use the mouse and keyboard to answer to the hazards. The size of monitor was not controlled as respondents completed the test using their personal devices in convenient time and space. Participants were instructed to complete the test on a device with a large monitor and to avoid using small, mobile devices to access the tests. While participants might have used either a PC or tablet, the tests were designed not to work on smartphones. It was required that the entire test would be completed in one session.

They were asked to carefully watch the (twenty-five) short video clips from a driver's perspective, in full knowledge that each clip would stop immediately prior to a hazardous situation and they will be asked to predict "what happens next?". They were also aware that they would have to choose one out of four possible options. They were advised to select the answer that most accurately describes how the driving situation is going to develop. There was no time limit placed on making this response. The sequence of clips and order of answer options was randomised. The answers were coded as '0' for 'incorrect' and '1' for 'correct' option. The sum of correct answers was used as a final score of hazard prediction skills. Study participants were not provided with the feedback on the accuracy of their answer, there was no possibility to observe how the situation actually developed during the test. Table 2 provides the description of hazards in the 12 video clips that provided the greatest differentiation between the driver groups and were subsequently selected to be taken forward for future research as the new Lithuanian hazard prediction test.

The entire procedure took approximately 40 minutes.

[Insert Table 2 here]

Data analysis

Student's t and chi-square analyses for group differences were applied in order to finalize the number of clips to be included in the test. Only clips with bigger differences in accuracy scores between novice and experienced drivers were selected for final test. The difference was calculated by Cohen's d according to Chinn's (2000) suggestion to use the formula d=LN(odds ratio)/1.81. Then psychometric qualities of the test were examined looking at internal consistency (Cronbach alpha) and confirmatory factor analysis for content and construct validity. In order to test the first hypothesis that experienced drivers will outperform novices in prediction accuracy and to check criterion validity t-tests were performed. The second hypothesis that the newly developed test will positively correlate with self-reported hazard perception skills, the scores of a hazard test using static images (HP-LIT) and crash involvement was tested with Spearman's correlation coefficients.

Results

Participants' total hazard prediction score across all 25 clips differed significantly between novice and experienced drivers (M_{novice} = 9.67 (39%), SD_{novice} = 2.61, $M_{experienced}$ = 12.07 (48%), $SD_{experienced}$ = 3.60, t(76.39) = -3.570, p = .001, with unequal variances between groups based on Levene’s test). This demonstrates the ability of the full hazard prediction test to differentiate between the two driver groups.

Chi-square analyses were conducted on individual clips to compare the frequency of drivers from both groups who answered correctly (or incorrectly). Thirteen clips did not differentiate between the two groups. The remaining clips (12) suggested that experienced drivers scored better than the novices (see Table 3). Unfortunately, only five of these clips reached the statistical threshold for significance, though in our refinement of these clips we accepted a significance level of $p<0.20$ as indication of potential use in a future iteration of the test. When restricted to these 12 clips, the difference between the two groups understandably widened (30% vs. 50%).

The Cronbach α score for the final selection of clips (12) was 0.56 showing that the internal consistency of this tool might benefit from improvement. It should be noted however, that the construct of hazard prediction is of a multidimensional nature. The violations of conventionally accepted α scores are quite common and justified in previous research (Crundall et al., 2021; Kroll et al, 2020; Sorrel et al., 2016). The confirmatory factor analysis yielded adequate results, suggesting that the twelve clips compose one latent factor and the test shows adequate construct validity. One factor solution revealed good fit to data ($\chi^2(54)=58.7$, $p=.31$; CFI=.90, TLI=.88, RMSEA=.03, 90%CIs [.00 - .08]).

[Insert Table 3 here]

All 12 clips were also subjected to a t-test to compare the mean accuracy scores of each group of drivers to mean chance expectancy (25%). Three clips Clip E; Clip K and Clip L did not significantly differ from chance which could indicate that participants provided random responses for these clips. However, detailed analysis of response frequencies revealed that some of the distracters were chosen more frequently than others (respectively for Clip E: 23%, 38%, and 8 %; for Clip K: 32%, 7%, and 38%; for Clip L: 32%, 16%, and 29%). This result may indicate that participants were not guessing *per se*, rather one of the distracters was perhaps too convincing. Modification of the distracter options might improve the future validity of these clips.

In order to test the first hypothesis and assess the criterion validity of the newly developed Lithuanian hazard prediction test, the mean scores of novice and experienced drivers were compared using a t-test. The general mean score of hazard prediction skill (mean number of correctly predicted hazards across all clips) was significantly higher among the experienced groups (M=6.12, SD=2.18) in comparison to the inexperienced ones (M=3.60, SD=1.64), t(77.94)=-6.09, p < .001, Cohen's d=1.31. In addition, gender differences in hazard prediction skills were subjected to a t-test and no significant differences were observed in either novice (M_{males} = 3.62, SD_{males} = 1.66, $M_{females}$ = 3.59, $SD_{females}$ = 1.66, t(43)=0.04, p>.97) or in experienced drivers' group (M_{males} = 6.62, SD_{males} = 2.26, $M_{females}$ = 5.47, $SD_{females}$ = 1.95, t(41)=1.76, p>.09). This suggests that the ability of the new hazard prediction test to differentiate between driver groups on the basis of experience is not obviously confounded with gender.

Differences between novice and experienced drivers were also found in self-reported HP (Student t test was used; Table 4). Experienced drivers reported higher scores of hazard perception skills than novices. No significant difference was found however in the scores of the static image-based HPT-LIT between these groups.

[Insert Table 4 here]

In regard to the criterion validity a correlation between hazard prediction scores and crash involvement was sought. The Spearman's rho score for the novice drivers' group was -.29, p = .05, suggesting marginal evidence for a correlation between hazard prediction skills and crash involvement with our novice drivers. The results in the experienced drivers' group were non-significant. Correlation analysis revealed that new Lithuanian hazard prediction test LHP_{12} was positively related to self-reported HP (Spearman's rho = .211, $p < .05$), but not related to static-image HPT-LIT (Spearman's rho = .027, $p > .05$). Thus, the hypothesis 2 was only partially confirmed.

Discussion

The main aim of this study was to develop a hazard prediction test suitable to differentiate between experienced and novice drivers in the Lithuanian driving context and explore its psychometric qualities. The video clips of the newly developed test captured a variety of road hazards such as pedestrians, cyclists, lane change, reversing cars, etc (see, Table 2). The final number of included clips was twelve and the suggested title of the instrument is LHP_{12}.

Both the LHP_{12}, and the 25 clips it is based on, successfully differentiated between novice and experienced drivers. Higher scores indicate better hazard prediction skills among experienced drivers when compared to novices. This supports good construct validity of this newly developed instrument and its potential for further use. The results confirm previous findings that hazard perception and prediction skills develop with driving experience naturally (Horswill et al., 2020; Crundall et al., 2021) and such tests are valid instruments for measuring this capacity in different countries (Ventsislavova, & Crundall, 2018; Ventsislavova et al, 2019). Also, the Lithuanian test revealed the association between correct hazard prediction and lower crash involvement in the novice driver group. This supports the previous findings that higher hazard prediction and perception abilities allow drivers to handle difficult traffic situations and prevent collisions (Horswill et al., 2020). Unexpectedly, the association was non-significant in the experienced driver group. This might suggest that crashes of experienced drivers are less related to hazard perception skills and more to other factors (like risk taking, distraction etc.; Sundfør et al., 2019). Alternatively, it might be hypothesized that the self-reported nature of crash-involvement data prompted a greater social desirability effect in this group. Still, it is hard to explain why novice drivers were not prone to this. Therefore, the relationship between our new hazard test and measures of crash risk (ideally, objective measures) should be pursued in future research.

It should be mentioned that the discriminative value of separate clips varied – in five clips out of twelve experienced drivers outperformed novices significantly, whereas in seven the differences were non-significant. This was similar to the findings of other researchers (Crundall et al., 2021) showing that the inclusion of clips that initially do not yield significant difference is common practice in hazard perception research. The rationale to include non-significantly differentiating clips in the final test was two-fold. First, if the significance level is close to .10, it might be expected that with a larger sample size the difference becomes significant in future studies. Furthermore, the inclusion of more clips in the test should reduce the chance of floor or ceiling effects, and the more items in a test will reduce the weight of a single item on the overall score and minimize the probability of errors (DeVellis, 2016). Despite the uneven discriminative power of the separate clips, the role of experience in the final score of test was highly significant, with an effect size of 1.31. In previous hazard prediction ("what happens next?") research this large effect size is also reported repeatedly (Crundall et al., 2021).

It might be argued that the newly developed Lithuanian hazard prediction test is quite hard as the mean scores are rather low for both groups. Nonetheless, we found nine percentage points difference between the correct answers of novices and experienced drivers (39% vs 48% respectively). Similar results were obtained by Crundall et al. (2021), in their study a comparable difference between learners and experienced drivers was 10 percent (45% and 55%

respectively). The scores were compared to chance expectancy and the results suggested that participants did not seem to randomly select the questions. Again, these results are in line with previous studies even among highly experienced groups, such as emergency or fire appliance drivers (Kroll et al., 2020). This confirms that the hazard prediction skill is complex and drivers who have not developed this skill might find it particularly difficult to predict hazards as it requires greater high order skills than just indicating that a hazard is present once it fully has materialised. Fortunately, these relatively low absolute scores do not remove the significant difference between drivers in terms of experience. However, the reasons underlying the consistently lower scores of hazard prediction paradigm compared to the hazard perception test raise an interesting question to be investigated in more detail in future research.

The internal reliability of our Lithuanian hazard prediction assessment tool is an ostensible problem, with Cronbach's Alpha for twelve clips only reaching .56, some way from the conventional level of internal consistency suggested in psychometrical literature (Sorrel et al., 2016). Some authors in road hazard prediction research report more respectable levels of reliability, up to and exceeding .70 in large samples (Horswill et al., 2020), however, it is more common to find reported Cronbach's Alphas of .50 (Horswill, & McKenna, 2004; Kroll et al., 2020). Researchers agree that hazard perception or prediction constructs are multidimensional and even a single item might measure several sub-processes (Ventsislavova et al., 2019). Therefore, it is hardly feasible to reach high intercorrelations among items. This suggests that LHP_{12} might benefit from some more developments or re-testing efforts in larger samples, although the reliability for scientific analysis is adequate.

The results yield some evidence about the criterion validity of the Lithuanian road hazard prediction test. Previous research reports no gender differences in hazard perception skills (Tūskė et al., 2019) and the same was found in the current study. This might serve as an argument for the appropriateness of the assessment tool. The correlation with self-reported hazard perception skill was significant although modest. As expected from our theoretical assumptions, drivers with better hazard perception skills report being capable to recognize hazards on the road more effectively. Some researchers argue that self-ratings of abilities are often over-estimated due to self-enhancement bias (Horswill et al., 2017), and this may have impacted on the strength of the correlation, with novices suffering greater relative bias than the experienced drivers. Therefore, the associations between subjective hazard perception skill evaluation and more objective measurement of skills with clips might be not very high. It should be noted that the self-reported hazard perception questionnaire was administered after completing the two hazard tests (dynamic and static), so the answers might be affected by actual performance on the tests. Though neither of the tests provided feedback, participants' perception of how well they performed might have served as the reference point for self-evaluation. Assessing self-ratings both before and after undertaking the tests will elucidate this issue in future research.

Contrary to expectations a significant correlation between scores on the static image and the new dynamic hazard test was not found. The static test was developed using a reaction time paradigm and measures hazard perception rather than prediction. Crundall et al. (2021) suggested that perception and prediction measurements focus on different underlying processes, therefore, the absence of relationship between two types of road hazard perception measurements is not an argument for poor validity of LHP_{12}. No differences between novice and experienced drivers in the static image test obtained in this study might advocate for a dynamic hazard prediction test. Nevertheless, further explorations on correlations between different assessment strategies are encouraged.

Some limitations of the current study should be noted. First, the internal reliability of new LHP_{12} was rather low, therefore the scores should be interpreted with caution. The instrument is useful for scientific purposes, though not yet applicable for individual assessment.

Low rates of crash involvement (especially, at fault crashes) likely limited the possibility of identifying any relationship between crash rates and prediction scores in the experienced group, though it is also possible that hazard prediction is less of a determinant of experienced drivers' crash risk. Also, the significant relation in the novice driver group between crash rates and prediction scores might be inconclusive due to the low variability of reported crashes. Larger sample sizes in future studies would help to solve this concern. In some scales (e.g., crashes, self-reported hazard prediction skills) social desirability bias might be evoked due to the self-reported nature of the measurement. This should be kept in mind when generalizing the results and future studies may look for objective measurements where possible.

The measure of driving experience based on years since licensing might have been imprecise. The bias towards the lower value might be expected, as not everyone begins to drive actively right after licensing. Finally, the size of the computer screen and resolution was not controlled, therefore the quality of video-material might be different across participants leading to differences in hazard prediction skills. Fortunately, the test was designed not to run on smartphones, thus eliminating these smallest of screens.

Despite limitations, the results of this study, and the newly developed LHP_{12}, extend the evidence for hazard prediction being relevant to yet more geo-specific contexts. The new test offers a low-cost method of measuring a complex cognitive skill that is extremely important for traffic safety. Furthermore, the development of a local measure of hazard skill is the first step in developing training resources to improve hazard awareness and reduce collisions within that context.

Conclusions

It might be concluded that newly developed Lithuanian hazard prediction test LHP_{12} is a valid instrument able to differentiate between hazard prediction skills of novice and experienced drivers. This conclusion is justified based on the difference in performance of our two groups on the full 25 clips, along with other results including the correlation of the test score with crash involvement among novice drivers, the correlation with self-reported hazard perception skills and non-significant gender differences. Despite a concern regarding internal consistency, the refined version of 12 clips offers a promising method for future exploration.

Acknowledgements

The study has received funding from the Research Council of Lithuania (LMTLT), agreement No. S-MIP-19-1

The authors would like to acknowledge the work Modesta Morkevičiūtė in synchronising and editing video clips.

References

1. Ābele, L., Haustein, S., Møller, M., & Martinussen, L. M. (2018). Consistency between subjectively and objectively measured hazard perception skills among young male drivers. *Accident Analysis & Prevention, 118,* 214-220.
2. Assailly, J. P. (2017). Road safety education: What works?. *Patient Education and Counseling*, *100*, S24-S29.
3. Castro, C., Padilla, J. L., Roca, J., Benítez, I., García-Fernández, P., Estévez, B., ... & Crundall, D. (2014). Development and validation of the Spanish hazard perception test. *Traffic Injury Prevention, 15(8)*, 817-826.
4. Chinn, S. (2000). A simple method for converting an odds ratio to effect size for use in metaanalysis. *Statistics in Medicine, 19*(22), 3127-3131.

5. 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.
6. Crundall, D. (2016). Hazard prediction discriminates between novice and experienced drivers. *Accident Analysis & Prevention, 86,* 47–58.
7. Crundall, D., & Kroll, V. (2018). Prediction and perception of hazards in professional drivers: Does hazard perception skill differ between safe and less-safe fire-appliance drivers?. *Accident Analysis & Prevention, 121,* 335-346.
8. Crundall, D., Van Loon, E., Baguley, T., & Kroll, V. (2021). A novel driving assessment combining hazard perception, hazard prediction and theory questions. *Accident Analysis & Prevention, 149*, 105847.
9. DeVellis, R. F. (2016). *Scale development: Theory and applications.* Los Angeles: Sage.
10. Di Stasi, L. L., Diaz-Piedra, C., Morales, J. M., Kurapov, A., Tagliabue, M., Bjärtå, A., ... & Catena, A. (2020). A cross-cultural comparison of visual search strategies and response times in road hazard perception testing. *Accident Analysis & Prevention, 148*, 105785.
11. Endriulaitienė, A., Šeibokaitė, L., Markšaitytė, R., Žardeckaitė-Matulaitienė, K., & Pranckevičienė, A. (2013). *Lietuvių rizikingas vairavimas: ką gali paaiškinti psichologiniai veiksniai. [Risky driving of Lithuanians: the contribution of psychological variables].* Monography. Kaunas: Vytautas Magnus University.
12. Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors, 37*(1), 32–64.
13. Europe 2020 Indicators. ec.europa.eu, 2020, visited on-line on December 10, 2020.

14. Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics* (Vol. 1). New York: Wiley.
15. 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.
16. Horswill, M. S., Garth, M., Hill, A., & Watson, M. O. (2017). The effect of performance feedback on drivers' hazard perception ability and self-ratings. *Accident Analysis & Prevention*, *101*, 135-142.
17. Horswill, M. S., Hill, A., & Jackson, T. (2020). Scores on a new hazard prediction test are associated with both driver experience and crash involvement. *Transportation Research Part F: Traffic Psychology and Behaviour*, *71*, 98-109.
18. Horswill, M. S., & McKenna, F. P. (2004). Drivers' hazard perception ability: Situation awareness on the road. In S. Banbury & S. Tremblay (Eds.). *A Cognitive Approach to Situation Awareness* (pp.155-175). Aldershot, UK: Ashgate.
19. Jackson, L., Chapman, P., & Crundall, D. (2009). What happens next? Predicting other road users' behaviour as a function of driving experience and processing time. *Ergonomics, 52(2),* 154-164.
20. Kroll, V., Mackenzie, A. K., Goodge, T., Hill, R., Davies, R., & Crundall, D. (2020). Creating a hazard-based training and assessment tool for emergency response drivers. *Accident Analysis & Prevention*, *144*, 105607.
21. Lim, P.C., Sheppard, E., Crundall, D. (2013). Cross-cultural effects on drivers' hazard perception. *Transportation Research, Part F, 21*, 194-206.
22. Lim, P.C., Sheppard, E., & Crundall, D. (2014). A predictive hazard perception paradigm differentiates driving experience cross-culturally. *Transportation Research Part F: Traffic Psychology and Behaviour, 26,* 210–217.

23. Mackenzie, A. K., & Harris, J. M. (2017). A link between attentional function, effective eye movements, and driving ability. *Journal of Experimental Psychology: Human Perception and Performance, 43*(2), 381-394. http://dx.doi.org/10.1037/xhp0000297

24. Moran, C., Bennett, J. M., & Prabhakharan, P. (2019). Road user hazard perception tests: A systematic review of current methodologies. *Accident Analysis & Prevention, 129*, 309-333.

25. Moran, C., Bennett, J. M., & Prabhakharan, P. (2020). The relationship between cognitive function and hazard perception in younger drivers. *Transportation Research Part F: Traffic Psychology and Behaviour, 74,* 104-119.

26. Pammer, K., Raineri, A., Beanland, V., Bell, J., & Borzycki, M. (2018). Expert drivers are better than non-expert drivers at rejecting unimportant information in static driving scenes. *Transportation Research Part F: Traffic Psychology and Behaviour*, *59,* 389-400.

27. Pradhan, A.K., & Crundall, D. (2016). Hazard avoidance in young novice drivers: definitions and a framework. In: Fisher, Donald L., Caird, Jeff K., Horrey, William J., Trick, Lana M. (Eds.), *Handbook of Teen and Novice Drivers: Research, Practice, Policy, and Directions* (Pp. 439-450). CRC Press, Boca Raton, Florida, United States. https://doi.org/10.1201/9781315374123.

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

29. Sagberg, F., Bjørnskay, T. (2006). Hazard perception and driving experience among novice drivers. *Accident Analysis and Prevention, 28*(2), 407–414.

30. Scialfa, C. T., Borkenhagen, D., Lyon, J., Deschênes, M., Horswill, M., & Wetton, M. (2012). The effects of driving experience on responses to a static hazard perception test. *Accident Analysis & Prevention, 45*, 547-553.

31. Scialfa, C. T., Deschênes, M. C., Ference, J., Boone, J., Horswill, M. S., & Wetton, M. (2011). A hazard perception test for novice drivers. *Accident Analysis & Prevention, 43*(1), 204-208.
32. Slavinskienė, J., Endriulaitienė, A., Šeibokaitė, L., & Markšaitytė, R. (2020). Is hazard perception related to risk taking attitudes and intentions? In *Transport means-2020: proceedings of the 24th international scientific conference, September 30-October 2, 2020, Kaunas, Lithuania. Part 2: proceedings. Kaunas: Kaunas University of Technology, 2019, 23, 2,* 721-725.
33. Slavinskienė, J., Markšaitytė, R., Endriulaitienė, A., & Šeibokaitė, L. (2019). The effectiveness of road hazard perception training: literature review. In *Transport means-2019: proceedings of the 23rd international scientific conference, October 2-4, 2019, Palanga, Lithuania. Part 2: proceedings. Kaunas: Kaunas University of Technology, 2019, 23, 2,* 879-883.
34. Smith SS, Horswill MS, Chambers B, Wetton M. (2009). Hazard perception in novice and experienced drivers: the effects of sleepiness. *Accident Analysis & Prevention*, *41,* 4, 729-33.
35. Sorrel, M. A., Olea, J., Abad, F. J., de la Torre, J., Aguado, D., & Lievens, F. (2016). Validity and reliability of situational judgement test scores: A new approach based on cognitive diagnosis models. *Organizational Research Methods*, *19*(3), 506-532.
36. Sun, L., Li, S., & Chang, R. (2018, October). Development and validity of Chinese hazard perception test. In *International Conference on Man-Machine-Environment System Engineering* (pp. 257-262). Springer, Singapore.
37. Sundfør, H. B., Sagberg, F., & Høye, A. (2019). Inattention and distraction in fatal road crashes–Results from in-depth crash investigations in Norway. *Accident Analysis & Prevention, 125,* 152-157.

38. Tūskė, V., Šeibokaitė, L., Endriulaitienė, A., & Lehtonen, E. (2019). Hazard perception test development for Lithuanian drivers. *IATSS Research*, *43*(2), 108-113.
39. Ventsislavova Petrova, P. (2019). *Development of a cross-cultural hazard perception test* (Doctoral dissertation, Nottingham Trent University).
40. Ventsislavova, P., & Crundall, D. (2018). The hazard prediction test: A comparison of free-response and multiple-choice formats. *Safety Science*, *109*, 246-255.
41. Ventsislavova, P., Crundall, D., Baguley, T., Castro, C., Gugliotta, A., Garcia-Fernandez, P., ... & Li, Q. (2019). A comparison of hazard perception and hazard prediction tests across China, Spain and the UK. *Accident Analysis & Prevention*, *122,* 268-286.
42. Wallis, T. S., & Horswill, M. S. (2007). Using fuzzy signal detection theory to determine why experienced and trained drivers respond faster than novices in a hazard perception test. *Accident Analysis & Prevention*, *39(6)*, 1177-1185.
43. Wells, P., Tong, S., Sexton, B., Grayson, G., Jones, E. (2008). *Cohort II: A Study of Learner and New Drivers.* Volume 1—Main Report (Vol. 81). Road Safety Research Report.
44. White, M. J., Cunningham, L. C., & Titchener, K. (2011). Young drivers' optimism bias for accident risk and driving skill: Accountability and insight experience manipulations. *Accident Analysis & Prevention*, *43*(4), 1309-1315.
45. World Health Organization. (2018). *Global status report on road safety 2018: Summary* (No. WHO/NMH/NVI/18.20). World Health Organization.
46. Theory exam. www.cbr.nl, retrieved 27/01/21.
47. 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.

Table 1. Demographic characteristics of study participants

	Novice drivers	Experienced drivers	Total
N total	45	43	88
Males	13	24	37
Females	32	19	51
Age in years (mean, SD)	19.76 (3.51)	28.28 (8.19)	23.92 (7.55)
Months since licensing (mean, SD)	6.8 (3.76)	112.6 (81.99)	58.49 (77.98)
Mileage/exposure (km/week) (N in percent)			
Less than 50	42.2	4.7	23.9
51-200	44.5	55.8	50
More than 200	13.3	39.5	26.1
Frequency (N in percent)			
Once per month	11.1	2.3	6.8
Less than once a week	17.8	2.3	10.2
Several days per week	48.9	23.2	36.4
Every day	22.2	72.2	46.6
Self-reported penalties in the past year	1.07 (0.25)	1.35 (0.57)	1.20 (0.46)

Self-reported collisions in the past year	1.04 (0.21)	1.09 (0.29)	1.07 (0.25)

Table 2. The list of twelve clips and description of hazards that were included in a final version of new Lithuanian hazard prediction test

Clip	Content	Duration of the clip until occlusion point (sec)	Hazard	Precursor
Clip A	In an urban street you are approaching the crossing. A pedestrian is visible on the right corner. Two cars in front of you turn right. The clip occludes as the pedestrian is about to step into the street from your right.	28	Pedestrian from the right crosses the street	Posture of pedestrian – leaning toward crossing
Clip B	You are driving on a busy suburban road. A bus stops at the bus stop in the opposite traffic lane on your left. The clip occludes as the car behind the bus drives into the opposite (into your) lane in order to overtake the bus.	14	The car behind the bus drives into your lane	The approaching car from opposite lane turns towards central line
Clip C	While travelling in the middle lane of a three-lane carriageway you are approaching a junction where cars from the right can drive into your road. The clip occludes as the bus is about to pull into your lane from the right.	13	The bus pulls into your lane from the right	Buss on the right slowly approaches your lane
Clip D	After turning right, you enter the two-lane old town street, you see immediate junction of two streets. You intend to turn left. You see a car parked on the street you intend to go, a pedestrian on the pavement and two people with bike. The	37	The cyclist goes with bike on the street you intend to turn in	The cyclist moving on the street with no intention to

	clip occludes as the cyclist steps with bike on the street you intend to turn in.			avoid approaching vehicles
Clip E	You are driving in the middle lane of a tree-lane one-way urban street. Several buses are driving on a bus lane on your right, you pass one of them. Suddenly the car in front of you brakes as the bus far in front is changing into second lane. The clip occludes as the parallel bus is about to change into your lane too.	30	The parallel bus changes into your lane	The breaking lights of the vehicle in front of you and of the bus on the right
Clip F	You are driving in the first lane of a two-lane urban street behind a tram. You change into the second lane intending to overtake the trolley. The clip occludes as the trolley is about to enter your lane just in front of.	24	The trolley enters the second lane in front of you	The breaking lights of the trolley in front of you
Clip G	While traveling in the second lane of the two-lane urban street you pass many cars that are parked on your left. On your right a bus is driving parallel with you. The clip occludes as the bus is about to start entering your lane.	9	The bus changes into your lane	The bus from the right turns very closely to the line between two lanes
Clip H	You are driving in the second lane of a two-lane one-way urban street behind another car. On the road further you see a van approaching your roads from the side street. The car in front brakes and changes to the second lane, you change afterwards too. The clip occludes as the car in front brakes again in order to park on the left.	17	Car in front of you parks in the left	Breaking lights of the car in front

Clip I	You are driving in the middle lane of a three-lane carriageway. On the road further a dark car in the first lane brakes and changes into your lane overtaking a white car that has stopped in the first lane. The clip occludes as another black car reverses from the right into your lane.	16	The car reverses from the right	The car in the first lane breaks, back of the reversing car is slightly visible
Clip J	While traveling on a narrow street in the old town you approach a truck of road workers that is parked on your right and see pedestrians walking in front of that truck from the right (the truck obstructs them). One worker stands at the back of the truck, another – on the left side of the street. A bit further several pedestrians are walking on the left side of the street. The clip occludes as pedestrians from the right are about to cross the street from behind the truck.	22	Pedestrians from the right cross the street	Pedestrians from the right approach the street, then their view is blocked with the parked truck (they are visible for a very short time)
Clip K	You are driving in the second lane of a two-lane urban street. Many cars are passing by on the opposite lane. The clip occludes as suddenly a car comes out from the left side road and intends to turn into your lane.	17	The car comes out from the left side road	White car from the left starts turning into your lane
Clip L	While traveling in the second lane of a two-lane busy urban street you pass a trolley. The clip occludes as a car is about to enter the second lane directly from the right (from behind the trolley).	14	The car enters the second lane directly from the right	The car from the right starts driving into your lane

Table 3. Differences between novice and experienced drivers in their performance on each video clip

	Clip	Number of correct predictions (percentage)		Chi-square	Significance level p	Odds ratio	95CIs	Cohen's d*
		Novice drivers	Experienced drivers					
1	Clip A	17 (37.8%)	31 (64.6%)	10.40	<.001	4.25	1.73-10.4	.80
2	Clip B	14 (31.1%)	21 (48.8%)	2.88	.09	2.11	.89-5.04	.41
3	Clip C	20 (44.4%)	26 (60.5%)	2.26	.14	1.91	0.82-4.46	.36
4	Clip D	18 (40.0%)	30 (69.9%)	7.86	<.01	3.46	1.43-8.37	.69
5	Clip E	11 (24.4%)	18 (41.9%)	3.02	.08	2.23	.90-5.53	.44
6	Clip F	10 (22.2%)	25 (58.1%)	11.80	<.001	4.86	1.92-12.3	.87
7	Clip G	23 (51.1%)	30 (69.8%)	3.20	.07	2.21	.92-5.29	.44
8	Clip H	27 (60%)	35 (81.4%)	4.84	<.05	2.92	1.10-7.71	.59
9	Clip I	3 (6.7%)	11 (25.6%)	5.88	<.05	4.81	1.24-18.7	.87
10	Clip J	4 (8.9%)	9 (20.9%)	2.53	.11	2.71	.77-9.59	.55

11	Clip K	7 (15.6%)	14 (32.6%)	3.50	.06	2.62	.94-7.33	.53
12	Clip L	8 (17.8%)	13 (30.3%)	1.88	.17	2.00	.74-5.47	.38

---

* Cohen’s d was calculated according to Chinn’s (2000) suggestion using the formula d=LN(odds ratio)/1.81

Table 4. Group comparison in HPT-LIT and self-reported HP scale

			HPT-LIT			
	Min	Max	M	SD	t (df)	p
Novice	7	24	16.71	3.66	t(77.73)=0.971	.331
Experienced	5	24	15.81	4.89		

			Self-reported HP			
	Min	Max	M	SD	t(df)	p
Novice	1.17	6.50	4.29	1.16	t(86)=-2.97	.014
Experienced	1.50	7.00	5.00	1.07		