

Proactive Listening to a Training Commentary improves Hazard Prediction

Castro, Candida ^a, Ventsislavova, Petya ^c, Peña-Suarez, Elsa^a, Gugliotta, Andres^a, García-Fernandez, Pedro ^b, Eisman, Eduardo ^a and Crundall, David^c.

^aCimcyc. Mind, Brain and Behaviour Research Centre. Faculty of Psychology. University of Granada, SPAIN

^b Electronic and Computer Technology. Faculty of Sciences. University of Granada, SPAIN

^cNottingham Trent University. UK

Email addresses: candida@ugr.es

Corresponding author:

Cándida Castro Ramírez

Experimental Psychology Department. Faculty of Psychology University of Granada
Campus Cartuja, s/n 18011 Granada. SPAIN

Cimcyc – Mind, Brain and Behaviour Research Centre

Universidad de Granada. Campus Cartuja, 18011. Granada. Spain. Room 330

Phone: + 34 958 240663 +34 655 99 40 59

Fax: +34 958 246239

ABSTRACT

The aim of this work was to explore the effect of Proactive Listening to a Training Commentary, using the recently developed version of the Spanish Hazard Perception test. Firstly, 16 videos were used in the pre-test session in its short version, cut to black just before the hazard appearance. The What Happens Next Assessment (at the pre-test stage) generates expectations about the outcome of the traffic situation. Then, the training (8 minutes in length) uses the complete version of the same 16 videos, revealing the hazards unfolding. It involves listening to a voice with relevant information about where to allocate attention in the complex driving scene in order to recognise and anticipate the hazard successfully. A total of 121 participants were included in this study. The sample consisted of learner, novice and experienced drivers, including re-offender and non-offender drivers. The participants were divided into 2 groups: a trained and an untrained group. Two assessment times were used: pre-test (16 videos) and post-test sessions (another 16 videos). The test presented a high internal consistency ($\text{Alpha} = 0.875$). This training shows significant positive effects for all types and groups of participants. No significant differences were found between the non-offender and the offender groups. Performance in gradual-onset hazard events can be improved after training but also by practice; however this training is essential and especially beneficial for training the ability to detect hazards that appear abruptly (which seems to be difficult to improve just by practice).

Keywords:

Hazard perception, Training, Proactive Listening, Commentary Training, Driving, Hazard Detection, Situation Awareness

1. Introduction

The ability to perceive hazards while driving is a factor that reduces the risk of having accidents (Wells et al., 2008; Horswill et al., 2010a). A hazard in the traffic context is any permanent or temporary object which remains in the road environment and has the potential to increase the risk of an accident (Jackson et al., 2009). Hazard perception is the skill of detecting, evaluating and

reacting to events on the road that have a high probability of producing a collision (Crundall et al., 2012) and is the only specific driving ability that correlates with a lower crash risk (Wells et al., 2008; Horswill et al., 2010a). Therefore, it is considered that improving the skill of hazard perception through training could decrease the crash risk. Beanland et al., (2013) assert that the training of higher-order cognitive skills, such as hazard perception, addresses the broader driving context, particularly anticipating or avoiding hazardous situations. Recognition of the importance of these cognitive skills has led to a proliferation of driver training programmes that directly target these skills.

In fact, it seems worth questioning whether training improves the ability to detect hazards of only learner drivers and drivers with less experience. It is possible that training could be effective for all groups of drivers, including those with considerable driving experience, safe drivers and re-offenders. However, training may not be equally effective when drivers are exposed to different types of hazard. It would also be interesting to find out whether practice in itself, using *What Happens Next* exercises (WHN) (McKenna and Crick 1997), would be enough to improve hazard perception test scores. In each of these exercises, trainees had to view video footage of a traffic situation, which was freeze-framed at a given point (usually just before a hazard was encountered) and at that point trainees were asked “what might be about to happen”.

Nevertheless, as Wetton et al., (2013) explained, the WHN exercises did appear to have a significant immediate training effect, independent of the expert commentary exercises, but the magnitude of this effect was reduced. That is, if one had to choose between using either WHN exercises or expert commentary exercises, then one would choose the latter. According to Endsley (1995), Situation Awareness operates at three levels that support hazard perception and make it possible to answer the three main questions: “What is the hazard?”, “Where is the hazard?” and “What happens next?” which means perceiving and understanding the hazardous situation and anticipating future driving events (Jackson et al. 2009).

Specifically, Wetton, et al. (2013) investigated what type of training would be most useful to improve hazard perception by using video-based tests featuring real driving footage at three points: immediately prior to the test (pre-); immediately post-test (post-); and after a one-week delay. They created four types of video training. The first was WHN, based on McKenna and Crick (1997) as referred to above. The second video was expert commentary training. The third type was hybrid commentary training (i.e., expert plus self-generated commentaries); and the fourth consisted of a full training package (i.e., WHN plus hybrid commentary training). All four types of training presented significant results compared with results from untrained groups. However, full training resulted in the greatest improvement and *WHN* Training the least. The addition of self-generated commentaries to the expert commentary training (hybrid commentary condition) did not significantly improve response times. All training effects decayed considerably after the delay, but the effect of full training remained significant. Although no benefit was found in adding self-generated commentaries to expert commentaries, the possibility remains that the *WHN* exercises may provide an additional benefit when combined with commentary training.

In another study, Isler et al., (2008), taught drivers how to identify hazards by detecting clues from the environment using commentary training while concurrently performing a secondary tracking task, simulating the steering of real driving. After the training, novice drivers detected a higher percentage of hazards and had faster response latencies compared to a baseline than those without training. Crundall et al., (2010) investigated whether learner drivers would benefit from being trained to produce a commentary drive. They compared one group of commentary-trained learner drivers to a control group. The results showed that the trained group had fewer crashes, reduced their speed sooner on approaching hazards and applied pressure to the brakes sooner than untrained drivers.

There is also evidence that training in hazard perception benefits both novice and older drivers as they both reduce their significantly speed when approaching hazards (Horswill, et al., 2010b). For instance, Horswill et al. (2010a) used a video of a driver's eye view of hazardous

traffic situations. The participants in the trained group heard an expert driving instructor giving a running commentary on the footage, indicating what he was paying attention to and giving general advice about anticipating hazards. The following excerpt from the commentary is typical:

“Scanning ahead. Looking over the crest of the hill. Car turning left. Approaching traffic. More cars coming towards us. Cars on the right. Checking amongst the trees.”

On the other hand, Meir et al., (2014) explored the formulation and evaluation of a new HP training test –the Act and Anticipate Hazard Perception Training (AAHPT) in young novice-drivers. There were three types of test mode (*Active, Instructional and Hybrid*) and a Control group. *Active members* observed video-based traffic-scenes and were asked to press a response button each time they detected a hazard. *Instructional members* underwent a tutorial which included both written material and video-based examples regarding HP. *Hybrid members* received a condensed theoretical component followed by a succinct *Active* component. The Control group was presented with a road safety tutorial. According to their results, one week later, the *Instructional* mode demonstrated inferiority in comparison to the other two modes; the *Active* and the *Hybrid* mode members were more aware of potential hazards relative to the control group.

However, the *Instructional* mode of training could be carried out as an active mode of training too. As McKenna et al (2006) pointed out, commentary training improved drivers’ situational awareness and led them to a better appreciation of the risks, by encouraging them to actively search for hazards. Although commentary training doesn’t necessarily require a simulator response, it still provides an active search guide to the participants. The fact that commentary training based on instructions can be applied without using a simulator means that a less expensive tool is available that doesn’t require great amounts of time, money or effort and that could be just as effective. During the hazard perception exercise and when the video footage is cut, the driver generates a process which consists of selection of information and decision making. Once commentary training begins (visualizing the complete driving scene), drivers initiate an active listening process, which directs the top-down allocation of attention and generates expectations in

drivers as to what may happen in the immediate future. Participants are eager to find out *WHN*, or in other words, they expect feedback from their performance, which is the best guide they could have. Indeed, these sequences of action assume an active role by drivers that culminates in expectations. It can also guide their attention as well as arousing expectations of receiving feedback on their performance.

Moreover, it seems worth exploring whether the training has a different effect on the improvement in their perception according to the kind of hazard. Underwood, et al., (2013) suggested the following classification of hazard situations: those where hazards appeared gradually vs. those where hazards appeared abruptly. The gradual onset hazard videos are those that show events unfolding (for example, a football flying out of a driveway can predict that children are nearby). (Horswill and McKenna, 2004; Underwood, 2007). However, the abrupt onset hazards are those that involve the abrupt capture of attention and exogenous events (for instance, a pedestrian appearing suddenly). This type of hazard is under consideration for inclusion in driving tests, but it could be considered that what they are testing is the viewer's speed of reaction rather than their ability to assess a scenario and anticipate how the situation will develop. Experienced drivers gained an advantage in those situations where the hazard appeared gradually. This is probably due to the fact that gradual onset provides clues that allow experienced drivers to figure out how the situation will develop. So, it is expected that experienced drivers may have a more developed awareness of events on the motorway and of the behaviour of other road users.

Furthermore, according to White et al. (2011) young drivers show an optimism bias for their driving skills and accident risk perceptions. In addition, when comparing their driving self-assessment with their actual behaviour, there are indications that they overestimate their driving skills (Craen et al., 2011). On the other hand, multiple road offenders obtain different hazard prediction scores from normal/safe drivers (see the classic study by Pelz and Krupat, 1974). The implication is that good drivers are more likely to avoid accidents than are drivers with a record of offending. According to Simon and Corbett (1996), results of accident history are positively related

to offending. The number of accidents and offences is higher among young men and their index of accidents is higher than those of women or older drivers (Laapoti et al., 2001; Yahya and Hammarstroöm, 2011). Lapham et al., (2006) stated that repeat offenders are more likely to be involved in fatal motor vehicle crashes, hit and run collisions with pedestrian fatalities and to have a high blood alcohol concentration (BAC) when driving (0.15 mg/dl and above) than first-time DUI offenders (Beirness et al, 1991; Fell, 1993, 1995; Solrick and Hernenway, 1994).

Research Aims

The first research aim was to assess the effect of the video-based Proactive Listening to a Training Commentary (PLTC) on participants' hazard prediction performance, and secondly, to compare the improvement of performance in hazard prediction skills of groups of different driving experience (,learner, novice and experienced drivers) and non-offenders vs. re-offenders, in various types of hazardous situation. That is, in order to check whether different hazardous situations distinguish between drivers, two hazardous situations were manipulated: hazards that appeared gradually (where the hazard could be predicted by using clues from the environment) and hazards that appeared abruptly (the abrupt hazards appear a few milliseconds before the video stops, requiring direct detection).

For these purposes, a new and improved version of the Hazard prediction Test adapted to a Spanish driving context (HP-WHN, Castro, et al. 2014) was used to measure the targets, emulating the WHN test (McKenna and Crick, 1997; Jackson, et al., 2009) for the assessment. Clips between 6 and 26 seconds long were presented to the participants and were stopped immediately prior to the hazardous situation and then three questions were asked for each video: What is the hazard? Where is it located? and WHN? The primary task was to detect and identify the impending hazardous traffic scenarios and following this, participants were exposed to the commentary training guidance provided by an expert, while watching the complete version of the pre-test videos and carrying out an active listening task. The commentary training had a guide function for participants, leading

their attention to what was relevant, i.e. where to look and how to use the visual information to make predictions.

2. Method

2.1 Design

Different mixed ANOVA designs were used in this study. The dependent measure was accuracy of the participants in the test (i.e. average in the HP test, max. 6). The 3 repeated measures independent variables were session (pre-test; post-test) and type of question (What is the hazard?, Where is the Hazard? and What Happens next?), and two types of hazard (gradual and abrupt onset hazards). The 3 independent variables measured between groups were training condition (trained group; untrained group), experience of the drivers (learner, novice and experienced drivers) and recidivism condition (non-offenders vs. re-offenders). The sphericity assumption was not achieved; therefore, the Greenhouse-Geisser correction was used.

2.2 Participants

A total of 121 drivers took part in the current study; 69 participants were men and 52 were women. Three experience groups were considered (see Table 1, top): (a) 20 (16.5%) learner drivers (18-37 years) who were attending lessons to obtain a driving licence for the first time, (b) 62 (50.4%) novice drivers (18-39 years) who were in possession of a driving licence and had less than eight years' driving experience, (c) 40 (31.1%) experienced drivers (26-53 years) who possessed different types of driving licence. Specifically, 20 of the novices and 20 of the experienced drivers were re-offenders. Re-offender drivers was recruited from collaborating driving schools in Granada (Spain) while they attended a driving re-education and recidivism prevention course (i.e., the course known as “*Re-obtaining the total number of points*”, which is compulsory in Spain for drivers who have been banned from driving after losing all the available points of their driving licence); their socio-demographic characteristics are shown in Table 1, bottom.

Insert Table 1 about here

All participants were recruited from either collaborating driving schools in Granada (Autoescuela la Victoria, Autoescuela Luna and Autoescuela Genil-Ogíjares, Granada, Spain) or the School of Psychology of the University of Granada. Ethical principles in the declaration of Helsinki for research involving human participants were followed in the current study.

Insert Table 2 about here

2.3 Materials

2.3.1 Videos

Thirty-two HD (high definition) videos with a resolution of 1920 X 1080, taped from the driver's viewpoint, were used for the test. For the recording of the videos, the protocol developed by the University of Nottingham, UK was used in order to control the bias involved in recording (see Wetton et al., 2011). A total of 300 videos were recorded in the city of Granada (Andalucía, Spain), during the autumn of 2012 and spring of 2013, by two experienced drivers who are part of the research team. Driving routes were chosen according to the Accidents Report of 2011 (National Department of Traffic, 2012). The videos included different road types, comprising 11% motorway outside the city (A-44 and A-92 highway) and 89% urban roads (the neighbourhoods of Sacromonte, Almanjayar and Zaidin, which are typical Andalusian). Hazard situations consisted of 50% cars, 25% pedestrians, 7% motorcycles, 11% trucks and 7% buses. These hazards appeared out of side streets, at junctions etc.

All videos were preserved in their original version and have not been retouched. There were no accidents during the recordings. Ethical principles in the declaration of Helsinki for research involving human participants were followed.

The presentation of each video was as follows: First, a black screen with the corresponding numerical code of the video appeared and then immediately following this, one of the driving scenes was presented at random. The clips' duration ranged between 6 and 26 s. and they were

stopped immediately prior to the hazardous situation, immediately cutting to a black screen which concluded the trial.

2.3.1.2 Proactive Listening to a Training Commentary (PLTC)

For the training session, a full version of the first 16 videos (pre-test) was developed, which included a voice that described in detail the complete traffic scene. The guide revealed the most important clues from the environment that would help to detect the hazards. An example of the PLTC that guided the participants' eye movements and visual search was as follows: *"We are on the A-92 heading towards Guadix at the height of Loma Verde, where it crosses the motorway from Granada to Jaen. The truck in front of us has just left the highway to Jaen. A red car on the motorway access lane is trying to join our lane. ATTENTION, perhaps we cannot facilitate its access to the lane because another vehicle behind us is approaching fast, so we would not be able to perform the manoeuvre."*

2.4. Questionnaire

For the study, a new and improved version of the HP-WHN test (Castro et al., 2014) was used. This new version of the HP-WHN test contains a brief demographic questionnaire and a total of 32 response forms (1 page per video) for all 32 video clips, including instructions on how to complete it. Participants were required to complete the response form at the end of each video clip by responding: (1) What is the hazard? (2) Where is it located? and (3) WHN?

For 'What is the hazard?' participants obtained 2 points if they gave an exact description of the hazard (e. g. red car in the left lane), 1 point if they gave a partially correct answer (e.g. a car on the left, but without giving any details of its characteristics or location or if there was more than one) and 0 points for an incorrect answer. For 'Where is the danger?' 2 points were given if the cross was marked in the restricted area where the hazard appeared, 1 point if the cross appeared near or around the restricted area and 0 points if the cross was outside the two previous areas. Finally, to score the 'WHN?' question, 2 points were given if an exact description of what would happen was provided: "The car yielded because it could not change lane and allow the red car to

merge smoothly with the traffic”, 1 point if the description was not complete, and 0 points if the answer was incorrect.

2.4 Procedure

A total of 121 participants took part in this study and all of them were pre-assigned to different groups: groups that were exposed to the **PLTC** and Control groups. Prior to the experiment, participants were provided with instructions and at the same time, an experimenter explained to them how to answer the questionnaire. They all completed a demographic information form before the beginning of the experiment.

Participants completed the test in group sessions. They filled in the socio-demographic questionnaire individually. Then the video clips were presented in groups while the participants were seated at a distance of between 3 and 5 metres from a projection screen. The videos were shown on the screen at a 1920 x 1080 resolution using a video projector connected to a standard computer. After each clip, participants were required to complete the response booklet following each video, answering ‘What is the hazard?’, ‘Where is the hazard?’ and ‘WHN?’

The study consisted of two experimental parts: During the first part, both trained and untrained groups were presented with the 16 Session A clips. Each clip started from a black screen and ended on a black screen immediately prior to the hazardous event, with a self-paced progression dependent on the amount of time participants required to fill in the questions between clips. Following the pre-test clips, the trained group were presented with the complete version of these videos (the guide voice described the complete version of the first 16 traffic scenes used for the assessment), providing all the details that lead the attention to what is relevant in the traffic scene, in order to detect the hazard easily, and also the outcome of the traffic environment providing delayed feedback. Participants had to carry out an active listening task, paying closer attention to the guide and the traffic scene. The control group took a break of 10 minutes. The trained group also took a break of 10 minutes after the intervention.

During the second part of the study and after the break, the post-intervention assessment was undertaken. Both groups (Trained and Untrained) had to watch the 16 Session B video clips (videos 17 to 32) and the procedure was identical to the first part. At the end of the experiment, both groups were asked whether they had any doubts or wanted more details about the study. For ethical reasons, when the Untrained group asked for an explanation, they were allowed to watch the PLTC videos too, and in that way they received the training at the end of the experiment, once the pre and post-session measures had been taken.

2.5 Data analysis

Once data collection was completed, to confirm the consistency of marking, 50% of the score sheets (randomly chosen) were scored by an independent researcher, and agreement was measured using Cohen's Kappa for each question. According to these analyses, the two evaluators generally agreed on the response correction $\kappa = .95$ for 'What is the hazard?', 'Where is the hazard?' $\kappa = 1$ and $\kappa = .94$ for 'WHN?' Disagreements were discussed and a conclusion was reached on each occasion. Given the high level of agreement, a single researcher scored the remaining scripts.

The scores obtained from pre-test and post-test clips were jointly submitted to classic item analysis and reliability analysis. In particular, a minimum acceptable item-total correlation was set at .30 and a high Cronbach's Alpha of 0.875 for each total scale was expected.

Only the videos that achieved discrimination indices greater than 0.30 were taken into account in the analysis. That is, 11 videos out of 32 (5 from the pre-test and 6 from the post-test) that did not accomplish this criterion, were withdrawn from further analysis. The final version of the hazard prediction test was composed of the 21 remaining hazardous situation video clips (11 videos are pre-test and 10 are post-test). Average scores were calculated for all 21 videos and then separately for the 11 pre-test and 10 post-test sessions. Average score was calculated separately for the 9 gradual-onset hazard events and 12 abrupt-onset hazard events. It should be noted that the

scores obtained with these two types of video clip were also divided into pre-test and post-test. This distinction was intended to measure separate issues and the interpretation of the scores was actually different for gradual-onset hazard events and abrupt-onset hazard events.

Effect size is listed as partial Eta squared (partial η^2), demonstrating the proportion of the total variance explained by a variable that is not explained by other variables in each mixed model ANOVA specified. Effect sizes (partial η^2 and η^2) of mixed model and paired-samples would be considered as follows: 0.01 a small effect size, 0.06 a medium effect size and 0.14 a large effect size (Cohen, 1988; Richardson, 2011). The level of statistical significance was set at 0.05 in all analyses. Planned comparisons were used as a post hoc test with Bonferroni correction. All statistical analyses were performed using IBM SPSS Statistics v19 for Windows.

3. Results

3.1 Psychometric properties

Taking into account the discrimination indices of the items are greater than 0.30, this version of the Hazard prediction Test consists of 21 videos (11 videos are pre-test and 10 are post-test), with an alpha coefficient of 0.875. Table 2 shows the descriptive statistics of the test items and the discrimination indices.

Insert Table 2 here

3.2. Training effects

A (2)×2 mixed-model ANOVA was conducted with session (pre-test and post-test) as the repeated measures factor and training (trained and untrained) as the between subjects factor. A significant main effect between pre- versus post-test [Wilks' Lambda=.666 F(1,119)=59.654 p=0.001 partial η^2 =0.334] and significant interaction between pre- versus post-test and training group was found [Wilks' Lambda=.918 F(1,119)=10.642 p=.001 partial η^2 =.082]. Paired-samples t-tests indicated that the HP scores of the trained group in **PLTC** improved between pre-test (2.19;

average scores out of 6) and post-test (3.05) [$t(60) = -7.532$ $p = .001$ $\eta^2 = .49$] and the untrained group also improved between pre-test (2,3) and post-test (2.6), [$t(59) = -3.265$ $p = .002$ $\eta^2 = .15$]. The trained group (3.05) outperformed the untrained group (2.6) in post-test [$t(119) = 2.008$ $p = .047$ $\eta^2 = .033$], (Figure 1).

Insert Figure 1 here

3.3 Effect of the training and type of stimulus

A $(2) \times (2) \times 2$ mixed-model ANOVA was conducted, where the repeated measures independent variables were session (pre- and post- session) and type of stimulus (gradual and abrupt hazards) and the independent variable measured between groups was training condition (trained and untrained).

The 3-way interaction between session, type of stimulus and training was not significant [Wilks' Lambda=1.00 $F(1,118) = .007$ $p = .934$ partial $\eta^2 = .00$]; for this reason one 2×2 mixed model ANOVA per type of stimulus was conducted.

The 2-way interaction between the session (pre- and post- session) and training condition (trained and untrained) was found to be significant [Wilks' Lambda=.913 $F(1,118) = 11.188$ $p = .001$ partial $\eta^2 = .087$].

The 2-way interaction between the session (pre- and post- session) and type of stimulus (gradual and abrupt hazards) was also found to be significant [Wilks' Lambda=.891 $F(1,118) = 14.424$ $p = .001$ partial $\eta^2 = .109$].

These were the only significant interactions found.

The main effect of type of stimulus was significant [Wilks' Lambda=.657 $F(1,118) = 61.489$ $p = 0.001$ partial $\eta^2 = .343$]. Gradual-onset hazard events ($M = 2.18$) are more difficult to detect than abrupt-onset hazard events ($M = 2.87$). It was for this reason that a specific analysis for each type of hazard (abrupt and gradual) was performed, in order to compare the results of the different groups of drivers.

The main effect of session was also significant [Wilks' Lambda=.532 F(1,118)=103.744 p=.001 partial η^2 =.468].

No more main effects were found to be significant.

3.3.1. Training gradual hazards

A (2)×2 mixed-model ANOVA was conducted with session (pre-test and post-test of gradual-onset hazard events) as the repeated measures factor and training (trained and untrained) as the between subjects factor. The results showed a significant main effect between pre- versus (1.74) post-test (2.62) in gradual-onset hazard events [Wilks' Lambda=.635 F(1,118)=67.947 p=.001 partial η^2 =.365] and a significant interaction between pre- versus post-test and training group was found [Wilks' Lambda=.954 F(1,118)=5.731 p=.018 partial η^2 =.046]. Paired-samples t-tests indicated that the scores of the trained group in gradual-onset events improved between pre-test (1.6) and post-test (2.8) [t(59)=-7.639 p=.001 η^2 =.50] although the untrained group also improved between pre-test (1.8) and post-test (2.4) [t(59)=-4.074 p=.001 η^2 =.22], (See Figure 2).

Insert Figure 2 here

3.3.1. Training abrupt hazards

A (2)×2 mixed-model ANOVA was conducted with session (pre-test and post-test of abrupt-onset hazard events) as the repeated measures factor and training (trained and untrained) as the between subjects factor. A significant main effect between pre-test (2.6) versus post-test (3.2) was found in abrupt-onset hazard events [Wilks' Lambda=.910 F(1,119)=11.751 p=.0001 partial η^2 =.090] and a significant interaction between pre- versus post-test and training group was found [Wilks' Lambda=.944 F(1,119)=7.113 p=.009 partial η^2 =.056]. Paired-samples t-tests indicated that the scores of the trained group in abrupt-onset hazard events improved between pre-test (2.6)

and post-test (3.2) [$t(60)=-4,255$ $p=.001$ $\eta^2=.23$] but there was no significant difference between pre-test (2.7) and post-test (2.8) in the untrained group [$t(59)=-.545$ $p=.587$ $\eta^2=.01$], Figure 2.

3.4 Experience effects

3.4 Effect of training and driving experience

A (2)×2×3 mixed-model ANOVA was used to examine the differences between session (pre-test and post-test) as the repeated measures factor, training (trained and untrained) and the groups of drivers with different traffic experience (learner, novice and experienced drivers) as the between subjects factors. Significant main effects between pre- versus post-test [Wilks' Lambda=.720 $F(1,115)=44.683$ $p=.001$ partial $\eta^2=0.280$] and the interaction between session and training group were found [Wilks' Lambda=.943 $F(1,115)=6.928$ $p=.010$ partial $\eta^2=.057$]. A significant main effect of experience was found [$F(2,115)=5.915$ $p=.004$ partial $\eta^2=.093$]. Paired comparisons showed that experts (3.24) outperform learner drivers (2.34) in the post-test. All the experience groups improved their performance after training, but this improvement was greater for those drivers who already held a driving license: novices and experts. The biggest improvement was found for the trained group of experienced drivers, who improved their performance between pre-test (2.3) and post-test (3.3). The trained group of novice drivers also improved their performance between pre-test (2.2) and post-test (3.0) but not as much. And finally, the trained group of learner drivers improved their performance the least between pre-test (1.8) and post-test (2.3), Figure 3.

Insert Figure 3 here

3.5. Effect of the training and recidivism

A (2)×2×2 mixed-model ANOVA was used to examine the differences between session (pre-test and post-test) as the repeated measures factor and, training (trained and untrained) and the groups of drivers with different driving records, recidivism (multiple road offenders and non-

offender drivers)) as the between subjects factors. To conduct this analysis, learners were excluded. Significant main effects between pre- versus post-test [Wilks' Lambda=.640 $F(1,97)=54.456$ $p=.001$ partial $\eta^2 =.360$] and the interaction between session and training group were found [Wilks' Lambda=.892 $F(1,97)=11.788$ $p=.001$ partial $\eta^2 =.108$]. The main effect of recidivism was not significant [$F(1,97)=.101$ $p=.752$ partial $\eta^2 =.001$], the averages of multiple road offenders were 2.38 in pre-test and 2.98 post-test; and the averages of non-offender drivers were 2.21 in pre-test and 2.81 post-test.

4.5 The effect of experience and type of question vs. type of hazard

A (3)×(2)×3 mixed-model ANOVA was used to explore the differences between the repeated measures type of question (What, Where and WHN) and type of stimulus (gradual and abrupt hazards) and the between groups measure of experience condition (learner, novice and expert).

The 3-way interaction between the three factors was not significant [Wilks' Lambda=.969 $F(2,116) = 1.638$ $p=.199$ partial $\eta^2 =.027$]

The 2-way interaction between the type of question and the type of stimulus was significant [Wilks' Lambda=.543 $F(1,116)=17.902$ $p=0.001$ partial $\eta^2 =.134$]. It was for this reason that a specific analysis for different types of hazard (abrupt and gradual) was performed.

No more interactions were found to be significant.

Three main effects were significant: The main effect of type of question [Wilks' Lambda=.435 $F(1,116)=20.486$ $p=.001$ partial $\eta^2 =.150$] was significant. The main effect of type of stimulus [Wilks' Lambda=.51 $F(1,116)=109.39$ $p=.001$ partial $\eta^2 =.48$] was significant. The main effect of experience [$F(2,116)=6.21$ $p=.003$ partial $\eta^2 =.097$] was significant.

4.5.1. Experience and type of question in gradual hazards

A (3)×3 mixed-model ANOVA for the gradual-onset hazards was used to examine the differences between type of question (What, Where and WHN) as the repeated measures factor and the groups of drivers with different traffic experience (learner, novice and experienced drivers) as the between subjects factors. Significant differences were found for gradual-onset hazards related to the type of question/hazard [Λ Wilks=.785 $F(1,117)=3.355$ $p=.050$ η^2 $p=.028$] and significant main effects for the type of experience [$F(2,116)=3.915$ $p=.023$ η^2 $p=.063$]. Planned comparisons showed significant differences between learners ($M=.574$ maximum 2 points) and experts ($M=.769$ maximum 2 points), $p=.020$. Specially, significant differences were found in ‘What is the hazard?’ ($p=.013$) and in ‘Where is the hazard?’ ($p=.028$) questions. In addition, a significant difference between learners and novices was found ($p=.028$) in ‘What is the hazard?’ and a marginal significance between these groups (learners and novices) in ‘Where is the hazard?’ ($p=.073$) (Figure 4 top).

4.5.2. Experience and type of question for abrupt hazards

A (3)×3 mixed-model ANOVA for the abrupt-onset hazards was used to examine the differences between type of question (what, where and WHN) as the repeated measures factor and the groups of drivers with different traffic experience (learner, novice and experienced drivers) as the between subjects factors.

Similar results were found for abrupt-onset hazards: there were significant differences for type of question related to the type of hazard [Λ Wilks=.380 $F(1,117)=30.582$ $p=.001$ η^2 $p=.207$] and significant main effect for the type of experience [$F(2,117)=6.811$ $p=.002$ η^2 $p=.104$]. Planned comparisons showed significant differences between experts ($M=1.09$ maximum 2 points) and learners (.82 maximum 2 points), $p=.002$; and between experts and novice drivers (.93 maximum 2 points), $p=.021$. Specially, these significant differences were found in ‘What is the hazard?’ and ‘Where is the hazard?’ questions (Figure 4 bottom).

4. Discussion

For the current research, a new and improved version of the Hazard prediction Test was created for a Spanish sample. The first research aim was to assess the effect of the video-based **PLTC** on participants' hazard prediction performance, and the second was to compare the improvement of performance in hazard prediction skills of groups with different driving experience (experts, novices, learner drivers), in various types of hazardous situation. That is, in order to check whether different hazardous situations discriminate between safe and unsafe drivers, two hazardous situations were manipulated: hazards that appeared gradually (where the hazard could be predicted by using clues from the environment) and hazards that appeared abruptly (requiring direct detection).

A new version of the video clips was developed, considering previous literature (e.g., Crundall et al., 2010, 2012; Jackson et al., 2009, Wetton et al., 2011) and previous work of this research team (Castro et al., 2014).

We have improved the selection of videos **(which ended with a sudden occlusion prior to the hazardous situation starting, yet with enough information for the viewer to predict or at least make an educated guess as to what might happen next)**. **The video photograms were cut precisely, in the exact millisecond when the hazards started to unfold**, set it up in different hazard categories (abrupt and gradual appearance) and improved the formulation of 'Where is the danger?' and 'What happens next?' questions. We asked participants to answer the questions immediately after each video (Jackson et al. 2009), in order to analyse hazard detection, situational awareness and the projection they had of the future traffic scene. For the training sessions a full version of the first 16 videos (pre-test) was developed, adapted to the Spanish driving context, including a voice that described in detail the complete traffic scene.

The results showed an acceptable psychometric reliability of the new version of the Hazard prediction Test HP-WHN in the Spanish driving context and it appears to be a useful tool for studying hazard prediction. The original version of the test was composed of 32 videos (16 pre-test and 16 post-test). However, taking into account the videos that achieve the criterion of a discrimination index higher than 0.30, the final version of the test used for the analysis was composed of 21 video clips (11 pre-test and 10 post-test sessions). This cleaned up version showed a Cronbach's Alpha of 0.875. The **PLTC** is effective for improving hazard prediction performance. The trained group showed higher means in comparison to the untrained group in a post-test session, and these results support the conclusions of Wetton et al., (2013) and Underwood et al., (2013): commentary training can improve drivers' hazard prediction response. Specifically, it appears that **PLTC** somehow guides the attention as well as arousing participants' expectations of receiving feedback on their performance. That is, this training provides useful feedback on whether the participants were right or wrong in the pre-test session, which then helps them improve their performance in perceiving the oncoming hazard.

One crucial theory that explains how we perform visual search tasks efficiently is Guided Search (Wolfe, 1994). According to this theory, both bottom-up (stimulus-driven) and top-down (goal-driven) factors may contribute to the topography of this activation. In many studies, participants are instructed to view scenes without any particular task in mind so that stimulus driven (bottom-up) processes guide visual attention (Hwang et al., 2009). However, whenever there is a search task, goal-driven (top-down) processes tend to dominate guidance. In other words, during visual search tasks, in which subjects are asked to find a particular target in a display, top-down processes play a dominant role in the guidance of eye-movement (e.g. Henderson et al. 2007, Petter and Itti, 2007, Pomplum, 2006, Zelinsky, et al., 2006).

The **PLTC** helps drivers not only by providing knowledge but also by increasing sensitivity to hazards. Learning becomes easier when expectations are involved. Drivers received the expected information (feedback) about the hazard occurrence and could appreciate whether their

performance succeeded or failed, which results in affective value. All these facts would provide drivers with useful patterns of visual search that could enable them to predict more accurately the appearance of possible hazards by gathering clues from the traffic scenes.

Specifically, gradual-onset hazard events seem to be more difficult to detect than abrupt-onset hazards and these results are similar to those of Underwood et al., (2013) who found better results for abrupt onset hazards (that gained faster responses) than gradual onset hazards. He found that the abrupt-onset hazards (attention-capturing hazards) certainly attracted faster responses (mean of 1.79 s) than the gradual-onset hazards (3.87 s). The gradual appearance of the hazard precursor seems to be more difficult to detect than the abrupt appearance of the real hazard, even when the clip cuts to black just as the hazard starts to emerge.

However, in the current study, training and practice improves prediction of gradual hazards more than it improves detection of abrupt hazards. Results for gradual-onset hazards always appeared better in the post-test sessions than in the pre-test sessions. For the gradual onset hazards, both the trained and the untrained groups showed an improvement in their performance after the post-test sessions, which indicates that the detection of gradual hazards could also be improved merely by practice. Gradual hazards have more precursors and therefore are more likely to benefit from practice and training. Nevertheless, as Wetton, et al. (2013) explained, the WHN practice did appear to have a significant immediate training effect, independent of the expert commentary exercises, but the magnitude of this effect was less and it would possibly not endure over time. That is, if one had to choose between using WHN exercises or Proactive Listening to a Training Commentary, then one would choose the latter or even better, the combination of both methods as we did.

On the other hand, abrupt-onset hazards seem to be easier for drivers to detect. However, to achieve a better performance for abrupt hazards, the implementation of **PLTC** is required. For the abrupt onset hazards, after the post-test session, only the trained group showed an improvement in their performance, which indicates that the detection of gradual hazards could not be improved

merely by practice. That is, for the abrupt-onset hazards, the trained group showed a significant improvement in performance after the post-test session whereas the untrained group did not show a significant improvement. This could be due to the fact that abrupt hazards capture the attention rapidly and there is not enough time to take advantage of the environmental clues. It is likely that either there is not enough time to perceive these clues or there is not enough time to process and consider them (Vargas et al., 2008). As a result, it is much more difficult to improve detection of abrupt hazards merely with practice, without the aid of the environmental clues provided in advance of the hazard's appearance. The PLTC seems to play a crucial role, mainly when the danger appears abruptly, in helping drivers to anticipate where it is worth allocating their attention (the most relevant part of the complex traffic scene).

As expected, learner drivers obtained lower average scores in comparison with novice and experienced drivers (e.g. Pollatsek et al., 2006; McKnight and McKnight, 2003; Fisher et al., 2006).

All driving groups benefit from the Proactive Listening to a Training Commentary, but the greatest improvement post-test was found for the experienced drivers, followed by the novice drivers and finally the learner drivers.

On the one hand, the fact that experienced drivers proved to be the group benefitting most could be due to the advantage they have in experience in comparison with the other groups. The importance of the "observer features" should be noted: a solid base of experience makes learning easier and, for that reason, novice and learner drivers, who haven't yet been exposed to the same number of traffic situations, obtained inferior results. Novice drivers process traffic situations more slowly than experienced drivers; therefore, if the situation becomes very complex, contains many cues or there is not enough time to process them, novices will be at a greater disadvantage. However, the last two groups also showed improvement, which means that the improvement in detecting hazards using PLTC is effective, albeit at different levels.

Nevertheless, it can be considered a counter-intuitive finding that learner drivers show the least benefit from Proactive Listening to a Training Commentary. Assuming a plateau in HP skills,

one would imagine that learners/novices have more ‘room for improvement’. This finding is important because novices are at greater risk and we need to devise training strategies that target them rather than their parents. But could PLTC be used as a tool to help learner and novice drivers reach an acceptable level of understanding of the driving situation and, at the same time, develop their sensitivity to hazards, taking advantage of their expectations to benefit from the full length videos provided in the training session? Or could it be that the experts benefit most because the commentary is pitched at them? Imagine a chemistry class of first year and third year undergraduates. If the guest lecturer talks in terms more familiar to the third year students, they will gain more than the first years. Or could it even be that actually processing a commentary is a secondary task – it might help at one level, but at another level it might hinder performance. According to a paper accepted in JEPA, listening to commentaries can negatively affect eye movements (Young, Chapman and Crundall 2014). Perhaps the benefit to experts of the extra information outweighs the costs, but is the opposite true for novices? Should commentary driving only be an advanced training tool (as it currently is in the UK)?

In addition, a peculiar pattern of learning was found with regard to the performance of novice drivers. When abrupt-onset hazards were presented, the performance of novices was similar to that of the learner drivers and significantly worse than that of experts. Conversely, for gradual-onset hazards, the performance of novice drivers approached that of experts and was significantly different to that of learners. Novices (as they are acquiring some expertise) benefit most because the commentary is now beginning to be pitched at them. They seemed to be able to make use of the environmental clues that help to anticipate the hazard, and so improved their performance. However, novice drivers have not yet registered enough memory records of traffic situations that could help them explore the traffic scene or inform them when hazard anticipation is needed for abrupt-onset hazards. Learner and novice drivers find it more difficult to guide their visual search intentionally (top-down). For this reason, we consider that these groups would need to develop their search strategies more because they would probably fail to perceive the source of the hazard or

not perceive it quickly enough to allow an understanding of the future situation (Jackson et al., 2009).

In any case, learner drivers showed the worst performance when compared with the other groups. According to Box and Wengraf (2013), once drivers have driven 1000 kilometres, their abilities can be considered equal to those of drivers with 3 or more years of driving experience and experience reduces the risk of accidents for all driving groups. Drivers need more practice in order to develop their skills and become expert in more complex tasks. And, in previous research (Castro et al, 2014), it was discovered that novices find it more difficult to detect complex driving situations, for example, differentiating between quasi-dangerous situations and dangerous situations. In this study, we observed parallel results: novices have difficulty in situations of greater uncertainty or those in which there are no clues or insufficient time to understand the traffic scene (e.g. abrupt hazards). Lack of experience makes the anticipation of hazards more difficult. Both novice and learner drivers strive more when driving and this implies more errors (Logan, 1988). This is because these drivers have fewer records in memory to use while driving; only driving experience would make possible the use of successful solutions, previously learned. As a result, the performance of experienced drivers seems to involve less effort. Therefore, there is less disruption caused by distraction and more consistent execution.

Experts outperformed the other groups particularly when the response to type of stimulus vs. type of question was measured. Specifically, there were differences between experts, novices and learners in detecting both gradual and abrupt onset hazards, so the best results were obtained for the “What is the hazard?” and “Where is the hazard?” questions. These results are similar to those of Jackson et al., (2009), regarding the pattern of responding to the questions, indicating that identifying the hazard was easier than noting the location and predicting what would happen next and, at the same time, support the idea that driving training increases Situation Awareness (Walker et al., 2009). Experienced drivers are more aware of the information concerning both types of hazard whereas learners, due to their inexperience, may need more time to process the hazard and

get used to being alert on the road. Also, learner and novice drivers might have assigned fewer intentional resources to the task.

It should be noted that there is little recent literature focusing on training re-offenders and non-offenders in hazard prediction tasks. We compared the performance of offenders to that of non-offender drivers and the results indicated that commentary training was equally effective for both. Offenders did not obtain better results in comparison to the non-offender group, which indicates that probably the re-offenders had similar patterns of processing, understanding the information and performing the Hazard prediction task but different decision-making and execution to those of non-offenders when driving on their own. Further research is required to ascertain the differences between these two groups, considering cognitive processes, decision-making and execution of the manoeuvres in real driving. It should be noted that both groups are quite singular and this fact could affect and impair the perception process in real driving, e.g. driving under the influence of alcohol, or at very high speeds.

For instance, recently Yahya and Hammarström, (2011) analysed a total of 1,995 records with the aim of determining to what extent differences exist between drunk and sober drivers. They found that 88% of the offenders were men and 12% women. The proportion of drunk drivers is 25% for men and 20% for women. The group of drivers below the age of 35 is over-represented in the register in relation to the population register. Due to this last fact, it would be a complex task to achieve equivalence for these groups. Richard et al., (2013) pointed out that driving at excessive speed still remains an unsolved problem. Speed is a contributory factor for accidents and even though there have been attempts to solve this problem; there has been no significant reduction in traffic accidents produced by speed.

Finally, complementary analyses were made to evaluate potential differences between the two groups of drivers recruited (see also Table 2). Results showed that the multiple offender group included more males. In addition, as expected, they had previously lost their driving licenses more times, had received more tickets during the last 12 months and had more traffic incidents reported

to an insurance company. They also tended to drive more kilometres per year and to have been involved in more accidents during the last 12 months (it should be noted that the multiple offender drivers had been banned from driving for some months before attending the re-education course). In consequence, the differences in these and other socio-demographic variables should be carefully considered in future studies. However, no differences were found between the ability of the non-offender and the re-offender drivers to correctly identify hazardous situations in the video clips.

5. Future research

It is necessary to continue this research in order to determine the effect that different hazards have on learners, novices and experienced drivers. In fact, even experienced drivers failed when accurate anticipation of what would happen next was required, for both abrupt and gradual-onset hazards. Therefore, we consider that all groups of drivers could benefit from using PLTC. However it must be considered the potential for repeated training sessions over time. For instance, gradual onset HP performance didn't exceed 50% even for trained participants.

Our participants' performance in this HP-WHN test is quite poor, specifically in the case of gradual onset HP. This could mean that our task is quite difficult. Anyway, the task seems to be sensible and significant differences can be found between groups of participants of different driving experience. It can therefore be used as a discrimination tool. It would be interesting to compare this level of performance with the results found in other research. For instance, Underwood et al. (2013) also used gradual and abrupt onset videos and they obtained a better performance. However the their task requirements were quite different. Underwood's participants performed the traditional Hazard Perception task that measures Reaction Time and Accuracy. And the videos were displayed in their complete form. The task required participants to respond to any potential hazards by tapping the spacebar on the computer keyboard, as soon as they were detected. This meant the results could not really be compared.

On the other hand, the ability of the experienced participants had not reached a ceiling despite decades of driving. According to Horswill, Taylor, Newman, Wetton and Hill (2013), even highly experienced drivers benefit from a brief hazard perception training intervention. (i.e. police drivers significantly outperformed highly experienced drivers in a validated video-based hazard perception test).

It has been shown that drivers can improve their scores in video-based hazard perception tests following training interventions. Horswill, Falconer, Pachana, Wetton and Hill (2013) and Horswill, Taylor, Newnam, Wetton and Hill (2015) found that after training, significant improvements in hazard perception are seen for even highly experienced drivers and drivers over 65 years. Horswill et al. (2015) showed that the effect of brief training in Hazard Perception remains after the intervention, and approximately 1 month and 3 months later without a significant decay in the training effect over this time period. Future studies may also include a prospective design. For instance, it could be ascertained by monitoring the frequency of driving "mishaps/motor vehicle collisions" between trained and untrained groups for monthly follow-ups for 6 or 12 months. Moreover, including driving "mishaps" (e.g., failing to notice merging traffic) in addition to actual incidents would also provide more data.

In conclusion, it appears that PLTC, using video-based training tests, is effective: the active training improves hazard prediction. However, further research should explore whether this kind of training would be effective long term and whether transfer occurs in hazard prediction during real driving.

In addition, the trade-off between PLTC and other concurrent driving tasks should be evaluated in order to implement effective training methods that could be widely used and affordable to all those drivers who need to improve their hazard prediction abilities.

Future research could also look at the possibility of achieving a more affordable version of the *Hazard prediction Test* that would allow a more effortless and factual correction, using closed

questions with different response alternatives. Finally, we are looking forward to comparing the results obtained in this test with reaction time data and eye movement recordings.

6. Acknowledgements

We thank the Spanish drivers who were volunteers in the trials. We are grateful to the anonymous reviewers for their helpful comments to improve this paper. Financial support was also provided by the Junta de Andalucía (PO7-SEJ-02613, and Proyecto Motriz P11-SEJ-7404; and BS-14-164 I+D+I Research project from the CEI-Biotic, Granada, Spain and the Spanish Government, MICINN (PSI2013-42729-P). Ph.D. Psychology Program, University of Granada. Also, we would like to thank the Spanish Dirección General de Tráfico –DGT, (Exp: 0100DGT21263) for supporting the present research, and the Autoescuelas la Victoria, Luna and Genil (Granada) for providing the sample of participants. None of the funding sources had direct involvement in the study design, in data collection, analysis or interpretation, in the writing of reports or in the decision to submit the paper for publication.

7. References

- Beanland, V., Goode, N., Salmon, P.M., and Lenné, M.G. 2013. Is there a case for driver training? A review of the efficacy of pre- y post-licence driver training. *Safety Science*, 51(1), 127-137. Doi: 10.1016/ 2012.06.021
- Beirness, D.J., Simpson, H.M., and Mayhew, D.R., 1991. *Diagnostic assessment of problem drivers: review of factors associated with risky and problem driving. Report to Transport Canada TP11549E*. Road Safety and Motor Vehicle Regulation, Transport Canada, Ottawa.
- Box, E., and Wengraf, I. 2013. *Young driver safety: Solutions to an age-old problem*. Download from internet, on 6 September 2013. Available in: RAC foundation http://www.racfoundation.org/assets/rac_foundation/content/downloadables/young_driver_safety-box_wengraf-july2013.pdf
- Castro, C. Padilla, J.L., Benítez, I. García-Fernández, P. Estévez, B, López-Ramón, M.F., Roca, J., and Crundall, D., 2012. Hazard Perception HP. *Aproximación Inicial a la Tarea de Detección de Peligros durante la conducción en el contexto español*. Dirección General de Tráfico (DGT)(Exp. 0100DGT21263)
- Castro, C., Padilla, J. L., Roca, J., Benítez, I, García-Fernández, P., Estévez, B., López-Ramón, M.F., and Crundall, D. 2014. A first development and validation of the Spanish Hazard Perception Test. *Traffic Inj. Prev.* DOI:10.1080/15389588.2013.879125
- Cohen, J., 1988. *Statistical power analysis for the behavioural sciences*. Erlbaum, Hillsdale, NJ.

- Crundall, D., Andrews, B., Van Loon, E., Chapman, P., 2010. Commentary training improves responsiveness to hazards in a driving simulator. *Accid. Anal. Prev.* 42, 2117-2124.
- Crundall, D., Chapman, P., Trawley, S., Collins, L, Van Loon, E., Andrews, B., and Underwood, G., 2012. Some hazards are more attractive than others: Drivers of varying experience respond differently to different types of hazard. *Accid. Anal. Prev.* 45, 600-609.
- De Craen, S., Twisk, D.A.M., Hagenzieker, M.P., Elffers, H. y Brookhuis, K.A. (2011). Do you novice drivers overestimate their driving skills more than experienced drivers? Different methods lead to different conclusions? *Accid. Anal. Prev.* 43, 1660–1665.
- Dirección General de Tráfico, DGT, 2012. *Indicador rápido de accidentalidad en carretera. Año 2011*. Ministerio del Interior. Disponible en:
http://www.dgt.es/was6/portal/contenidos/documentos/seguridad_vial/estadistica/accidentes_24horas/resumen_anual_siniestralidad/resumen_siniestralidad039.pdf
- Endsley, M.R. 1995. Toward a theory of situation awareness in dynamic-systems. *Human Factors*, 37(1), 32-64.
- Fell, J.C., 1993. Repeat DWI offenders: their involvement in fatal crashes. In: Utzelmann, H-D., Berghaus, G., Kroj G. (Eds.), *12th International conference on Alcohol, Drugs and Traffic Safety-T92*. Cologne, Germany, National Highway Traffic Safety Administration.
- Fell, J.C., 1995. *Repeat DWI Offenders in the United States, Washington, DC*. National Highway Traffic Safety Administration, NHTSA Traffic Tech. No. 85.
- Fisher D.L., Pollatsek AP, and Pradhan A., 2006. Can novice drivers be trained to scan for information that will reduce their likelihood of a crash? *Injur. Prev.* 12, 25-29.
- Henderson, J.M., Brockmole, J. R., Castelhana, M. S., and Mack, M. L. 2007. Visual saliency does not account for eye movements during visual search in real-world scenes. In R. van Gompel, M. Fischer, W. Murray, & R. W. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 537–562). Amsterdam: Elsevier.
- Horswill, M.S., Taylor, K., Newnam, S., Wetton, M., and Hill, A. (2013). Even highly experienced drivers benefit from a brief hazard perception training intervention. *Accid Anal Prev*, 52, 100-110.
- Horswill, M.S., Falconer, E.K., Pachana, N.A., Wetton, M. and Hill, A. (2015). The long-term effects of a brief hazard perception training intervention in older drivers. *Psychol Aging*, 30(1), 62-77
- Horswill, M.S., Anstey, K.J., Hatherly, C., and Wood, J., 2010a. The crash involvement of older drivers is associated with their hazard perception latencies. *JINS* 16, 939-944.

- Horswill, M.S., Kemala, C. N., Wetton, M. A., Scialfa, C. T., and Pachini, N. A., 2010b. Improving Older Drivers' Hazard Perception Ability. *Psychol Aging*, 25, 464-469.
- Hwang, A.D., Higgins, E. C., and Pomplun, M. 2009. A model of top-down attentional control during visual search in complex scenes. *J. Vision*, 9 (5): 25, 1– 18.
- Isler, R.B., Starkey, N.J., and Williamson, A.R., 2008. Video-based road commentary training improves hazard perception of young drivers in a dual task. *Accid. Anal. Prev.* 41, 445-452.
- Jackson, L., Chapman, P., and Crundall, D., 2009. What happens next? Predicting other road users' behavior as a function of driving experience and processing time. *Ergonomics*, 52 2, 154-164.
- Laapotti, S., Keskinen, E., Hatakka, M., and Katila, A. 2001. Novice Drivers accidents and violations. A failure on higher or lower hierarchical levels of driving behavior. *Accid. Anal. Prev.*, 33, 759-769.
- Lapham, S.C., Ring-Kapitula, L. C' de la Baca, J., and McMillan, G.P. 2006. Impaired-driving recidivism among repeat offenders following an intensive court-based intervention. *Accid. Anal. Prev.* 38 (1), 162-169.
- Logan, G.D. 1988. Toward an instance theory of automatization. *Psychol. Rev.*, 95492–527.527.
- McKenna, F.P., Crick, J.L. 1997. *Developments in hazard perception*. Department of Transport, DOT. London, UK.
- McKenna, F.P., Horswill, M.S., and Alexander, J.L., 2006. Does anticipation training affect drivers' risk taking? *J. Exp. Psycho.*, 12 (1), 1–10.
- McKnight, A.J., and McKnight, A.S., (2003). Young Novice Drivers: Careless or Clueless? *Accid. Anal. Prev.*, 35 (6), 921-925.
- Meir, A., A., Borowsky, A. and Oron-Gilad, T., 2014. Formation and Evaluation of Act and Anticipate Hazard Perception Training AAHPT intervention for young-novice drivers. *Traffic Inj. Prev*, 15(2), 172-180.
- Pelz, D.C. and Krupat, E. 1974. Caution profile y driving record of undergraduate males. *Accid. Anal. Prev.*, 6, 45-58.
- Peters, R.J., and Itti, L. 2007. Beyond bottom-up: Incorporating task-dependent influences into a com-putational model of spatial attention. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2007)*, 1–8.
- Pollatsek, A., Fisher, D.L., and Pradhan, A.K., 2006. Identifying and Remediating Failures of Selective Attention in Younger Drivers. *Curr Dir Psychol Sci.* 15 (5), 255-259.
- Pomplun, M. 2006. Saccadic selectivity in complex visual search displays. *Vision Res.*, 46, 1886–1900.

- Richard, C., Campbell, J.L., Brown, J.L. Lichty, M.G., Chrysler, S.T. and Atkins, R. 2013. Investigating Speeding Behavior with Naturalistic. Methodological Lessons Learned. *Transportation Research Record: Journal of the Transportation Research Board*, 2365, Transportation Research Board of the National Academies, Washington, D.C., 58–65
- Richardson, J.T.E. 2011. Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review*. 6, 135–147.
- Simon, F. and Corbett, C. 1996. Road traffic offending, stress, age and accident history among male and female drivers. *Ergonomics*, 39 (5), 757-780.
- Solnick, S.J., and Hemenway, D., 1994. Hit the bottle and run: the role of alcohol in hit-and-run pedestrian fatalities. *J. Stud. Alcohol*, 55, 679–684.
- Underwood, G., Ngai, A., and Underwood, J., (2013). Driving experience and situation awareness in hazard detection. *Saf. Sci.* 56, 29-35.
- Vargas, C., Moreno-Rios, S., Castro, C., and Underwood, G. (2011). Encoding time y signs exposure time in the representation of diagrammatic deontic meanings. *Acta Psychol.*, 137, 106-114.
- Wang, Y.B., Zhang, W., and Salvendy, G.A. 2010. A comparative Study of Two Hazard Handling Training Methods for Novice Drivers. *Traffic Inj. Prev.* 11, 483-491.
- Walker, G.H., Stanton, N.A., Kazi, T.A., Salmon, P.M., and Jenkins, D.P.: Does advanced driver training improve situation awareness? *Applied Erg.*40(4), 678–687 (2009)
- Wells, P., Tong, S., Sexton, B., Grayson, G., and Jones, E., 2008. *Cohort II: A Study of Learner and New Drivers*. Department of Transport, DOT. London, UK.
- Wetton, M.A., Hill, A. and Horswill, M.S. 2011. The development y validation of a hazard perception test for use in driver licensing. *Accid. Anal. Prev.*, 43(5), 1759-1770.
- Wetton, M.A., Hill, A., and Horswill, M.S., 2013. Are What Happen Next exercises and self-generated commentaries useful additions to hazard perception training for novice drivers? *Accid. Anal. Prev.* 54, 57-66.
- White, M., Cunningham, L.C. Titchener, K. (2011). Young drivers’ optimismo bias for accident risk and driving Skull: Accountability and insight experience manipulations. *Accid. Anal. Prev.* 43, 1309-1315.
- Wolfe, J. M. 1994. Guided Search 2.0: A revised model of guided search. *Psychon. Bull. Rev.*, 1 , 202–238.
- Wolfe, J.M. and Horowitz, T.S. 2004. What attributes guide the deployment of visual attention and how do they do it? *Nat. Rev. Neurosci.*, 5, 495-501.
- Yahya, M. and Hammarström, U. 2011. The offender’s drink-driving as accident cause in traffic – An analysis of data in the register of legal proceedings. VTI. Rapport 710

Young, A.H., Chapman, P., and Crundall, D. (2014). Producing a commentary slows concurrent hazard prediction responses. *Journal of Experimental Psychology: Applied*, 20(3), 285-294.

Zelinsky, G. J., Zhang, W., Yu, B., Chen, X., and Samaras, D. 2006. The role of top-down and bottom-up processes in guiding eye movements during visual search. In Y. Weiss, B. Scholkopf, and J. Platt (Eds.), *Advances in neural information processing systems* (vol. 18, pp. 1569–1576). Cambridge, MA: MIT Press.

Table 1. Socio-demographic information for the three experience groups of drivers (learner, novice and experienced drivers) non recidivist drivers and (novice and experienced drivers) multiple offenders.

Table 1. Socio-demographic information for the three experience groups of drivers (learner, novice and experienced drivers) non recidivist drivers and (novice and experienced drivers) multiple offenders.

Non recidivist drivers		Learner drivers					Novice drivers					Experienced drivers				
Socio-demographic information		N	Min	Max	Mean	SD	N	Min	Max	Mean	SD	N	Min	Max	Mean	SD
Age		20	18	37	20,30	4,231	41	18	29	21,93	2,611	20	27	53	37,10	8,372
Gender ^a		18	1	2	2	--	41	1	2	2	--	20	1	2	1	--
Level of education ^b		20	3	6	4	--	41	3	6	4	--	19	2	6	4	--
Years driving regularly ^c		20	1	1	1	--	41	2	3	2	--	20	4	4	4	--
Years since obtaining driving license		5	0	4	1,00	1,732	40	0	10	3,55	2,396	20	11	37	20,35	7,849
Driving frequency ^d		16	1	5	5	--	41	1	5	2	--	20	1	2	1	--
Kilometres driven last 12 months		10	0	1000	280,20	435,690	35	0	60000	4922,06	10956,898	20	150	100000	35087,50	34305,385
Accidents last 12 months		13	,00	1,00	,0769	,277	37	,00	2,00	,189	,518	20	,00	2,00	,2000	,523
Quasi-accidents last 12 months		15	0	3	,60	,986	41	0	20	2,37	3,839	20	0	26	7,40	7,465
Traffic incidents- Insurance company		16	0	0	,00	,000	41	0	3	,17	,543	20	0	2	,30	,571
N° of times losing driving license		10	0	0	,00	,000	41	0	1	,02	,156	20	0	0	,00	,000
Traffic tickets received		14	0	0	,00	,000	41	0	2	,10	,374	20	0	2	,35	,671
Multiple Offenders																
Age		--	--	--	--	--	19	21	39	25,32	4,498	20	26	51	38,70	8,498
Gender ^a		--	--	--	--	--	20	1	2	1	--	20	1	2	1	--
Level of education ^b		--	--	--	--	--	20	1	6	3	--	20	1	6	3	--
Years driving regularly ^c		--	--	--	--	--	20	3	3	3	--	20	4	4	4	--
Years since obtaining driving license		--	--	--	--	--	19	4	12	6,68	2,110	20	10	33	20,55	8,624
Driving frequency ^d		--	--	--	--	--	20	1	5	1	--	19	1	5	1	--
Kilometres driven last 12 months		--	--	--	--	--	14	0	150000	27310,71	41096,080	19	0	130000	20489,53	32467,944
Accidents last 12 months		--	--	--	--	--	18	,00	2,00	,2778	,57451	18	,00	2,00	,1667	,51450
Quasi-accidents last 12 months		--	--	--	--	--	18	1	17	5,17	4,605	17	0	7	2,59	1,698
Traffic incidents- Insurance company		--	--	--	--	--	17	0	2	,47	,800	17	0	2	,18	,529
N° of times losing driving license		--	--	--	--	--	19	0	2	1,00	,577	18	0	2	,94	,539
Traffic tickets received		--	--	--	--	--	19	0	3	1,42	1,216	18	0	3	1,11	1,079

Median valued reported:

(a) 1 = Female. 2 = Male. Median value reported.

(b) 1 = Primary. 2 = Secondary (compulsory). 3 = Secondary (non-compulsory). 4 = Vocational. 5 = Grade. 6 =

Master.

(c) 1 = Learning to drive. 2 = Up to 2 years. 3 = Between 3-7 years. 4 = 8 or more years. Median value reported.

(d) 1 = Daily. 2 = Weekly. 3 = Monthly. 4 =

Never.

Table 2. Descriptive Statistic of the items in the new Hazard prediction Test.

Video	Mean	Standard Deviation	Discrimination Index
1	2.75	1.61	0.62
5	3.61	1.41	0.39
6	3.25	1.21	0.44
8	1.02	1.75	0.46
10	0.47	1.08	0.40
11	3.80	1.55	0.42
13	1.66	2.05	0.38
14	3.49	1.08	0.43
15	2.74	1.60	0.50
16	2.50	1.73	0.36
17	4.64	1.39	0.75
18	4.36	1.51	0.58
19	3.53	1.61	0.53
20	2.80	1.96	0.46
21	2.83	1.98	0.55
22	3.37	1.46	0.53
23	3.05	1.89	0.47
27	4.04	1.62	0.73
30	2.97	1.17	0.32
31	3.22	1.52	0.45
Total	3.00	1.56	

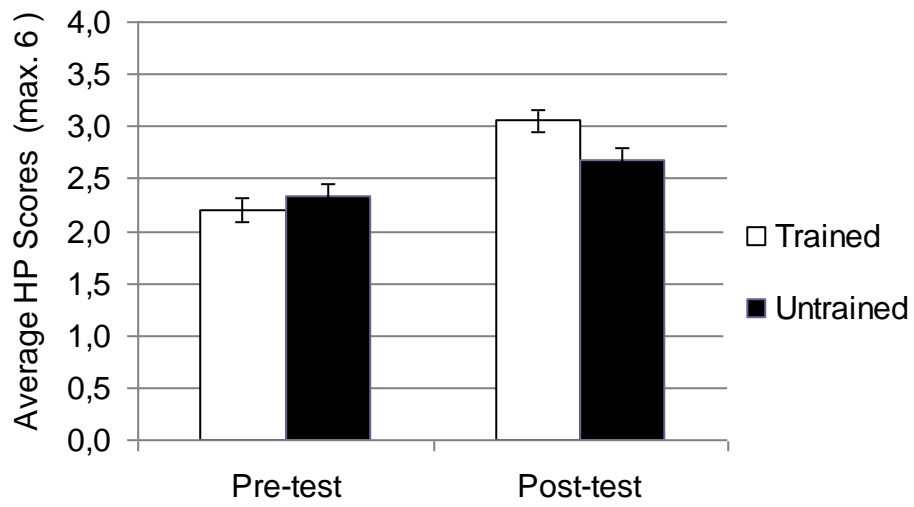


Figure 1. Mean hazard prediction scores (the scale was composed of six points) pre-test and post-test for the trained and untrained groups. Error bars represent standard errors of the mean

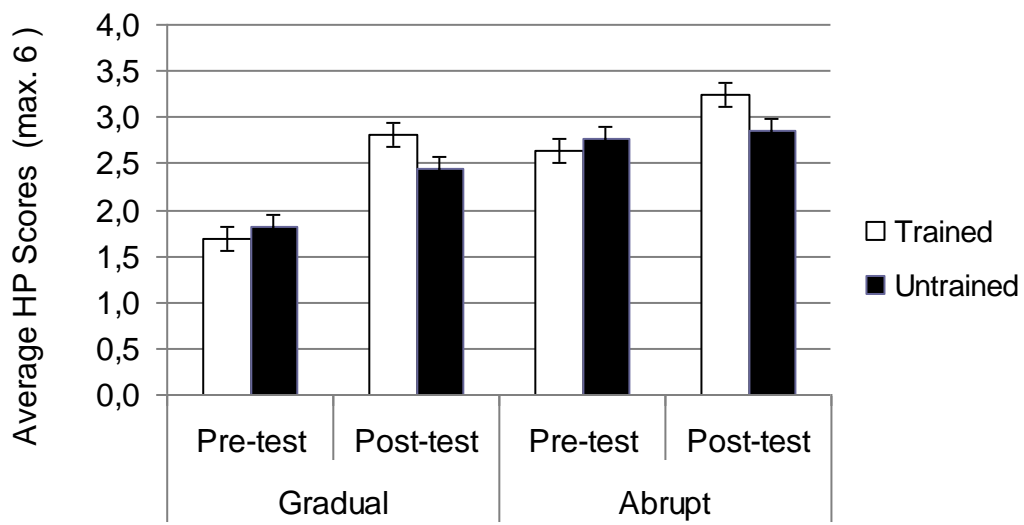


Figure 2. Mean gradual and abrupt-onset hazard events (the scale was composed of six points) pre-test and post-test for the trained and untrained groups. Error bars represent standard errors of the mean

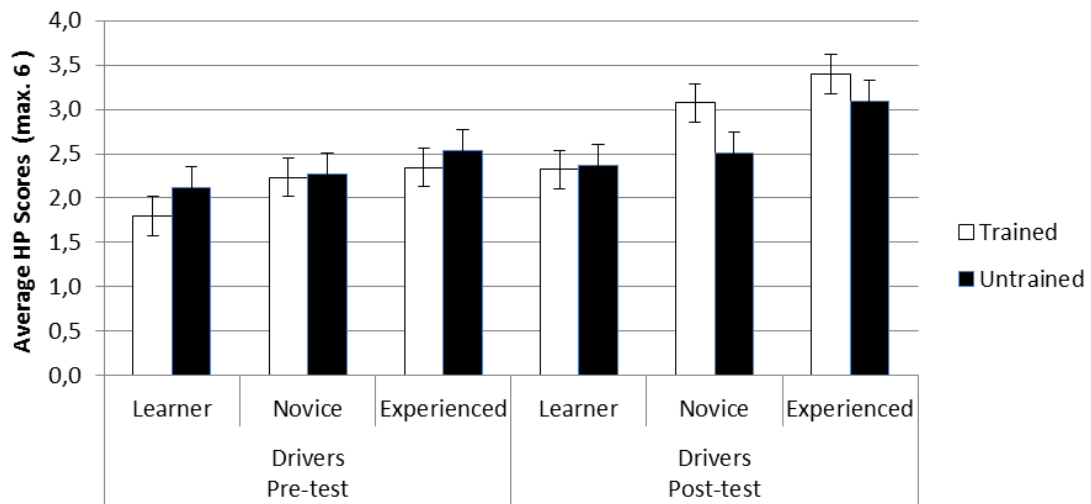


Figure 3. Mean hazard prediction scores (the scale was composed of six points) of learner, novice and experienced drivers in the pre-test and post-test for the trained and untrained groups. Error bars represent standard errors of the mean.

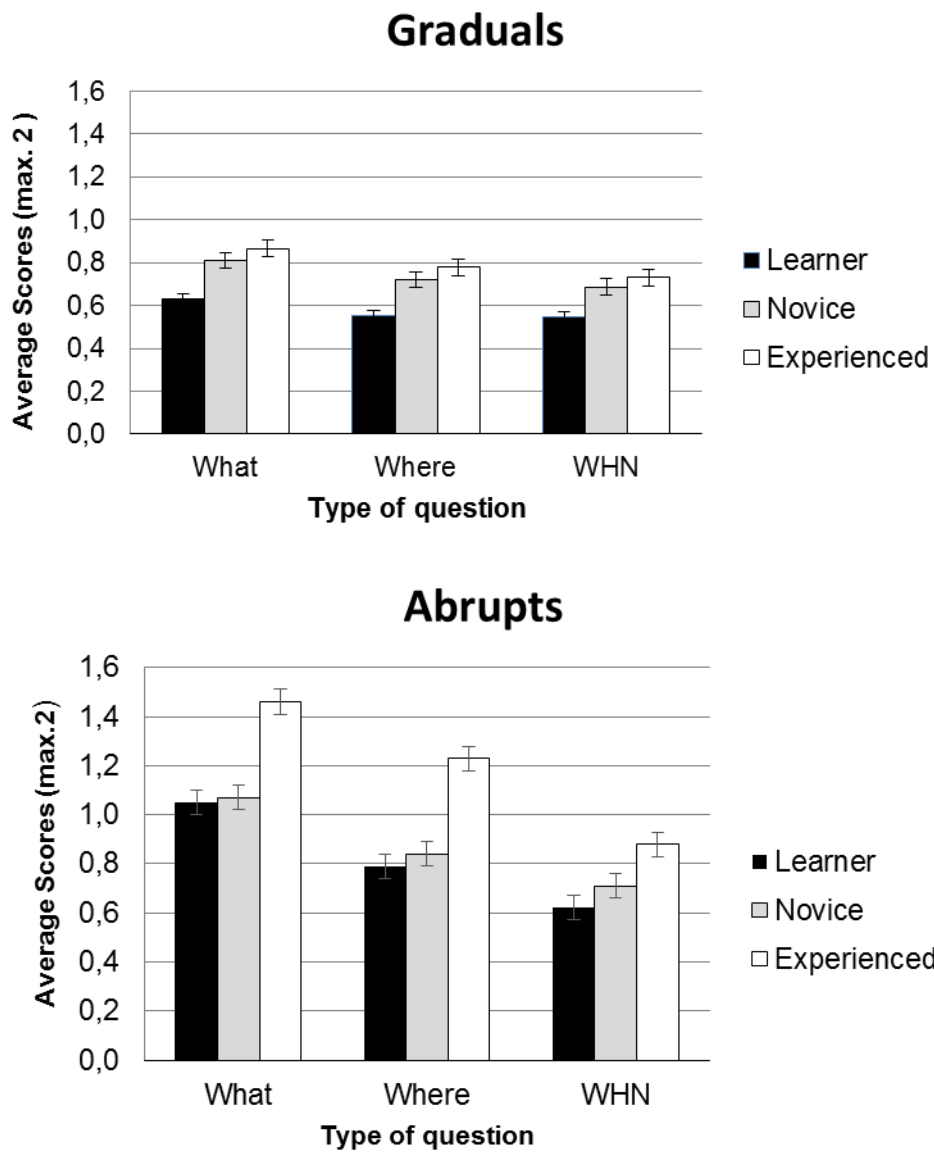


Figure 4. (Top/Down) Mean hazard prediction scores of gradual and abrupt onset hazards (average max. 2) per question, for learner, novice and experienced drivers. Error bars represent standard errors of the mean.