Detection or Appraisal – Do Their Eye Movements Reveal What Causes Novices' Poor Performance in a Dynamic Hazard Perception Test?

Sarah Malone and Roland Brünken

Department of Education
Saarland University
Saarbrücken, Germany

ABSTRACT

According to Grayson et al. , risk behavior in driving consists of *hazard detection*, *threat appraisal*, *action selection* and *implementation*. Hazard perception tests often include the task to react quickly to hazards within traffic scenarios. Thus, two components of risk behavior are included in one measure and therefore confounded: hazard detection and threat appraisal. Tracking the eye movements, researchers found evidence for novices having deficits regarding hazard detection . In contrast, Hystegge et al. revealed, that novice drivers were as fast as expert drivers in looking at still hazards but needed more time to evaluate them. The aim of the present eye tracking experiment was to investigate, whether experienced drivers outperform novices with regard to hazard detection or threat appraisal. 22 experienced drivers and 15 learner drivers were presented 32 animated traffic scenarios in a computer based hazard perception test. The depended variables were accuracy and speed of hazard detection (first fixation) and threat appraisal (reaction after detection). Experts outperformed novices clearly in hazard detection: They focused on more hazards and detected them faster than the novices. Moreover, after having detected a hazard, experts react to it more reliable but not faster than the novices.

Keywords: hazard perception, novice drivers, eye tracking

INTRODUCTION

Although the overall number of fatal traffic accidents is decreasing in most countries, novice drivers still have a very high risk to die in traffic crashes and higher near-crash rates compared to more experienced drivers . Fortunately, driving expertise seems to evolve relatively fast. After two years of solo driving, the drivers reach a stable low risk level . Nevertheless, the main purpose of traffic safety research is to identify quantifiable aspects of the driving skill that can explain the high accident liability of novice drivers . Understanding the novice drivers' main deficits that cause their high accident risk and having identified possibilities to assess related skills appropriately can help to improve driving education on the one hand and driving assessment on the other hand. One skill that has been already identified as a relevant and quantifiable component is hazard perception. Hazard perception refers to the driver's quick reaction to emerging hazards within traffic situations and is therefore related to the response to risks in traffic. According to Grayson et al. (2003), risk behavior consists of four separate components: *hazard detection*, *threat appraisal*, *action selection* and *implementation*. Computer based assessment of hazard perception often includes dynamic visualizations of traffic scenarios and the task to react as soon as possible to emerging hazards . Thus, only two of the four components of risk behavior are assessed: hazard detection and threat appraisal. Furthermore, both components are included in one measure and therefore confounded. In general, experienced drivers outperform inexperienced drivers in hazard perception tests concerning accuracy (e.g. McKenna & Crick, 1994) and speed of reaction (e.g. Quimby & Watts, 1981), but the assessment cannot definitely reveal the reason for the novices' failure. Eye tracking experiments can help to understand the novices' deficits by separating the hazard detection process from the threat appraisal process clearly.

THEORETICAL AND EMPIRICAL BACKGROUND

Hazard perception and its assessment

According to McKenna and Crick , hazard perception is the driver's ability to identify emerging hazards and potential dangerous situations within traffic scenarios. It is one of the most frequently investigated driving related skill. The popularity of this concept in traffic safety research is due to the simplicity of its quantification by the means of hazard perception tests. Hazard perception tests usually consist of several traffic scenarios that are presented one by one to the participants, whose reactions to appearing hazards are recorded .

Comparisons between experienced and novice drivers revealed that the experts were able to identify more hazards accurately than the novice drivers . Novices miss mainly latent or hidden hazards . Moreover, experts outperformed novices concerning the average time needed to react to emerging hazards . With a few exceptions, when experts outperformed novices in only a few scenarios , hazard perception tests can be regarded as valid measures of driving expertise.

Among others, existing hazard perception tests differ with respect to the presentation mode of the displayed traffic scenarios (static vs. animated) and the categorization of the included hazards. This leads to the question, how hazard perception tests should be designed to provide the best measures of driving expertise. Traffic scenarios can be presented either in a dynamic way as videos of real traffic situations or computer animations or in a static way as still pictures . Two studies were conducted in order to compare the two presentation modes for hazard perception assessment , finding no clear advantage of one presentation mode. Malone (2012) expected the integration of dynamic instead of static scenarios to enhance the validity of a hazard perception test, because animated traffic scenarios should be more realistic than static presentations and therefore, more specific for driving. As a stable finding of expertise research is that experts outperform novices more clearly when the task is specific for the respective domain , it was deduced that the difference between driving experts and novices would be bigger if dynamic instead of static traffic scenarios are included in a hazard perception test. Contrary to the expectations, the inclusion of dynamic presentations impeded the reaction task for both, experts and novices, whereas their performance difference was nearly the same for both presentations modes. The inclusion of dynamic instead of static traffic scenarios in hazard perception assessment increases the difficulty of the test, which could be due to the animations' transience.

The application of animated traffic scenarios did not lead to higher differences in the overall hazard perception performance between experts and novices. However, it is possible, that the two presentation modes differ in their ability to support the assessment of different skills underlying overall hazard perception test performance. There are several possible reasons for the relatively poor performance of novice drivers in hazard perception tests. Grayson et al. postulate a four component model of risk behavior consisting of hazard detection, threat appraisal, action selection and implementation. Errors can occur on every stage, but action selection and implementation are not measured typical hazard perception tests. That means these components cannot account for the novices‘ low performance in these tests. The two remaining components could account for hazard perception performance, but they are usually confounded in hazard perception tests: If the participant reacts to a hazard he must have seen and appraised it already. Therefore, usual hazard perception assessment alone cannot differentiate between different deficits that lead to the novices' poor performance.

Uncovering novice drivers' deficits – The potential of eye tracking

Analyzing people's eye movements during real or simulated driving or watching traffic scenarios, can reveal deficits relating to searching the road for safety relevant information . Eye tracking studies, including dynamic traffic situations, revealed that novices might have problems with hazard detection. Watching traffic scenarios, novices display inadequate visual search: They look less far ahead and less often in the driving mirrors . Besides, beginner drivers display a low variance of horizontal search and don't seem to adapt their visual search to changing traffic conditions . In contrast to these findings, Huestegge et al. (2010) found no evidence for novice drivers having deficits in hazard detection. In the context of hazard perception assessment, the authors used eye tracking to separate the two components hazard detection and threat appraisal. In the experiment, a hazard perception test with static pictures of traffic scenarios was used. The participants were supposed to push the space bar of a computer keyboard as soon as they became aware of a hazard on the displayed photo. During the test, the participants' eye movements were recorded. Hazard detection was operationalized as the time interval between the onset of a traffic scenario and the participant's first fixation on the included hazard. Threat appraisal time was defined by the interval between the first fixation on the hazard and the reaction by the participant.

In contrast to the findings of previous research using eye tracking in the traffic context, Huestegge et al. found that in their static hazard perception test, novices were as fast as experts to fixate the hazards but needed more time to react to them, which the authors interpreted as a deficit on the threat appraisal stage.

HYPOTHESES

Hazard Perception tests show, that beginner drivers obviously lack important skills to accomplish the task as good as experienced drivers. Previous eye tracking findings concerning the reasons for the novices' failure are not consistent. Therefore, the general research intention was to investigate, what accounts for the novices' low performance in a dynamic hazard perception test: hazard detection or hazard appraisal?

According to previous research in the field of hazard perception it was assumed, that experts identify more hazards and react faster to them than novices in a dynamic hazard perception test (hazard perception performance). Eye tracking research in real traffic or using dynamic presentations of traffic scenarios has revealed that experienced drivers exceed beginner drivers in scanning and searching the road. Therefore, it was assumed that experts fixate more hazards and fixate the hazards earlier than novices in a dynamic hazard perception test (hazard detection performance). As it was shown in a static hazard perception test, experts outperform novices in hazard appraisal, too. For the present study it was deduced from this finding, that having detected a hazard, experts react to it more reliable and faster than novices in a dynamic hazard perception test (hazard appraisal).

METHOD

The sample consisted of 22 experienced drivers (experts) and 15 learner drivers (novices). The experts had on average 6.53 years (*SD* = 2.48), but at least two years experience in solo driving. 43 % of the participants were male.

The study design was a 2 by 3 mixed design with the between subjects factor expertise (experts vs. novices) and the within subjects factor hazard type (implicit hazard vs. explicit hazard vs. distractor). The dependent variables were *overall performance in a dynamic hazard perception test* (accuracy and speed), *hazard detection* (accuracy and speed), which was deduced from the participants eye movements during the test and hazard appraisal (accuracy and speed). Data for the latter was inferred from eye movements and final reactions of the participants after having detected a hazard.

The computer based experiment started with a questionnaire about demographical data, driving habits and driving education. After that, the participants took a dynamic hazard perception test presented on a computer screen. A Tobii 300 eye tracker was employed for observing the eye movements during task presentation. Each participant completed a computer-based reaction time task that consisted of 32 items, that had been created in order to assess the accuracy and speed of reactions to different traffic conditions. One of the items, always presented first after the instruction, served as an instructional example; The participants' performance on this item was not included in their overall test score. The scenarios were developed by the means of the Vicom Editor© – a software, which offers the creation of animated traffic scenarios. The participants were instructed to scan the presented scenarios for hints that indicate a reason for the ego-vehicle to slow down or brake. Immediately after the detection of such a hint, the space bar of the keyboard had to be pressed once. Each scenario contained maximum one stimulus that emphasized the need to slow down.

Concerning the kind of the applied traffic situations, three different types were employed: distractors (no hazard), implicit hazards and explicit hazards. Within the seven distractor items, the participants were expected not to react at all. The distractor items had different scopes: two scenarios from each of three different road types (freeway, rural road, urban road). All these scenarios had in common that there were rather few other road users included and that the environment was sparsely-built.

Seven further scenarios included one implicit hazard each. The participants were to react to the relevant hints by pressing the spacebar as soon as they became aware of them. The hints which were shown in these situations did not pose a real risk but indicated the need to slow down because of current traffic rules. Noncompliance to reduce ones speed in these situations won’t lead to a traffic accident in any case, but it definitely increases the crash risk.

An explicit hazard was included in each of another 17 scenarios. The respective traffic situations can be considered as dangerous, because missing the relevant hints and therefore, not slowing down, may automatically lead to a crash. In Figure 1 a frame from a traffic scenario with an explicit hazard is displayed. The ego-vehicle is driving on the left lane while another car merges from the right lane. The ego-vehicle is supposed to slow down, because the brake lights of the vehicle ahead indicate slow or stopped traffic ahead.

Figure 1. Computer-generated image of traffic situation. The image shows an example for explicit hazard.

Results

First of all, the results for the overall performance in the hazard perception tests are reported. Descriptive results for the variables accuracy and speed are provided in Table 1, broken down by expertise group and hazard type. Accuracy is operationalized as proportion of correct answers (hits and correct rejections) in relation to the total number of items. Speed is defined as the time, between the first appearance of a hazard and the first reaction by the participant. The response times were converted in standardized z-values.

Regarding the dependent variable accuracy, the ANOVA revealed a main effect of expertise ($F(1, 35) = 31.18$; $p < .001$; $\eta_p^2 = .47$). Also, a significant main effect of hazard type could be found ($F(2, 70) = 141.51$; $p < .001$; $\eta_p^2 = .80$): The distractors were less difficult than the other two item types.

Table 1: Results for overall performance in hazard perception (accuracy and speed)

Expertise	Hazard type	HP performance accuracy		HP performance speed		
		M	*SD*	*M*	*SD*	*n*
experts	distractor	.87	.02	/	/	22
	implicit	.55	.20	-.15	.11	22
	explicit	.53	.12	-.2	.07	22
	overall	.65	.02	-.19	.07	22
novices	distractor	.85	.04	/	/	15
	implicit	.26	.20	.37	.14	15
	explicit	.35	.12	.44	.09	15
	overall	.49	.02	.04	.09	15
overall	distractor	.86	.03	/	/	37
	implicit	.43	.25	.11	.09	37
	explicit	.53	.12	.11	.06	37
	overall	.57	.02	.11	.06	37

A significant interaction effect of expertise and hazard type ($F(2, 70) = 10.49$; $p < .001$; $\eta_p^2 = .23$) was found for the dependent variable accuracy, which indicates, that expert-novice differences exist for the two item types that include hazards but not for the distractors. The interaction effect is visualized in Figure 2.

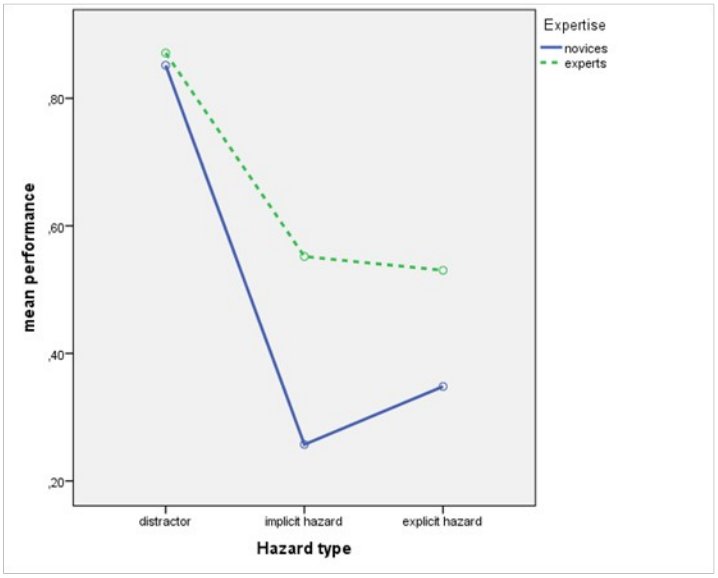

Figure 2. Interaction effect between the factors expertise and hazard type for the dependent variable accuracy (hazard perception performance)

Moreover, the ANOVA revealed a main effect of expertise for the dependent variable speed in the overall hazard perception performance ($F(1, 35) = 28.56$; $p < .001$; $\eta_p^2 = .47$). The experienced drivers reacted faster to emerging hazards than the learner drivers. However, a significant main effect of hazard type could not be found ($p \geq .05$). The interaction between expertise and hazard type was not significant for the dependent variable speed ($p \geq .05$).

The descriptive results for performance in hazard detection are shown in Table 2. By means of eye tracking it was recorded whether the participants fixated the relevant hazards (hazard detection accuracy) at least for once within a critical timeframe and how fast they were in detecting them after they had appeared on screen (hazard detection speed). A main effect of the factor expertise was found, indicating that the experts outperformed the novices in accuracy ($F(1, 35) = 11.44$; $p < .01$; $\eta_p^2 = .25$).The main effect of the factor hazard type for this variable indicates, that more explicit hazards were detected than implicit hazards ($F(1, 35) = 25.10$; $p < .001$; $\eta_p^2 = .42$). The ANOVA revealed also an interaction effect of the factors expertise and hazard type ($F(1, 35) = 4.34$; $p < .05$; $\eta_p^2 = .11$). The difference between the experts and the novices was more obvious for items with implicit hazards than for items with explicit hazards (see Figure 3).

Table 2: Results for performance in hazard detection (accuracy and speed)

Expertise	Hazard type	HP detection accuracy		HP detection speed		
		M	*SD*	*M*	*SD*	*n*
experts	implicit	.66	.04	-.10	.41	22
	explicit	.74	.03	-.03	.58	22
	overall	.70	.03	-.07	.09	22
novices	implicit	.42	.05	.32	.45	15
	explicit	.62	.04	.31	.48	15
	overall	.52	.04	.31	.11	15
overall	implicit	.54	.03	.11	.07	37
	explicit	.68	.03	.14	.09	37
	overall	.61	.03	.12	.07	37

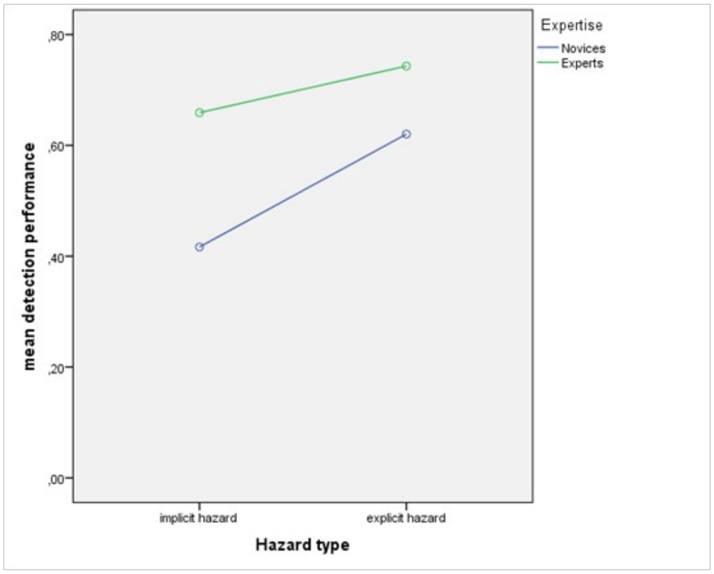

Figure 3. Interaction effect between the factors expertise and hazard type for the dependent variable accuracy (hazard detection performance)

The ANOVA revealed a main effect of expertise for the dependent variable speed in hazard detection (F(1, 35) = 6.64; $p < .05$; $\eta_p^2 = .16$). On average, the experts were faster than the novices in detecting the hazards. No significant main effect of hazard type could be found ($p \geq .05$). The interaction between expertise and hazard type was not statistically significant ($p \geq .05$).

The descriptive results for hazard appraisal are displayed in Table 3. A main effect oft the factor expertise for the dependent variable accuracy could be found (F(1, 35) = 17,66; $p < .001$; $\eta_p^2 = .34$). If the novices looked as a hazard they less often recognized it as a hazard than the experts. No further significant main effects or interactions for both dependent variables could be found (all $p \geq .05$).

Table 3: Results for performance in hazard appraisal (accuracy and speed)

Expertise	**Hazard type**	**HP appraisal accuracy (min. 1, max. 2)**		**HP appraisal speed**		
		M	*SD*	*M*	*SD*	*n*
experts	implicit	1.59	.28	-.01	.50	22
	explicit	1.55	.16	-.01	.40	22
	overall	1.57	.21	-.01	.45	22
novices	implicit	1.26	.29	-.22	.53	15
	explicit	1.38	.19	-.07	.70	15
	overall	1.32	.24	-.14	.61	15
overall	implicit	1.46	.32	-.06	.50	37
	explicit	1.48	.19	-.05	.48	37
	overall	1.47	.25	-.06	.49	37

CONCLUSIONS

Experts outperformed novices in the developed dynamic hazard perception test. Experts identified more hazards than novices and reacted faster to them than the novices. Therefore, the test, with the different types of scenarios and hazards, can in general be seen as a valid measure of driving competence.

As intended, the experiment allowed to separate two components of risk behavior that are usually confounded in hazard perception tests: hazard detection and threat appraisal. The results from eye tracking indicate that experts outperformed novices in hazard detection: Experts focused on more hazards and focused on the hazards faster than novices. This result matches the findings of previous eye tracking studies using dynamic traffic scenarios, that revealed different search patterns for experts and novices . This result stands in contrast to the findings of Huestegge et al. , who found no differences between expert and novice hazard detection performance in their static hazard perception test. It is possible, that the presentation mode of a hazard perception test determines whether hazard detection accounts for performance or not. An appropriate visual search strategy and consequently, a quick detection of a traffic hazard seems to be only relevant if dynamic traffic scenarios are displayed.

Experts outperformed novices also in hazard appraisal: After having detected a hazard, experts reacted to it more reliable but not faster than novices. With respect to the underlying model of risk behavior , the results indicate, that both components of risk behavior, hazard detection and threat appraisal, account for the performance in dynamic hazard perception tests.

The results of the present experiment can have implications for driver training as well as for driving assessment. The results from eye tracking deliver important information about the underlying skills that cause performance differences in a hazard perception test. Novices are inferior to experts in hazard detection and hazard appraisal. Therefore, driving education should provide more specific training of hazard detection and appraisal.

From a diagnostic point of view, the results of the present experiment prove the advantage of using dynamic instead of static scenarios in hazard perception tests. In contrast to the static hazard perception test used by Huestegge et al. , the dynamic hazard perception test revealed expertise related differences in hazard detection. In real traffic, detecting a hazard in time is the first required step in crash avoidance. If a hazard is detected too late or completely missed, all of the other steps wouldn’t be initiated at all. This leads to the conclusion that dynamic hazard perception tests are more valid than static hazard perception tests and therefore, more appropriate for driving assessment.

Limitations of the present study can be seen with respect to the lack of an experimental variation of the presentation mode (dynamic vs. static). Only applying exactly the same traffic scenarios in a dynamic and a static version allows to compare the two presentation modes.

REFERENCES

Biermann, A., Skottke, E.-M., Anders, S., Brünken, R., Debus, G., & Leutner, D. (2008). Entwicklung und Überprüfung eines Wirkungsmodells: Eine Quer- und Längsschnittstudie. In R. Brünken, G. Debus & D. Leutner (Eds.), *Wirkungsanalyse und Bewertung der neuen Regelungen im Rahmen der Fahrerlaubnis auf Probe (Berichte der BASt, Heft M 194)* (pp. 46-111). Bremerhaven: Wirtschaftsverlag NW.

Borowsky, A., Shinar, D., & Oron-Gilad, T. (2010). Age, skill, and hazard perception in driving. *Accident Analysis and Prevention, 42*(4), 1240-1249.

Crundall, D., Chapman, P., Trawley, S., Collins, L., van Loon, E., Andrews, B., & Underwood, G. (2012). Some hazards are more attractive than others: Drivers of varying experience respond differently to different types of hazard. *Accident Analysis and Prevention, 45*(0), 600-609.

Crundall, D., Underwood, G., & Chapman, P. R. (1998). How much do novice drivers see? The effects of demand on visual search strategies in novice and experienced drivers. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 395-417). Oxford: Elsevier.

Glaser, R., & Chi, M. T. H. (1988). Overview. In M. T. H. Chi, R. Glaser & M. J. Farr (Eds.), *The nature of expertise* (pp. xv-xxviii). Hillsdale: NJ: Erlbaum.

Grayson, G. B., Maycock, G., Groeger, J. A., Hammond, S. M., & Field, D. T. (2003). Risk, hazard perception and perceived control. Crowthorne, Berkshire: TRL Limited.

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: Theory and application* (pp. 155-175). Aldershot: Ashgate.

Huestegge, L., Skottke, E.-M., Anders, S., Muesseler, J., & Debus, G. (2010). The development of hazard perception: Dissociation of visual orientation and hazard perception. *Transportation Research, 13F*, 1-8.

Lee, S. E., Simons-Morton, B. G., Klauer, S. E., Ouimet, M. C., & Dingus, T. A. (2011). Naturalistic assessment of novice teenage crash experience. *Accident Analysis & Prevention, 43*(4), 1472-1479.

Malone, S. (2012). *Computerbasierte Messung von Teilaspekten der Fahrkompetenz. Besonderheiten des Expertiseerwerbs beim Autofahren.* Universität des Saarlandes, Saarbrücken.

McKenna, F. P., & Crick, J. L. (1994). Hazard perception in drivers: A methodology for testing and training. Crowhtorne, Berkshire: TRL Limited.

McKnight, A. J., & McKnight, S. A. (2003). Young Novice Drivers: Careless or Clueless. *Accident Analysis and Prevention, 35*, 921-925.

Mourant, R. R., & Rockwell, T. H. (1972). Strategies of visual search by novice and experienced drivers. *Human Factors, 14*(4), 325-335.

Pollatsek, A., Narayanaan, V., Pradhan, A. K., & Fisher, D. L. (2006). Using eye movements to evaluate a pc-based risk awareness and perception training program on a driving simulator. *Human Factors, 48*, 447-464.

Pradhan, A. K., Hammel, K. R., DeRamus, R., Pollatsek, A., Noyce, D. A., & Fisher, D. L. (2005). The use of eye movements to evaluate the effects of driver age on risk perception in an advanced driving simulator. *Human Factors, 47*, 840-852.

Quimby, A. R., & Watts, G. R. (1981). Human factors and driving performance. Crowthorne, Berkshire: Transport and Road Research Laboratory.

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

Schade, F. D. (2000). Verkehrsauffälligkeit von Pkw-Fahrern und ihre Entwicklung mit dem Lebensalter – ein Modell. *Zeitschrift für Verkehrssicherheit, 46*, 9-18.

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

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 and Prevention, 45*(0), 547-553.

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 and Prevention, 43*(1), 204-208.

Smith, S. S., Horswill, M. S., Chambers, B., & Wetton, M. A. (2009). Hazard perception in novice and experienced drivers: The effects of sleepiness. *Accident Analysis & Prevention, 41*(4), 729-733.

Sümer, N., Berfu, Ü. A., & Birdal, A. (2007). *Assessment of Hazard Perception latencies using real life and animated traffic hazards: comparison of novice and experienced drivers.* Paper presented at the Fourth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, Stevenson, Washington.

TÜV | DEKRA arge tp 21. (latest update: 24.09.2008). Vicom Editor edition 0.7.1.

Underwood, G. (2007). Visual attention and the transition from novice to advanced drivers. *Ergonomics, 50*(8), 1235-1249.

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