**Document Version**
Final published version

**Citation (APA)**
Kuipers, J., Mulder, M., & Voskes, M. (2026). Training for the unexpected: Enhancing driver preparedness through hazard awareness. A 15-year cohort study. *Accident Analysis and Prevention*, *232*, Article 108536. https://doi.org/10.1016/j.aap.2026.108536

Accident Analysis and Prevention 232 (2026) 108536

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

# Training for the unexpected: Enhancing driver preparedness through hazard awareness. A 15-year cohort study

Jorrit Kuipers [a,b,c,*], Max Mulder [b], Maran Voskes [d]

[a] Green Dino BV, the Netherlands
[b] Department Control & Operations, Faculty of Aerospace Engineering, Delft University of Technology, the Netherlands
[c] Department of Learning, Data Analytics and Technology, Faculty of Behavioural, Management and Social Sciences, University of Twente, the Netherlands
[d] Saxion University of Applied Sciences, the Netherlands

## ABSTRACT

This cohort study investigates the long-term effects of simulator-based hazard awareness training (HAT) on learner and novice drivers in the Netherlands, using a dataset of 2,372 participants over a 15-year period. Most prior studies on HAT have measured only immediate post-training outcomes; no longitudinal cohort study with a control group has previously examined both supervised and unsupervised driving outcomes over a multi-year horizon. Although the HAT and control groups showed small but statistically significant differences in gender composition, education level, and fear of driving at the start of training, the effect sizes were negligible ($d \leq 0.09$), and these characteristics are addressed as covariates in the analyses.

HAT improved performance during simulator training and supervised driving: HAT students' viewing skills were better during the intersection test, required fewer on-road training hours, passed the driving exam in fewer attempts, and achieved a higher first-attempt pass rate than the control group. These benefits did not persist into unsupervised post-licensing driving. Violations, errors, and accident involvement were comparable between HAT and control group drivers in the first and last year after licensing. Personal characteristics — including gender, licensing age, self-assessed driving competence, and subjective driving difficulty — were stronger and more lasting predictors of post-licensing behaviour than training type. These findings suggest that hazard awareness is a trainable skill, but that training effects on risk-taking behaviour are moderated by individual characteristics that emerge most clearly once drivers operate independently, aligning with findings of a previous study on the same dataset. Teaching higher safety margins during supervised driving may offer a more durable route to reducing accident risk for novice drivers than higher-order skill training alone.

## 1. Introduction

In the Netherlands, the likelihood of being involved in a car crash is 4.5 times higher for novice drivers aged 18 to 24 than for more experienced drivers aged 30 to 59 (SWOV, 2021). In the first months after licensing, the crash rates are the highest. The crash risk drops substantially over the first two years of driving, with the largest decline occurring in the first year (ECMT, 2006).

A broad coalition of stakeholders in the Netherlands aims to innovate professional driving education to deliver safer, more competent drivers, rather than those who merely demonstrate what the examiner assesses during the examination. Current driver education based on uncontrolled curricula will be replaced by an obligatory national curriculum (Roemer, 2021).

The educational design document for the Dutch National Curriculum Driver Training B (Roelofs et al., 2023) mentions the use of driving simulators to support practical lessons and the training of higher-order skills, such as hazard detection and anticipation. Both are in line with the European Commission's revised Directive on driving licenses (European Commission, 2023). This EC directive recommends incorporating driving simulators into testing risk awareness for both novice and experienced drivers. The driver training, testing, and probation rules for all driving categories must ensure that novice drivers drive safely.

The Dutch National Curriculum Driver Training B (car) requires the acquisition of good viewing skills (conscious awareness). Good viewing skills are essential for detecting the needed information to predict and recognise hazards. McKnight and McKnight's (2003) analysis of police reports supports this requirement: failures in visual scanning, attention maintenance, and speed management were responsible for approximately 87.1% of crashes among young drivers. Errors rather than violations caused the accidents. Vlakveld (2011) noted that these results suggest that poor hazard detection and anticipation skills, rather than reckless driving, are key determinants of the higher accident risk faced by young drivers.

Although educational innovations, such as driving simulators, are expected to mitigate the problem of young drivers, the Netherlands'

---

* Corresponding author at: Green Dino BV, the Netherlands.
*E-mail address:* j.kuipers@greendino.nl (J. Kuipers).

https://doi.org/10.1016/j.aap.2026.108536
Received 16 January 2025; Received in revised form 20 March 2026; Accepted 29 March 2026

National Scientific Institute for Road Safety Research (SWOV, 2019) concluded that scientific evidence remains lacking. Inspired by Vlakveld (2011), SWOV (2019) suggested hazard-detection training using simulators to improve viewing skills and reduce accident risk. Vlakveld (2011) researched computer and simulator methods for training and testing hazard detection and anticipation skills, and gave the following definition of hazard anticipation: The detection and recognition of road and traffic situations that could increase the possibilities of a crash, including the prediction of how these situations can develop into acute threats. The feelings of risk that are evoked by these predictions and the execution of actions that will reduce the feelings of risk and will ensure a safety margin that is large enough to avert a crash should the latent hazard materialise. Hazard anticipation can range from 'automatic' to 'controlled'.

Vlakveld (2011) proposed that moderated risk perception supports the development of schemata for recognising covert (not visible) latent hazards[1] — what Damasio (1996) termed ‘somatic’ markers. These schemata enable early detection, recognition, and swift action, creating a safety margin sufficient to avert a crash. Vlakveld further introduced error-learning: through trial and error on the simulator, students discover how to prevent a crash next time.

Vlakveld (2011) suggested that, in addition to cognitive (trainable) aspects, changes in risk awareness arising from the maturation of emotions and motivations also affect hazard-anticipation performance. In support of Vlakveld, Kuipers, De Winter and Mulder (2023) found respondents' fear levels provided a logical explanation for students' progression through driver education. Drivers who start driving at a higher age are expected to use higher safety margins to mitigate accident risk. In contrast, younger drivers are expected to be more sensation-seeking and to overestimate their capabilities more than older novice drivers, leading them to use lower safety margins. Vlakveld illustrated this with self-reported speed choice for a 50 km/h road. The group aged 20–29 years reported violating the maximum speed by more than 10%. Only after the age of 60 years was the maximum speed obeyed. However, Vlakveld (2011) found no differences in safety margins between novice drivers of younger and older ages.

This study distinguishes between two constructs, following Vlakveld (2011). Hazard awareness refers to the conscious, trainable skill of detecting and anticipating specific hazards — deciding where to look and how to respond to an identified threat. Risk awareness refers to the generalised, largely unconscious tendency to maintain safety margins, driven by emotion and motivation. This distinction generates a directional prediction: HAT, which targets hazard awareness, should improve supervised driving outcomes but leave risk-taking behaviour after licensing relatively unchanged. Vlakveld presented detection and anticipation as trainable skills and emotion and motivation as personal characteristics that influence risk perception and the choice of safety margins. Where hazard awareness involves making conscious decisions to avoid collisions in a specific traffic situation, risk awareness involves making unconscious decisions to increase safety margins in general. This two-component model also implies an important dose–response prediction: if hazard awareness is genuinely trainable, greater training intensity (more lessons, more varied scenarios, spaced repetition) should produce proportionally larger and more durable effects on supervised and unsupervised driving outcomes, whereas risk-awareness outcomes should remain relatively unresponsive to training dose regardless of intensity.

Several overview studies report positive correlations between hazard awareness training and improvements in hazard awareness skills immediately after the training (e.g., McDonald et al., 2015; Omran et al., 2023; Prabhakharan et al., 2024). However, no study has yet examined the long-term impact of HAT on both supervised and unsupervised driving, using a control group (Kuipers et al., 2023). This paper is the first to address this retention question and to investigate the effects of simulator hazard awareness training over a 15-year period (2008–2023) on the driving performance of learner and novice drivers.

Vlakveld (2011) demonstrated in the SimRapt experiment that simulator-based hazard-anticipation training improved visual search (detection) of latent hazards. Whether this effect persisted over an extended period was not investigated, as participants were tested within an hour of the training.

Vlakveld emphasised the importance of error learning in hazard anticipation training, combined with a driving simulator that offers a wide field of view, enabling scanning of the entire traffic situation. In 2013, Green Dino introduced a simulator-based hazard-awareness training program, drawing on the SimRapt scenarios developed by Vlakveld (2011). This hazard awareness training should promote safer driving on the road and reduce involvement in accidents. Teaching higher safety margins, standard viewing (detection) procedures, and influencing emotion and motivation (risk awareness) through virtual car crashes, Green Dino aimed to improve the detection and anticipation skills needed for hazard awareness.

To evaluate the long-term effects of the hazard awareness training, Green Dino provided data from simulator students, encompassing their simulator training scores and responses to two questionnaires. Four research hypotheses were formulated based on the goals of Green Dino and tested on this unique dataset: Hypotheses 1 and 2 test the trainability claim (hazard awareness responds to instruction); Hypothesis 3 tests the persistence claim (trainability extends beyond the supervised phase); Hypothesis 4 tests the boundary condition mentioned (individual differences in emotion and motivation moderate training transfer).

1) Hazard awareness training on simulators improves the viewing skills of learner drivers.
2) Hazard awareness training on simulators facilitates acquiring a safe driving style during education.
3) Hazard awareness training on simulators stimulates a safer driving style after licensing.
4) Emotion and motivation influence hazard awareness.

Kuipers, De Winter, and Mulder (2023) studied the same dataset Green Dino provided and sought measures that explained educational progress. They conclude that personal characteristics such as fear explain progress in education and advise studying the safety and effectiveness of accompanied driving and simulator training. This study follows up.

## 2. Method

The dataset used in this study comprises simulator scores from student drivers who underwent simulator training between 2008 and 2023, as well as responses to two questionnaires: one administered in 2015 and reported in Kuipers, De Winter, and Mulder (2023), and the second in 2023. Green Dino composed the 2015 questionnaire before engaging academic collaborators. Unlike validated instruments such as the Driver Behaviour Questionnaire (DBQ), the in-house questionnaire has not been independently validated, which limits direct comparability with other studies; this limitation is discussed further in Section 4. The 2023 questionnaire was not changed to maintain consistency, but included additional questions on responsibility for an accident,[2] self-comparison of driving skills with those of other drivers, and perceived difficulty driving. Additionally, questions related to violations and errors were included for the last 12 months of driving.

Simulator data were collected from 33 driving schools across the

---

[1] A latent hazard is a typical traffic situation that has the potential to escalate into a hazardous one.

[2] In the questionnaires, "accident" is used instead of "collision" or "impact". This has been continued in this paper.

Netherlands using four different types of simulators (515 students did not provide a driving school name). The main differences between the simulators were the use of beamers, TVs, or computer screens, real car parts for vehicle handling, or Logitech game devices. All simulators had a wide field of view, were constructed with 3 screens, and ran the same software application. The software was updated annually with extra lessons and options. The functional specifications did not differ between the simulator types and software versions. A description of the driving simulators and the hazard awareness training program is included in Appendix I.

### 2.1. Participants

After filtering, the simulator and questionnaire data of 2,372 participants remained for analysis. Of these, 911 (38.4%) identified as male and 1,460 (61.6%) as female; one participant did not report gender. Males had a mean age of 24.4 years and females 24.1 years. Of the 2,370 participants who reported their age, 106 (4.5%) were 17–18 years old, 1,301 (54.9%) were 19–23 years old, and 963 (40.6%) were 24 years or older. Regarding educational attainment, 70.6% of the sample held a higher education qualification (University of Applied Sciences or Research University), while 28.3% had lower levels of education (Secondary Vocational Education or below). The overrepresentation of females and highly educated participants likely influences the results. Kuipers, De Winter, and Mulder (2023) demonstrated gender-related performance differences.

A total of 672 respondents (28.3% of the full sample) completed training at a single driving school. The gender distribution of this school closely resembles that of the broader sample (64.0% female vs. 61.6% overall), making meaningful gender distortion unlikely. However, participants from this school were more highly educated than the broader sample: 78.0% held a University of Applied Sciences or Research University qualification, compared with 70.6% overall, with Research University-level participants particularly overrepresented (46.9% vs 33.7%). Correspondingly, participants in Secondary Vocational Education and below were underrepresented (12.9% vs 19.1%). Given that this school accounts for more than a quarter of the dataset, its composition likely skews the overall education distribution upward and should be considered when interpreting education-related findings.

### 2.2. Study design

To study the effects of Hazard Awareness Training, data were divided into two groups: an HAT group and a control group. The HAT group comprised student drivers who completed both Hazard Detection Training and Testing (HDTT) and Hazard Anticipation Training and Testing (HATT). HDTT focuses on error-reduction training for standard viewing procedures — where to look — with direct visual and audio feedback. HATT builds on this by using training scenarios based on the SimRapt methodology introduced by Vlakveld (2011), incorporating specific scenario precursors, structured feedback, and virtual car crashes. HATT is completed after HDTT, once students have mastered the required viewing procedures. The total duration of HAT is approximately four hours, depending primarily on the number of additional rehearsal lessons needed.

The control group did not actively choose to forgo HAT; rather, the simulators available to them did not offer it, as HAT was only available on simulators equipped with a webcam and HATT from 2013 onwards. All participants, regardless of group, completed the same vehicle handling and intersection lessons on the simulator.

Only respondents who completed both a vehicle handling test and an intersection test on a simulator between 2008 and 2023 were included in the analyses, thereby mitigating the self-selection effect — in contrast to the approach taken by Kuipers, De Winter, and Mulder (2023). All remaining respondents were included in the analysis of differences in simulator test performance across vehicle-handling and intersection tasks. Students who completed both simulator tests but did not receive HAT, including those who underwent simulator training prior to 2015, formed the control group. While the HAT and control groups showed statistically significant differences in gender composition ($\chi^2 = 4.73$, $p = 0.030$), education level ($\chi^2 = 9.71$, $p = 0.002$), and fear of driving at the start of training (Table 3: $d = 0.13$, $p = 0.040$), all effect sizes were small to negligible. The HAT group had slightly more female participants (63.0% vs 58.0%) and a higher proportion of highly educated participants (24.0% vs 17.9% HBO/WO). These differences are noted as potential confounds and are taken into account when interpreting group comparisons.

A distinction is drawn between hazard awareness, treated as a conscious and therefore trainable skill, and risk awareness, treated as an unconscious process driven by emotion and motivation, and therefore less amenable to direct training.

### 2.3. Data preparation and statistical analysis

The data preparation and most analyses were conducted in IBM SPSS Statistics. In round two of the peer review, Claude.ai (Anthropic) was used to calculate Cronbach's alpha, and ChatGPT (OpenAI) was used to check these calculations. Continuous variables were derived from existing records: simulator training hours were calculated by dividing the number of simulator lessons by 2 (30 min per lesson), and these were combined with on-road training hours to produce total driving education hours. Licensing age was computed as a continuous variable (current age minus years licensed, with a 0.5 correction for mid-year acquisition), and kilometres driven were rescaled to thousands for interpretability. Composite scores for traffic violations and driving errors were computed by summing relevant items separately for the first 12 months and the most recent 12 months post-licensure.

Driving anxiety was categorised into three levels. HDTT and HATT respondents were combined into a single HAT group because only small effects were observed when analysed separately; HDTT and HATT are elaborated in Appendix I. Education level was not included in the analyses due to a strongly skewed distribution in the sample.

Several filters were applied to prepare the data sample for analysis of specific topics. Most analyses used the combined data gathered between 2008 and 2023. Where only the latest inventory data (2015–2023) was applicable or available, a remark is added. The sample size varies across analyses due to differences in the filters applied. Due to very small group sizes, accident details — such as whether multiple drivers were involved or whether the respondent was the offender — were not analysed.

Group differences in continuous outcomes were tested using independent samples t-tests or one-way ANOVA when predictors had three or more categories. To examine within-person change over time, T1 and T2 scores were compared using paired samples t-tests. Associations between categorical variables were assessed using chi-square tests. Where distributional assumptions were violated, nonparametric equivalents (Mann-Whitney U and Kruskal-Wallis) were used. Effect sizes were reported as Cohen's d for mean comparisons (interpreted as small ($d = 0.20$), medium ($d = 0.50$), and large ($d = 0.80$)) and as Spearman's r for nonparametric and ordinal associations. Statistical significance was set at $p < 0.05$.

## 3.

The internal consistency of composite scores was assessed using Cronbach's alpha (α). The simulator training programme showed excellent reliability across its 23 lesson scores ($\alpha = 0.91$), supporting its use as a coherent measure of the acquisition of progressive driving skills. The three core intersection test scores — safety, risk mitigation, and driving skill — demonstrated acceptable reliability ($\alpha = 0.73$). The eco driving score was excluded from this composite because its item-total correlation was markedly lower ($r = 0.13$ vs. 0.48–.52 for the other three items); its removal raised alpha to 0.73, and it is therefore treated

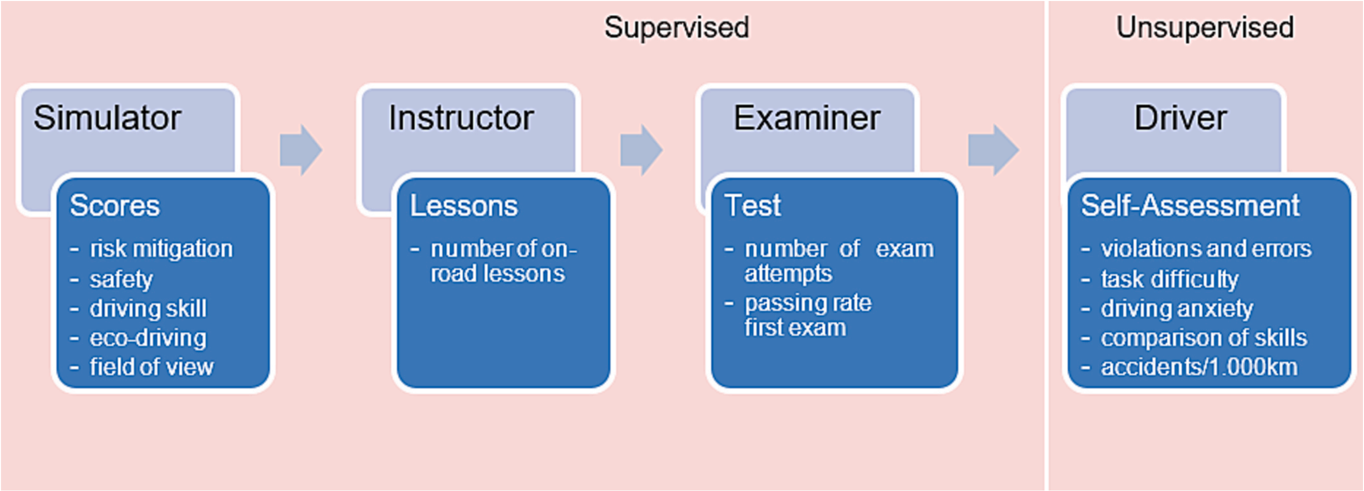


**Fig. 1.** Driver safety indicators.

**Table 1**
Simulator scores, entire sample. Bold p-values are significant ($p < 0.05$).

| | Driving skill test | | | Intersection test | | | | |
|---|---|---|---|---|---|---|---|---|
| | **Mean** | ***SD*** | ***n*** | **Mean** | ***SD*** | ***n*** | **Cohen's *dr*** | ***p*** |
| Simulator risk mitigation score | 5.04 | 0.65 | 390 | 5.46 | 0.71 | 236 | 0.53 | **< 0.0001** |
| Simulator safety score | 6.32 | 1.34 | 390 | 7.25 | 0.98 | 236 | 0.58 | **< 0.0001** |
| Simulator driving skill score | 6.45 | 1.59 | 390 | 7.94 | 1.00 | 236 | 0.78 | **< 0.0001** |
| Simulator eco driving score | 6.26 | 1.14 | 390 | 7.22 | 1.50 | 236 | 0.64 | **< 0.0001** |
| Simulator field of view score | 3.75 | 2.28 | 140 | 5.94 | 1.89 | 197 | 1.05 | **< 0.0001** |

as a separate outcome throughout. The driving skill test composite (driving skill, safety, risk mitigation) showed acceptable internal consistency ($\alpha = 0.66$) when the eco-driving subscale was excluded. The violation score (8 items, $\alpha = 0.55$) and error score (3 items, $\alpha = 0.44$) reflect moderate internal consistency, which is expected for behavioural frequency scales measuring conceptually related but distinct risky acts rather than a single latent construct. The simulator scores are summarised in Appendix II of this paper. The self-reported measures are summarised by Kuipers, De Winter, and Mulder (2023).

The first indicator is the simulator assessment, based on the difference between five simulator scores obtained between the vehicle handling test at the beginning of the simulator curriculum and the intersection test. The intersection test is conducted approximately 3.5 h after training on the intersection procedures. This information is used to study Hypotheses 1 and 4.

After the simulator training, students received further instruction on the road from a driving instructor. The second indicator is the driving instructor's assessment, reflected in the total number of (self-reported) 60-minute on-road lessons the student completed before the instructor determined that the student driver was ready to drive unsupervised and required no further training to apply for the driving exam. This information is used to study Hypotheses 2 and 4.

The third indicator is the examiners' assessment of whether the student driver can drive unsupervised, based on the success of the first exam attempts and the total number of exam attempts. This information is also used to study Hypotheses 2 and 4.

The fourth indicator is the drivers' self-assessment of errors, violations, task difficulty, driving anxiety, skills and accident risk, obtained after the training when driving unsupervised. This information is used to study Hypotheses 3 and 4.

The self-assessment is studied for four periods of driving license possession: (i) the first six months, (ii) seven to twelve months, (iii) the first twelve months, and (iv) the last twelve months. The last twelve months did not overlap with the other periods. The split in the first year was made to investigate differences between the first and second halves of the first year of driving, which were found between the first and last years of driving. The gap between the first and last year of driving varied between participants but was not considered further in the analysis.

## 4. Results

The results are split into subsections, 'Driving Supervised' and 'Driving unsupervised' (see Fig. 1). First, the objective performance measurements obtained in the driving simulator and the (semi-objective) performance assessments of the driving instructor and examiner, as reported by the respondents, are presented. Then, the respondents' subjective self-assessment measures regarding their driving behaviour after licensing follow.

### 4.1. *Driving supervised*

#### 4.1.1. *Simulator*

To determine the effect of HAT on hazard detection skills (Hypothesis 1), differences in the 'field of view score' between the driving skill test (at the end of the first module, 'Vehicle Operation') and the intersection test (at the end of the second module, 'Intersections') were analysed.

According to the simulator's skill assessment metrics, higher scores indicate more advanced detection skills. Simulator scores for safety and economic driving skills were analysed to determine the effects of simulator training on anticipatory skills. For the analysis of the field of view score, only data from the HAT group were used. Data for the control group were not available because they drove on simulators without a head-tracking device. For the analysis of the other scores, data from both groups were used.

The risk mitigation score reflects the applied safety margins; the safety risk score reflects (near) traffic accidents (collisions); the eco driving score reflects fuel efficiency (low motor revolutions); and the

**Table 2**
Correlation matrix, entire sample. Bold r-values are significant (p < 0.05).

| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | HT (0 = no, 1 = yes) | | | | | | | | | | | | | | |
| 2 | Gender (0 = male, 1 = female) | 0.07 | | | | | | | | | | | | | |
| 3 | Licencing age | **−0.12** | 0.01 | | | | | | | | | | | | |
| 4 | Driving anxiety | −0.07 | **0.26** | **0.21** | | | | | | | | | | | |
| 5 | Comparison driving skills | −0.07 | **−0.23** | **−0.33** | **−0.32** | | | | | | | | | | |
| 6 | Subjective driving difficulty | **0.17** | **0.23** | **0.38** | **0.47** | **−0.61** | | | | | | | | | |
| 7 | Risk mitigation score at intersection test | 0.00 | **0.21** | 0.06 | **0.11** | 0.03 | 0.07 | | | | | | | | |
| 8 | Safety score at intersection test | −0.01 | **−0.24** | **−0.11** | **−0.17** | 0.11 | −0.10 | **0.28** | | | | | | | |
| 9 | Driving skill score intersection test | 0.08 | **−0.14** | **−0.13** | **−0.12** | 0.05 | 0.00 | **0.27** | **0.78** | | | | | | |
| 10 | Eco driving score intersection test | **−0.20** | **0.12** | **0.10** | **0.12** | 0.08 | −0.03 | **0.30** | 0.00 | 0.02 | | | | | |
| 11 | Field of view score intersection test | | 0.00 | 0.03 | 0.04 | −0.05 | 0.05 | 0.14 | **0.22** | **0.16** | 0.13 | | | | |
| 12 | Simulator training hours | **0.27** | 0.00 | **−0.09** | 0.03 | −0.05 | 0.06 | 0.07 | **0.18** | **0.18** | **−0.09** | 0.08 | | | |
| 13 | On road training hours | **−0.10** | **0.14** | **0.21** | **0.27** | **−0.20** | **0.26** | 0.05 | **−0.19** | **−0.19** | 0.03 | **−0.23** | 0.00 | | |
| 14 | Total duration driving education | −0.07 | **0.14** | **0.20** | **0.27** | **−0.21** | **0.26** | 0.06 | **−0.16** | **−0.17** | 0.02 | **−0.22** | **0.14** | **0.99** | |
| 15 | Number of driving exams | **−0.10** | 0.07 | 0.01 | **0.12** | −0.01 | 0.02 | −0.01 | **−0.15** | **−0.14** | 0.00 | −0.11 | **−0.08** | **0.32** | **0.31** |

**Table 3**
Differences between the control group (No) and HAT (Yes). Bold p-values are significant (p < 0.05).

| | No | | | Yes | | | | |
|---|---|---|---|---|---|---|---|---|
| | **Mean** | ***SD*** | ***n*** | **Mean** | ***SD*** | ***n*** | **Cohen's *dr*** | ***p*** |
| Simulator risk mitigation score (1—10) | 5.44 | 0.67 | 474 | 5.46 | 0.72 | 197 | −0.03 | 0.72 |
| Simulator safety score (1—10) | 7.31 | 1.08 | 474 | 7.23 | 0.96 | 197 | 0.08 | 0.34 |
| Simulator driving skill score (1—10) | 7.76 | 1.13 | 474 | 7.92 | 1.00 | 197 | −0.14 | 0.09 |
| Simulator eco driving score (1—10) | 7.74 | 1.53 | 474 | 7.16 | 1.48 | 197 | **0.39** | **0<.0001** |
| Simulator field of view score (1—10) | | | | | | | | |
| Number of simulator training hours | 6.15 | 2.60 | 521 | 7.50 | 2.37 | 197 | −0.53 | **0<.0001** |
| Number of on-road training hours | 43.27 | 19.95 | 518 | 39.18 | 17.10 | 188 | 0.21 | **0.01** |
| Number of total training hours | 49.42 | 20.20 | 518 | 46.63 | 17.25 | 188 | 0.04 | 0.09 |
| Number of driving test attempts | 1.78 | 1.01 | 521 | 1.53 | 0.82 | 195 | 0.26 | **0<.0001** |
| Passed first exam (0 = no, 1 = yes) | 0.52 | 0.50 | 521 | 0.64 | 0.48 | 195 | −0.25 | **0<.0001** |
| Violations first 12 months | 2.05 | 2.16 | 444 | 2.06 | 2.13 | 118 | 0.00 | 0.97 |
| Errors first 12 months | 1.77 | 1.26 | 435 | 1.63 | 1.20 | 117 | 0.11 | 0.28 |
| Fear at start of driving education (1—5) | 2.82 | 1.22 | 521 | 2.66 | 1.21 | 195 | 0.13 | 0.13 |
| Violations last 12 months | 2.90 | 2.02 | 30 | 3.05 | 2.92 | 93 | −0.06 | 0.79 |
| Errors last 12 months | 1.23 | 1.41 | 30 | 1.26 | 1.33 | 92 | −0.02 | 0.92 |
| Accident involvement first 12 months(0 = No, 1 = Yes) | 0.09 | 0.29 | 462 | 0.10 | 0.30 | 124 | −0.02 | 0.84 |
| Accident involvement last 12 months(0 = No, 1 = Yes) | 0.11 | 0.31 | 352 | 0.09 | 0.28 | 94 | 0.07 | 0.57 |
| Accident risk first 12 months (accidents/1000 km) | 0.11 | 0.54 | 435 | 0.15 | 1.33 | 117 | −0.06 | 0.55 |
| Accident risk last 12 months (accidents/1000 km) | 0.05 | 0.16 | 33 | 0.02 | 0.06 | 88 | 0.30 | 0.14 |

field of view score reflects viewing appropriate areas (see Appendix II for more details).

The paired-samples t-tests show that students improved across all five scores and learned from the simulator intersection lessons (Table 1). Compared to the vehicle handling test, student's (1) risk mitigation score increased; they were more risk aware (d = 0.53, p < 0.001), (2) safety risk score increased; they anticipate better (d = 0.58, p < 0.001), (3) driving skill score increased; they improved their vehicle handling skills (d = 0.78, p < 0.001), (4) eco driving score increased; they drove more environmentally friendly (d = 0.64, p < 0.001), and (5) field of view score increased; they improved their detecting skills (d = 1.05, p < 0.001). The effect sizes were medium for most scores, but large for the field-of-view score.

Table 2 presents the Pearson correlation matrix of several dependent measures obtained in this study. HAT training (column 1) shows a negative correlation with the eco-driving score (row 10, r = -0.20) at the intersection test, indicating that students in the HAT group scored lower on the eco-driving assessment. This effect was further investigated with an independent *t*-test comparing the HAT and control groups, which showed a significant difference (d = 0.39, p < 0.001; Table 3). The HAT group did not differ significantly from the control group in safety, driving skill, or risk mitigation scores on the vehicle handling test, confirming comparable performance at the start of the education phase.

Gender differences were observed (column 2 in Table 2). Higher safety (row 8, r = -0.24) and driving skill scores (row 9, r = -0.14) are associated with being male, whereas females show better risk mitigation (row 7, r = 0.21) and eco-driving scores (row 10, r = 0.12). Males exhibited a more fluent driving style and were better at anticipating intersections (as indicated by the safety score and driving skill score). Women drove more carefully and risk-aware (as reflected in the risk mitigation score).

Self-reported driving anxiety (column 4 in Table 2) is negatively correlated with safety (row 8, r = -0.17) and driving skill scores (row 9, r = -0.12). Risk mitigation (row 7, r = 0.11) and eco driving scores (row 10, r = 0.12) demonstrate opposite effects.

The field of view score positively correlates (row 11, Table 2) with the safety score (column 8, r = 0.22) and the driving skill score (column 9, r = 0.16). No significant differences were observed with the risk mitigation score (column 7) and between field of view scores of males and females (column 2, row 11) or simulator hours (column 12, row 11).

**Table 4**

Correlations for the first twelve months of driving. Bold values are significant ($p < 0.05$).

| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | HAT (0 = no, 1 = yes) | | | | | | | | | | | | | | | | | |
| 2 | Gender (0 = male, 1 = female) | 0.05 | | | | | | | | | | | | | | | | |
| 3 | Licencing age | **−0.12** | 0.00 | | | | | | | | | | | | | | | |
| 4 | Driving anxiety (1—5) | −0.08 | **0.25** | **0.20** | | | | | | | | | | | | | | |
| 5 | Comparison driving skills (0—4) | −0.04 | **−0.28** | **−0.37** | **−0.35** | | | | | | | | | | | | | |
| 6 | Subjective driving difficulty (0—4) | 0.18 | **0.27** | **0.39** | **0.52** | **−0.65** | | | | | | | | | | | | |
| 7 | On road training hours | −0.07 | **0.14** | **0.18** | **0.25** | **−0.24** | **0.25** | | | | | | | | | | | |
| 8 | Number of driving exams | −0.03 | 0.06 | 0.00 | **0.12** | −0.08 | 0.00 | **0.28** | | | | | | | | | | |
| 9 | Risk mitigation score at intersection test (0—10) | −0.03 | **0.19** | 0.07 | **0.11** | 0.03 | 0.05 | **0.10** | 0.02 | | | | | | | | | |
| 10 | Safety score at intersection test (0—10) | −0.03 | **−0.26** | **−0.12** | **−0.17** | 0.17 | −0.12 | **−0.15** | **−0.11** | **0.27** | | | | | | | | |
| 11 | Driving skill score intersection test (0—10) | 0.04 | **−0.13** | **−0.13** | **−0.10** | 0.14 | 0.02 | **−0.15** | **−0.10** | **0.29** | **0.78** | | | | | | | |
| 12 | Eco driving score intersection test (0—10) | **−0.19** | **0.09** | **0.09** | **0.12** | 0.11 | −0.01 | 0.05 | 0.00 | **0.31** | −0.01 | 0.01 | | | | | | |
| 13 | Field of view score intersection test (0—10) | | 0.03 | 0.01 | 0.07 | 0.02 | 0.00 | **−0.30** | −0.10 | **0.27** | **0.22** | 0.11 | 0.14 | | | | | |
| 14 | Km driven first twelve months | −0.03 | **−0.18** | **0.09** | **−0.13** | 0.19 | **−0.20** | **−0.09** | −0.02 | −0.02 | **0.09** | 0.06 | 0.00 | 0.04 | | | | |
| 15 | Violation score first twelve months (0—20) | 0.00 | **−0.15** | **−0.14** | **−0.15** | 0.15 | −0.12 | **−0.11** | −0.02 | **−0.17** | 0.08 | 0.04 | **−0.11** | −0.04 | **0.23** | | | |
| 16 | Error score first twelve months (0—12) | −0.04 | 0.00 | 0.00 | **0.14** | 0.10 | −0.04 | 0.03 | 0.07 | 0.02 | −0.05 | **−0.11** | 0.02 | 0.02 | **0.13** | **0.18** | | |
| 17 | Accidents first twelve months (0 = no, 1 = yes) | −0.04 | 0.01 | 0.08 | 0.02 | −0.02 | 0.12 | 0.00 | −0.02 | −0.03 | 0.00 | −0.01 | 0.08 | −0.11 | 0.05 | **0.09** | 0.07 | |
| 18 | Accident risk first twelve months (accidents/1.000 km) | 0.03 | 0.05 | 0.01 | 0.05 | −0.05 | 0.14 | 0.01 | −0.01 | −0.02 | −0.04 | 0.00 | 0.05 | 0.01 | −0.07 | 0.03 | 0.03 | **0.49** |

Apparently, detection skills are associated with vehicle handling and anticipation skills, but not with risk awareness.

The driving skill and safety scores at the intersection test are strongly correlated (column 8, row 9, r = 0.78 in Table 2). The driving skill score also positively correlates with the risk mitigation score (column 7, row 9, r = 0.27). Being more skilled in vehicle handling procedures is associated with being more risk-aware and better at anticipating intersections. These correlations can be partly explained by the fact that the metrics use the same variables to calculate both scores, such as inappropriate steering (see Appendix II).

### 4.1.2. *Instructor*

Significant positive correlations are observed between on-road training hours (row 13, Table 2) and gender (column 2, r = 0.14), licensing age (column 3, r = 0.21), driving anxiety (column 4, r = 0.27) and subjective driving difficulty (column 6, r = 0.26. This suggests that females, older students, students with higher anxiety levels, and students who experienced greater driving difficulty required more on-road training hours before the instructor considered them sufficiently competent to drive unsupervised. On-road training hours correlate negatively with HAT (column 1, r = -0.10) and comparison of driving skills (column 5, r = -0.20), suggesting that students who followed HAT and those who perceived themselves as better drivers than their peers required less on-road training.

On-road training hours and the total duration of driving education are strongly correlated (column 13, row 14, r = 0.99) and are nearly equally correlated with all other measures. The number of simulator training hours (column 12) did not affect the number of on-road driving lessons (row 13, r = 0.00). However, it increased the total duration of driving education (row 14, r = 0.14), suggesting that driving instructors did not recognise, or ignored, the simulator experience reported in paragraph 3.1.1.

Some measures do not correlate at all (Table 2): HAT and the risk mitigation score (items 1 and 7), gender and the field of view score (items 2 and 11), gender and the simulator training hours (items 2 and 12), the subjective driving skill score and the driving skill score (items 6 and 9), the safety score and the eco-driving score (items 8 and 10), the eco-driving score and the number of driving exams (items 10 and 15), and the amount of simulator training hours and on-road training hours (items 12 and 13).

Negative correlations indicate that the instructor's judgment on driving proficiency is comparable with that of the simulator (Table 2). Students who scored higher on the safety, driving skill and field of view scores in the simulator were considered ready for the driving exam after fewer on-road training hours (column 8, row 13, r = -0.19; column 9, row 13, r = -0.19; column 11, row 13, r = -0.23).

Table 3 shows that students in the HAT group spent 1.35 h more on the simulator (d = -0.53, $p < 0.001$). However, they then required significantly fewer on-road hours, 39.18, compared to 43.27 for the control group (d = 0.21, p = 0.01). As a result, the average duration of the total driver education was 46.63 h for the HAT group and 49.42 h for the control group (d = 0.04, p = 0.09).

### 4.1.3. *Examiner*

Table 2 shows that reporting higher levels of driving anxiety before driving education correlates positively with needing more exam attempts to pass (column 4, row 15, r = 0.12).

Students who scored higher on the simulator's safety and driving skill assessment required fewer exam attempts (column 8, row 15, r = -0.15; column 9, row 15, r = -0.14), indicating that the simulator's automated assessment and the examiner's judgment of driving competence are aligned.

Table 3 shows that control-group students required 1.78 attempts to pass the driving test, whereas the HAT group required 1.53 attempts (d = 0.26, $p < 0.001$). Similar results are observed for the first exam passing rate. The control group's passing rate on the first exam is 0.52,

**Table 5**
Differences in driving between the first and last twelve months. Bold p-values are significant ($p < 0.05$).

| | First 12 months | | | Last 12 months | | | | |
|---|---|---|---|---|---|---|---|---|
| | **Mean** | ***SD*** | ***n*** | **Mean** | ***SD*** | ***n*** | **Cohen's *dr*** | ***p*** |
| Violation score (0——20) | 1.97 | 2.08 | 789 | 2.69 | 2.60 | 796 | −0.39 | **0<.0001** |
| Error score (0——12) | 1.44 | 1.29 | 786 | 1.15 | 1.20 | 802 | 0.30 | **0<.0001** |
| Accident involvement (0 = no, 1 = yes) | 0.10 | 0.30 | 1543 | 0.08 | 0.28 | 1536 | 0.04 | 0.11 |
| Accident risk (accidents/1.000 km) | 0.16 | 1.20 | 1489 | 0.07 | 1.37 | 600 | 0.03 | 0.52 |

**Table 6**
Differences between drivers licensed less than six months and drivers licensed between six and twelve months. Bold p-values are significant ($p < 0.05$).

| | <6 months | | | 6–12 months | | | | |
|---|---|---|---|---|---|---|---|---|
| | **Mean** | ***SD*** | ***n*** | **Mean** | ***SD*** | ***n*** | **Cohen's *dr*** | ***p*** |
| Comparison driving skills (0——4) | 1.88 | 0.87 | 110 | 2.16 | 0.73 | 71 | −0.34 | **0.02** |
| Subjective driving difficulty (0——4) | 1.95 | 1.07 | 110 | 1.76 | 1.08 | 71 | 0.17 | 0.26 |
| Violation score first 12 months (0——20) | 1.74 | 1.70 | 265 | 2.26 | 2.39 | 275 | −0.25 | **0<.0001** |
| Error score first 12 months (0——12) | 1.51 | 1.36 | 270 | 1.51 | 1.23 | 275 | 0.00 | 0.99 |
| Km driven first 12 months | 1487 | 2987 | 264 | 2955 | 5304 | 277 | −0.34 | **0<.0001** |
| Accident involvement first 12 months (0 = No, 1 = Yes) | 0.06 | 0.24 | 277 | 0.08 | 0.27 | 283 | −0.06 | 0.45 |
| Accidents risk first 12 months (accidents/1.000 km) | 0.24 | 2.20 | 263 | 0.15 | 0.95 | 277 | 0.05 | 0.53 |

compared to 0.64 for the HAT group (d = -0.25, $p < 0.001$).

#### 4.1.4. *Summary of supervised driving results*

The analysis of the measures of supervised driving showed significant results: (1) Students improve their driving skills during the simulator training on all measured aspects, especially on detection skills. (2) The HAT group scored lower on eco driving; no other significant differences were found in the intersection test simulator scores. (3) HAT shortened the duration of driving education to become a sufficiently competent driver. The HAT group drivers required fewer on-road training hours ($M1 - M2 = 4.09$) and passed the exam with fewer attempts ($M1 - M2 = 0.25$). (4) The simulator, instructor, and examiner similarly assessed safe and competent driving. Students with higher safety, driving skills, and field-of-view simulator scores required fewer on-road training hours and passed the exam with fewer attempts.

### 4.2. *Driving unsupervised*

#### 4.2.1. *Violations*

In the first twelve months, the violation score (item 15 in Table 4) showed a negative correlation with gender (column 2, r = -0.15), licensing age (column 3, r = -0.14), and driving anxiety (column 4, r = -0.15). This demonstrates that males, drivers who were licensed at an earlier age, and drivers who experienced less anxiety reported higher levels of violating/risk-taking behaviour.

A significant negative correlation is found between violation scores and on-road training hours (row 15, column 7, r = -0.11). This implies that having fewer on-road training hours is associated with *more* risk-taking in the first twelve months of driving. Drivers judged by the instructor to be ready in an earlier phase of driving education reported more violating behaviour than those considered to need more lessons.

The risk mitigation simulator score (column 9, r = -0.17) and eco driving score (column 12, r = -0.11) have a negative correlation with violation scores (row 15), which suggests that students with higher risk mitigation and eco-driving scores reported lower violation scores.

The violation score (row 15) correlated positively with kilometres driven in the first twelve months (column 14, r = 0.23), error score (column 15, row 16, r = 0.18), and accident involvement in the first twelve months (column 15, row 17, r = 0.09), indicating that drivers that had a riskier driving style in the first twelve months drove more kilometres, made more errors and were also more likely to be involved in an accident.

The violation scores increased from the first year of driving (M = 1.97) to the last (M = 2.69) (Table 5). A similar effect is observed when the first twelve months are further divided into two groups (Table 6). Drivers licensed for less than six months had a significantly lower violation score (M = 1.74) than drivers licensed for six to twelve months (M = 2.26).

The significant correlations found in the first twelve months (Table 4) do not replicate exactly the correlations in the last twelve months (Table 7). While the violation score remains negatively correlated with licensing age (column 3, row 15, r = -0.21), neither gender (column 2, row 15) nor driving anxiety (column 4, row 15) are associated with the violation score in the last year. Subjective comparison of one's driving skills (column 5, row 15, r = 0.33) and subjective driving difficulty (column 6, row 15, −0.24) remain correlated with the violation score in the last twelve months of driving.

There was no difference in the violation score between HAT (M = 2.03) and the control group (M = 2.04) in the first twelve months (Table 3). This suggests that drivers demonstrate comparable low-level risk-related behaviour, regardless of the type of training. The difference over the last 12 months could not be calculated because the control group's violation score sample was too small.

#### 4.2.2. *Errors*

The error score (row 16 in Table 4) in the first twelve months positively correlates with driving anxiety (column 4, r = 0.14), indicating that students who report greater anxiety before driving education are more likely to make errors in the first year after obtaining their license. Furthermore, a negative correlation with the driving skill simulator score (column 11, r = -0.11) suggests that error-prone drivers can already be detected during simulator training.

Positive correlations are observed with kilometres driven (column 14, row 16, r = 0.13) and violation scores (column 15, row 16, r = 0.18). The positive correlation between the error and violation scores was the only one maintained over the last 12 months (column 15, row 16; r = 0.32 in Table 7).

Error scores of all drivers are similar between the first and second halves of the year (M = 1.51, Table 6). The error score decreased significantly from the first to the last twelve months ($M1 - M2 = 0.29$ in Table 5), indicating that fewer errors were made only *after* the first twelve months of driving unsupervised.

**Table 7**

Correlations for the last twelve months of driving. Bold r-values are significant ($p < 0.05$).

| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | HDTT/HATT (0 = no, 1 = yes) | | | | | | | | | | | | | | | | | |
| 2 | Gender (0 = male, 1 = female) | 0.04 | | | | | | | | | | | | | | | | |
| 3 | Licencing age | **−0.11** | 0.03 | | | | | | | | | | | | | | | |
| 4 | Driving anxiety (1——5) | −0.08 | **0.22** | **0.21** | | | | | | | | | | | | | | |
| 5 | Comparison driving skills (0——4) | −0.02 | **−0.27** | **−0.38** | **−0.35** | | | | | | | | | | | | | |
| 6 | Subjective driving difficulty (0——4) | 0.18 | **0.26** | **0.38** | **0.56** | **−0.64** | | | | | | | | | | | | |
| 7 | On road training hours | −0.08 | **0.13** | **0.19** | **0.28** | **−0.27** | **0.23** | | | | | | | | | | | |
| 8 | Number of driving exams | 0.01 | 0.09 | 0.01 | **0.16** | −0.12 | 0.06 | **0.33** | | | | | | | | | | |
| 9 | Risk mitigation score at intersection test (0——10) | −0.05 | **0.16** | 0.06 | **0.12** | −0.01 | 0.08 | **0.10** | 0.06 | | | | | | | | | |
| 10 | Safety score at intersection test (0——10) | **−0.13** | **−0.33** | −0.07 | **−0.19** | 0.19 | −0.13 | **−0.18** | **−0.16** | **0.25** | | | | | | | | |
| 11 | Driving skill score intersection test (0——10) | 0.01 | **−0.20** | **−0.13** | **−0.13** | 0.18 | 0.02 | **−0.20** | **−0.11** | **0.27** | **0.81** | | | | | | | |
| 12 | Eco driving score intersection test (0——10) | **−0.28** | **0.11** | **0.10** | **0.15** | 0.12 | 0.03 | 0.07 | 0.05 | **0.29** | −0.03 | −0.03 | | | | | | |
| 13 | Field of view score intersection test (0——10) | | 0.03 | 0.04 | 0.09 | −0.01 | 0.04 | −0.21 | −0.08 | **0.23** | **0.22** | 0.17 | 0.12 | | | | | |
| 14 | Km driven last twelve months | −0.16 | −0.10 | 0.01 | **−0.19** | **0.32** | **−0.41** | −0.18 | −0.11 | −0.09 | 0.12 | −0.02 | −0.05 | 0.00 | | | | |
| 15 | Violation score last twelve months (0——20) | 0.06 | 0.00 | **−0.21** | −0.09 | **0.33** | **−0.24** | −0.14 | −0.03 | −0.12 | 0.14 | 0.03 | −0.07 | −0.02 | **0.39** | | | |
| 16 | Error score last twelve months (0——12) | 0.10 | 0.13 | 0.00 | 0.10 | −0.05 | 0.07 | −0.03 | −0.01 | −0.10 | 0.02 | −0.08 | −0.05 | 0.00 | −0.02 | **0.32** | | |
| 17 | Accidents last twelve months (0 = no, 1 = yes) | −0.02 | −0.07 | 0.02 | 0.00 | 0.08 | 0.01 | 0.07 | −0.02 | −0.03 | 0.03 | 0.02 | 0.01 | −0.05 | 0.14 | 0.08 | 0.04 | |
| 18 | Accidentrisk last twelve months (/1000 km) | −0.12 | −0.06 | −0.13 | −0.05 | 0.04 | 0.07 | 0.00 | −0.17 | −0.03 | 0.13 | 0.16 | 0.01 | −0.15 | −0.10 | 0.01 | 0.08 | **0.71** |

HAT training did not significantly affect error rates during the first 12 months (Table 3), indicating that driving proficiency remained equivalent between the groups. The difference in the last twelve months could not be calculated because the control group sample size was too small.

### 4.2.3. *Comparison of driving skills*

For the self-comparison of driving skills with those of other drivers, the dataset from the 2023 questionnaire was used (sample 2015–2023). The 2023 questionnaire contains additional information to the 2015 questionnaire regarding the comparison of driving skills.

Table 2 shows significant negative correlations between the self-comparison of driving skills (row 5) and gender (column 2, r = -0.23), licensing age (column 3, r = -0.33), driving anxiety (column 4, r = -0.32), and subjective driving difficulty (column 5, row 6, r = -0.61). Drivers who were male, younger, less anxious, and experiencing less driving difficulty found themselves to be better skilled than their peers.

The self-assessment of one's driving skills compared to others increased from 1.88 in the first six months to 2.16 in the second six months on a scale of 0–4 (d = -0.34) (Table 6). After a few months of driving unsupervised, drivers seem to become more convinced that they are better drivers than others. The longer one is licensed, the more advanced one judges their driving competence compared with peers.

For the last twelve months, comparison of driving skills shows significant correlations with the violation score (column 5, row 15, r = 0.33) (Table 7). Confidence in driving skills appears to stimulate violations of driving rules.

### 4.2.4. *Subjective driving difficulty*

For the analysis of subjective driving difficulty, the dataset from the 2023 questionnaire was used (sample 2015–2023). The 2023 questionnaire contains additional information on experienced driving difficulty compared to the 2015 questionnaire.

The subjective driving difficulty was negatively correlated with the number of kilometres driven in the first twelve months (column 6, row 14, r = -0.20) (Table 4), and the last twelve months (column 6, row 14, r = -0.41) (Table 7), suggesting that drivers who experienced higher driving difficulty drove fewer kilometres.

The subjective driving difficulty score correlates with the violation score (column 6, row 15, r = -0.24) (Table 7), higher distance driven (column 6, row 14, r = -0.41) and judgment of having higher driving skills than others (column 5, row 6, r = -0.64). It indicates that the less driving difficulty drivers experienced, the more confident they were in their driving skills, the more kilometres they drove, and the more violations they committed.

### 4.2.5. *Accidents and accident risk*

Accident involvement did not significantly differ between the first and second half years of driving (Table 6, sample 2015–2023) or between the first and last year of driving (Table 5, sample 2008–2023). This suggests that accident involvement remained approximately the same over the unsupervised driving years.

Drivers involved in an accident drove more kilometres than those not involved in the first twelve months (p = 0.029, d = -0.17) (Table 8) and the last twelve months (p = 0.002, d = -0.52) (Table 9). The risk of being involved in an accident dropped considerably between the first six months and the second six months (0.24 > 0.15) (Table 6) and between the first year and the last year (0.16 > 0.07) (Table 5), both with nine accidents per 100.000 km (M1 – M2 = 0.09). Nevertheless, these differences were insignificant, mainly due to the large deviation in reported kilometres.

In the first twelve months, a positive correlation was observed between accident involvement and reported violation score (column 15, row 17, r = 0.09) (Table 4), demonstrating that drivers with riskier driving behaviour were more likely to be involved in an accident. Additionally, drivers involved in an accident had a higher violation

**Table 8**
Differences between drivers involved in an accident (Yes) and those not involved in an accident (No) in the first twelve months. Bold p-values are significant ($p < 0.05$).

| | No<br>Mean | *SD* | *n* | Yes<br>Mean | *SD* | *n* | Cohen's *dr* | *p* |
|---|---|---|---|---|---|---|---|---|
| HDTT/HATT (0 = no, 1 = yes) | 0.56 | 1.17 | 491 | 0.38 | 1.01 | 47 | 0.00 | |
| Gender (0 = male, 1 = female) | 0.61 | 0.49 | 1784 | 0.65 | 0.48 | 190 | 0.00 | |
| Licencing age | 21.12 | 4.97 | 1772 | 21.95 | 6.00 | 191 | 0.00 | |
| Driving anxiety before driving education | 2.75 | 1.25 | 1786 | 2.69 | 1.28 | 192 | 0.00 | |
| Comparison driving skills | 2.34 | 0.91 | 595 | 2.39 | 0.77 | 52 | 0.00 | |
| Subjective driving difficulty | 1.50 | 1.19 | 596 | 1.58 | 1.21 | 52 | 0.00 | |
| Simulator risk mitigation score (1—10) | 5.50 | 0.67 | 806 | 5.44 | 0.61 | 97 | 0.00 | |
| Simulator safety score (1—10) | 7.37 | 1.05 | 806 | 7.44 | 0.87 | 97 | 0.00 | |
| Simulator driving skill score (1—10) | 7.87 | 1.09 | 806 | 7.96 | 1.04 | 97 | 0.00 | |
| Simulator eco driving score (1—10) | 7.58 | 1.52 | 806 | 7.88 | 1.47 | 97 | 0.00 | |
| Simulator field of view score (1—10) | 5.53 | 2.06 | 91 | 4.60 | 1.73 | 6 | 0.00 | |
| Number of simulator training hours | 5.93 | 3.06 | 1396 | 6.22 | 2.86 | 149 | 0.00 | |
| Number of on-road training hours | 40.93 | 18.93 | 1733 | 39.57 | 14.52 | 191 | 0.00 | |
| Number of total training hours | 47.46 | 19.12 | 1371 | 46.62 | 15.04 | 149 | 0.00 | |
| Number of driving test attempts | 1.68 | 0.98 | 1785 | 1.65 | 0.95 | 192 | 0.00 | |
| Passed first exam (0 = no, 1 = yes) | 0.57 | 0.50 | 1785 | 0.58 | 0.49 | 192 | 0.00 | |
| Violations first 12 months (0—20) | 1.99 | 2.11 | 1699 | 2.57 | 2.56 | 185 | −0.27 | **0<.0001** |
| Errors first 12 months (0—12) | 1.57 | 1.22 | 1678 | 1.71 | 1.15 | 180 | 0.00 | |
| Km driven first 12 months | 4225 | 6941 | 1735 | 5385 | 6771 | 183 | −0.17 | **0.03** |

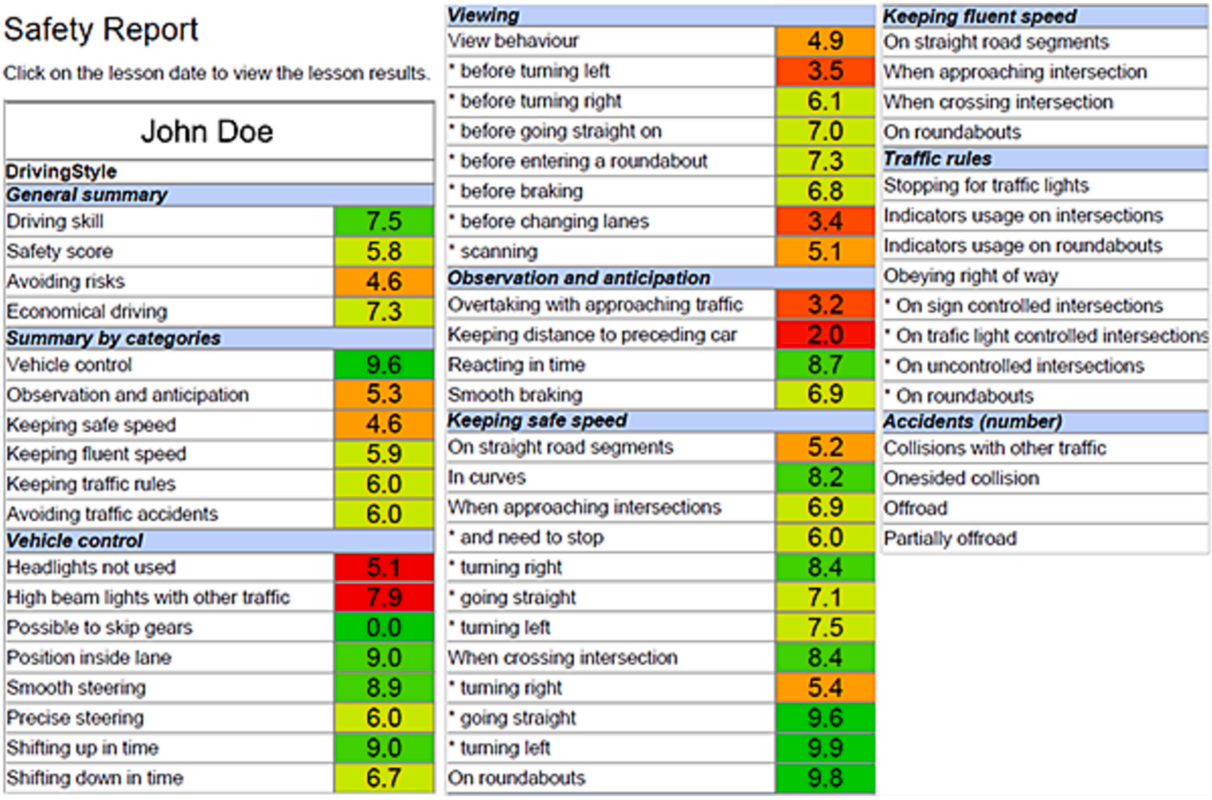
**Table 9**
Differences between drivers involved in an accident (Yes) and those not involved in an accident (No) in the last twelve months. Bold values are significant ($p < 0.05$).

| | No<br>Mean | *SD* | *n* | Yes<br>Mean | *SD* | *n* | Cohen's *dr* | *p* |
|---|---|---|---|---|---|---|---|---|
| HDTT/HATT (0 = no, 1 = yes) | 0.60 | 1.20 | 368 | 0.53 | 1.16 | 45 | 0.00 | |
| Gender (0 = male, 1 = female) | 0.61 | 0.49 | 1404 | 0.55 | 0.50 | 128 | 0.00 | |
| Licencing age | 21.23 | 5.21 | 1396 | 21.13 | 4.66 | 128 | 0.00 | |
| Driving anxiety before driving education | 2.75 | 1.25 | 1408 | 2.62 | 1.34 | 128 | 0.00 | |
| Comparison driving skills | 2.35 | 0.89 | 528 | 2.33 | 0.90 | 42 | 0.00 | |
| Subjective driving difficulty | 1.49 | 1.18 | 529 | 1.41 | 1.21 | 42 | 0.00 | |
| Simulator risk mitigation score (1—10) | 5.53 | 0.65 | 622 | 5.48 | 0.59 | 69 | 0.00 | |
| Simulator safety score (1—10) | 7.49 | 1.02 | 622 | 7.63 | 0.94 | 69 | 0.00 | |
| Simulator driving skill score (1—10) | 7.94 | 1.08 | 622 | 8.08 | 1.07 | 69 | 0.00 | |
| Simulator eco driving score (1—10) | 7.70 | 1.50 | 622 | 7.60 | 1.65 | 69 | 0.00 | |
| Simulator field of view score (1—10) | 5.47 | 1.84 | 74 | 5.15 | 2.43 | 8 | 0.00 | |
| Number of simulator training hours | 6.02 | 3.03 | 1053 | 6.00 | 2.62 | 93 | 0.00 | |
| Number of on-road training hours | 40.00 | 17.40 | 1364 | 44.55 | 24.46 | 125 | −0.25 | **0.04** |
| Number of total training hours | 46.94 | 17.98 | 1034 | 50.57 | 22.67 | 92 | 0.00 | |
| Number of driving test attempts | 1.67 | 0.96 | 1407 | 1.73 | 1.03 | 128 | 0.00 | |
| Passed first exam (0 = no, 1 = yes) | 0.57 | 0.50 | 1407 | 0.56 | 0.50 | 128 | 0.00 | |
| Violations last 12 months | 2.85 | 2.65 | 517 | 4.12 | 3.12 | 41 | −0.47 | **0.01** |
| Errors last 12 months | 1.06 | 1.12 | 518 | 1.71 | 1.25 | 41 | −0.58 | **0<.0001** |
| Km driven last 12 months | 7911 | 9483 | 568 | 12,863 | 9843 | 47 | −0.52 | **0<.0001** |

**Fig. 2.** Driving simulator types “Classic” and “Drive Master B”.

score than drivers not involved in an accident in both the first ($p = 0.003$, $d = -0.27$) (Table 8) and the last twelve months ($p = 0.015$, $d = -0.47$) (Table 9). Moreover, a significant difference in error scores in the last twelve months ($p = 0.002$, $d = -0.58$) indicates that drivers involved in an accident made more errors than those not involved. This effect was not observed for accident risk (column 16, row 18) (Tables 4 and 7).

No significant differences were found in the violation scores and the subjective driving skill scores between drivers involved and not involved in an accident in the first twelve months and last twelve months (Tables 8 and 9), suggesting that experiencing an accident did not affect



Safety Report

Click on the lesson date to view the lesson results.

John Doe

| DrivingStyle | |
|---|---|
| ***General summary*** | |
| Driving skill | 7.5 |
| Safety score | 5.8 |
| Avoiding risks | 4.6 |
| Economical driving | 7.3 |
| ***Summary by categories*** | |
| Vehicle control | 9.6 |
| Observation and anticipation | 5.3 |
| Keeping safe speed | 4.6 |
| Keeping fluent speed | 5.9 |
| Keeping traffic rules | 6.0 |
| Avoiding traffic accidents | 6.0 |
| ***Vehicle control*** | |
| Headlights not used | 5.1 |
| High beam lights with other traffic | 7.9 |
| Possible to skip gears | 0.0 |
| Position inside lane | 9.0 |
| Smooth steering | 8.9 |
| Precise steering | 6.0 |
| Shifting up in time | 9.0 |
| Shifting down in time | 6.7 |

| ***Viewing*** | |
|---|---|
| View behaviour | 4.9 |
| * before turning left | 3.5 |
| * before turning right | 6.1 |
| * before going straight on | 7.0 |
| * before entering a roundabout | 7.3 |
| * before braking | 6.8 |
| * before changing lanes | 3.4 |
| * scanning | 5.1 |
| ***Observation and anticipation*** | |
| Overtaking with approaching traffic | 3.2 |
| Keeping distance to preceding car | 2.0 |
| Reacting in time | 8.7 |
| Smooth braking | 6.9 |
| ***Keeping safe speed*** | |
| On straight road segments | 5.2 |
| In curves | 8.2 |
| When approaching intersections | 6.9 |
| * and need to stop | 6.0 |
| * turning right | 8.4 |
| * going straight | 7.1 |
| * turning left | 7.5 |
| When crossing intersection | 8.4 |
| * turning right | 5.4 |
| * going straight | 9.6 |
| * turning left | 9.9 |
| On roundabouts | 9.8 |

| ***Keeping fluent speed*** |
|---|
| On straight road segments |
| When approaching intersection |
| When crossing intersection |
| On roundabouts |
| ***Traffic rules*** |
| Stopping for traffic lights |
| Indicators usage on intersections |
| Indicators usage on roundabouts |
| Obeying right of way |
| * On sign controlled intersections |
| * On trafic light controlled intersections |
| * On uncontrolled intersections |
| * On roundabouts |
| ***Accidents (number)*** |
| Collisions with other traffic |
| Onesided collision |
| Offroad |
| Partially offroad |

**Fig. 3.** Safety Report.

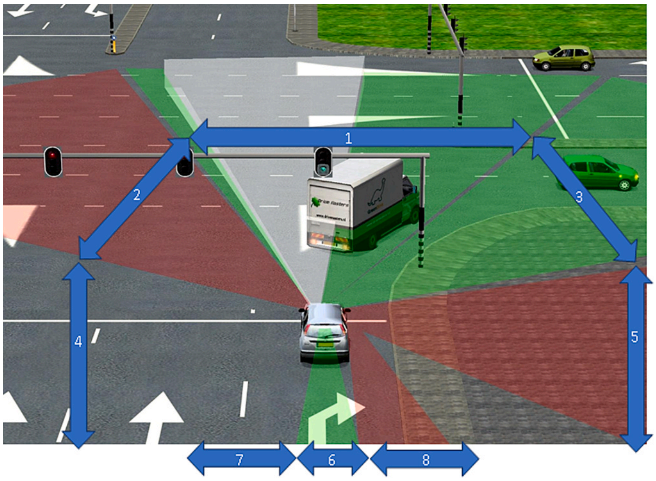


**Fig. 4.** Detection areas: fields of view (Source: Green Dino). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

one's risk-taking behaviour and subjective driving proficiency.

The HAT and control groups did not differ significantly in accident involvement or accident risk (Table 3).

### 4.2.6. *Summary of unsupervised driving results*

The data analysis reveals significant differences and correlations among measures.

Violation score: (1) A higher violation score in the first 12 months correlates positively with being male, younger licensed, less anxious, less fuel-consuming, and less risk mitigating. (2) Drivers judged by the instructor (not the examiner) to be ready for driving unsupervised in an

**Fig. 5.** Graphical user interface with red areas marking detection faults (Source: Green Dino). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Fig. 6.** Hazard anticipation training example (Source: Green Dino). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

earlier phase of education (they needed fewer lessons) reported more violations. (3) Drivers with a higher violation score drove more kilometers, made more errors and were more likely to be involved in an accident. (4) Violations increased from the first to the second half of the year and from the first to the last year of driving. (5) Early-age licensees reported more violations in the last 12 months. (6) Drivers who consider themselves more competent and experience less effort operating a vehicle reported more violating behaviour.

Error score: (1) A higher error score correlates with higher driving anxiety, more kilometers and lower driving skills. (2) A higher error score correlates with a higher violations score, independent of the months of licensing. (3) Errors decreased from the first 12 to the last 12 months.

Driving skills & driving difficulty scores: (1) Males and early-age licensees saw themselves as more competent than their peers. They experienced less anxiety and followed fewer on-road training hours. (2) The comparison driving skills score correlates negatively with the subjective driving difficulty score in the first 12 months. Drivers who reported higher driving difficulty drove less in the first year. (3) Higher comparison driving skills and lower subjective driving difficulty scores correlate with a higher violation score in the last 12 months of driving. (4) The longer one was licensed, the more advanced one judged their own driving skills.

Accidents & accident risk: (1) Accident involvement remained the same over the first and last years of unsupervised driving. (2) Drivers reporting accidents drove more kilometers in the first twelve months and last twelve months of driving. (3) The risk of being involved in an accident dropped from the first six months to the second six months and from the first twelve to the last 12 months (non-significant). (4) Drivers reporting more risk-taking (higher violations score) had more accidents, independent of the first or last twelve months of driving. (5) Drivers involved in accidents reported more errors. (6) Experiencing an accident did not affect risk-taking behaviour and self-calibration. (7) Drivers reporting accidents in the last twelve months required more on-road training hours. (8) Faster learners (fewer on-road training hours) were less involved in accidents.

No significant effects or correlations are found for HAT for the unsupervised driving phase.

## 5. Discussion

Simulator manufacturer Green Dino provided a unique cohort dataset to study the long-term effects of a simulator hazard-awareness training program. Four hypotheses on benefits and obstacles were analysed. The results showed positive effects on supervised driving, but not on unsupervised driving. The results reveal that hazard detection and anticipation are trainable skills, but emotions and motivations seem to temper the long-term effects of HAT after licensing.

### *5.1. Hazard awareness: trainable skills*

For all respondents, simulator scores increased after the intersection training, which included hazard detection and anticipation instruction, as well as error feedback. It demonstrates the educational value of simulator-based training for acquiring driving skills. Risk mitigation, safety, vehicle handling, and eco driving skills improved significantly, with medium effect sizes.

For the HAT group, the field of view score also increased significantly, with a large effect size. This result supports Hypothesis 1: Hazard awareness training on simulators improves the viewing skills of learner drivers. The significant increase in the field of view score demonstrates that error feedback is crucial for maintaining and enhancing detection skills. Although Dutch driving students should have some prior knowledge of hazard detection at intersections from riding a bicycle in traffic, their viewing score remains relatively low after training ($M = 5.94$ out of 10). This suggests that these skills are challenging to acquire and remain in the early stages of development, warranting continued attention during on-road driving lessons.

It is unclear how much the HAT error feedback contributed to the overall learning effect in the HAT group. An experiment measuring students' field of view scores with and without HAT is necessary to gain a deeper understanding. Also, it is unclear if the control group improved their viewing skills between the two tests, as the simulator they trained on didn’t provide field of view scores.

The eco driving score is positively correlated with being female, higher driving anxiety, licensing age, and the risk mitigation score. This suggests that emotion and internal motivation positively affected fuel consumption among these students. In contrast, the HAT group students showed a negative correlation with the eco driving score during simulator training. This interpretation is further supported by the reliability analysis: the eco driving score showed a markedly lower item-total correlation with the other intersection test scores ($r = 0.13$), and removing it raised the internal consistency of the intersection test composite from $\alpha = 0.55$ to $\alpha = 0.73$. Eco driving therefore measures a distinct dimension of driving behaviour, and its lower scores in the HAT group should not be interpreted as contradicting the positive hazard-awareness findings. One plausible mechanism is that HAT students, exposed to crash scenarios emphasising speed management, adopted a more conservative speed approach that used more fuel, penalised by the eco-driving metrics (which reward low engine revolutions rather than cautious, speed-reducing driving).

Gender, licensing age, and driving anxiety correlated with detection and anticipation simulator scores. Emotion and motivation seem to influence hazard awareness from an early stage of education. These results support Hypothesis 4: Emotion and motivation influence hazard awareness.

Many factors influence the number of on-road training hours necessary to obtain unsupervised driving permission (Kuipers et al., 2023). This study showed that the HAT group needed significantly fewer on-road training hours (39.18) than the control group. The total education hours (46.63) was also lower. The number of on-road driving hours differed by 4.09 h, and the total education hours differed by 2.79 h. The control group needed 12.4% more on-road lessons, and its total education time was 6.8% longer. In addition, HAT students passed the driving exam in fewer attempts (1.56) than non-HAT students (1.78). The control group needed 14.1% more attempts. The HAT group's success rate on the first driving exam was higher (0.64) than the control group's (0.51). The control group's success on the first exam was 25.5% lower.

This demonstrates that the instructors' and examiners' judgements about safe driving performance were aligned: both were convinced earlier in driving education that the HAT group had sufficient safe driving skills for unsupervised driving compared to the control group. According to Anderson's (1982) learning theory, instructors and examiners judge the HAT group to have procedural-level cognitive skills earlier than the control group. HAT simulator training appeared to facilitate and reinforce the development of a safe driving style that educational experts recognised. These results support Hypothesis 2: Hazard training on simulators facilitated the acquisition of a safe driving style during the educational process.

No significant learning effects or correlation measures were found for HAT in unsupervised driving. The results do not support Hypothesis 3: Hazard awareness training on simulators stimulates a safer driving style after licensing. No evidence was found to support the assumption that HAT leads to safer driving after obtaining a license. The pronounced facilitating effect of HAT on driving education disappeared after obtaining a license.

Although HAT positively influenced driving education, HAT students did not show a safer driving style than the control group in the simulator intersection test. Performance in the early phase is identified as slow and error-prone (Fitts & Posner, 1967). Additionally, Anderson (1982) stated that cognitive task performance remains relatively unstable early in driving education, as the focus is consciously on isolated components of the driving task and possible strategies are tested and rejected. Once associations between the isolated components are formed and strengthened, procedures are generated that can be applied in traffic situations that are recognised as similar. This development takes time. Differences observed between the HAT and control groups in on-road training hours, and exam attempts suggest that the HAT students can proceed to the subsequent phases more quickly. However, this facilitating effect is only demonstrated during driving education, not when driving unsupervised.

The HAT scenarios were practised only once in an early phase of education, which may imply fewer benefits. The benefits of learning increase rapidly with repetition and stabilise after a specific period (Fitts & Posner, 1967). The positive effects of HAT training observed in this study could be reinforced through repetition and differentiation of the traffic scenarios. From a learning perspective, a single exposure to HAT fades without consolidation. Repetition of HAT scenarios — distributed across the driving curriculum rather than delivered before starting on-road driving lessons — would test whether the observed post-licensing null effect is a fundamental ceiling on hazard-awareness training or merely an artefact of insufficient practice exposure.

### 5.2. Risk Awareness: Emotional and motivational factors

Drivers with higher violation scores perceived themselves as better drivers than others and reported experiencing fewer driving difficulties. Lower subjective driving difficulty was correlated with greater distance driven in the last 12 months and with a judgment of having higher driving skills than others. The higher trust drivers have in their driving skills, the more they report violating behaviour.

Risk-seeking driving behaviour among all respondents increased after licensing and stabilised after more than 24 months of licensing. Drivers who obtained their licenses at a younger age reported more risk-taking behaviour. They were more confident and had more exposure. They also found themselves to be better drivers, which aligned with the assessments of the simulator, driving instructor, and examiner.

Significant negative correlations were found between the self-comparison of driving skills and gender, and driving anxiety. Also, a significant negative correlation was found between the comparison driving skills score and the subjective driving difficulty score. These results indicate that males, drivers licensed at a younger age, and less anxious drivers tend to perceive themselves as more competent drivers. Those who experienced more driving difficulty rated their driving skills lower.

Violating behaviour increased with experience. Accident involvement did not change self-assessments of task difficulty and safety performance. Koppel et al. (2022) replicated a study by Svenson (1981) and confirmed that drivers overestimate their driving skills and perceive themselves as less risky than their peers. Learning novice drivers to self-reflect (calibrate) appears to be very difficult, which aligns with the findings of De Craen (2010) and Kruger and Dunning (1999).

The learning effects and correlations related to emotion and motivation support Hypothesis 4: Emotion and motivation influence hazard awareness. These effects and correlations were already visible on the simulator and persisted after licensing, raising questions about the learning value of current-day driving education. That is, while known to be 'less trainable', perhaps more emphasis should be placed on mitigating the potential effects of personality traits (high confidence and self-appraisal, less risk aware) on driving safely when unsupervised.

Kuipers, De Winter and Mulder (2023), who studied the same cohort, found that fear of driving before starting simulator lessons provided a logical explanation for students' progression through driver education. They advised using information-processing styles, incorporating fear, and education level in driving education. The number of accidents per 1,000 km was 39% higher among lower-educated individuals. This suggests investigating whether there is a correlation among education level, HAT, and accident risk.

### 5.3. Safety margin

The length of driver's license possession and driving kilometres were positively correlated with violations and errors. On average, respondents lowered their safety margins as they gained more experience. Violations in the first 12 months were positively correlated with errors during the same period, suggesting a more accident-prone driving style. The correlation between violations and errors is even more pronounced over the last 12 months of driving, indicating lower safety margins. HAT did not suppress this safety-lowering tendency. These results do not support Hypothesis 3: Hazard awareness training on simulators stimulates a safer driving style after licensing. It is worth noting that these correlations are likely conservative estimates of the true underlying relationships. Because the violation ($\alpha = 0.55$) and error ($\alpha = 0.44$) composites carry moderate measurement error, correlations involving them are systematically attenuated — pulled toward zero by noise in the scores. Statistically, the observed correlation approximates the true relationship multiplied by the square root of the scale's reliability. Applied to the violation–error relationship, the observed growth from $r = 0.18$ in the first twelve months to $r = 0.32$ in the last twelve months corresponds to estimated true correlations of approximately 0.29 and 0.52, respectively, after correcting for attenuation. This suggests that the co-occurrence of violations and errors strengthens substantially with driving experience — from a weak to a strong association — and that the safety margin erosion documented in this study is likely more pronounced in reality than the observed coefficients indicate. Similarly, the correlations between violations and accident involvement ($r = 0.09$) and between errors and driving anxiety ($r = 0.14$) should be understood as floor estimates; the true relationships are plausibly stronger. These scale limitations apply symmetrically across all post-licensing analyses. Where any effect — including HAT — was small or non-significant, the moderate reliability of these composites means the true effect size may be somewhat larger than observed, though the consistency of null findings across multiple outcomes makes a substantially masked effect unlikely.

Drivers who had a riskier driving style in the first twelve months drove more kilometres, made more errors and were also more likely to be involved in an accident. This group was identified as faster learners and required fewer lessons in driving education. Although the faster learners anticipated better and were better at handling the vehicle in the simulator, they drove with lower safety margins than other drivers, as indicated by the risk-mitigating and eco-driving scores. It appears that a more fluent driving style, characterised by greater hazard awareness rather than a risk-averse approach, is rewarded by the simulator and instructors. Over the last twelve months, the difference in violating behaviour disappeared, as all drivers reported an increase in violating behaviour. These findings support Hypothesis 4: Emotion and motivation influence hazard awareness.

Currently, driving instructors prepare their students to pass the driving test (Roemer, 2021) and instruct them to drive fluently, following the national Driving Procedure B (car) guidelines. This changes the risk-avoiding driving style learned in the simulator to a more accident-prone one, reducing the time to detect and anticipate, and increasing the impact of a crash. Teaching and permitting lower safety margins (e.g., shorter following distance, higher speed, and a narrower field of view at intersections) might reinforce violating behaviour while driving unsupervised. It is also possible that instructors and examiners — being highly skilled drivers themselves — misjudge students' risk detection and anticipation skills by reference to their own automated, lower-safety-margin driving style (cf. Kruger & Dunning, 1999, on the difficulty skilled performers have in recognising others' incompetence). This could mean that instructors inadvertently students' risk detection and anticipation skills, driven by their own high competencies in driving safely, which may involve operating with lower safety margins.

### 5.4. Limitations

Potential self-selection effects are limited in the present study. Participation in the hazard awareness training was not voluntary but occurred as part of simulator-based driver education. Students, therefore, did not actively choose whether to participate in the training. In addition, there was no option for students to choose whether eye-tracking (gaze registration) was available; it depended solely on the simulator's technical configuration during the lesson. Furthermore, all participants followed the same core training modules, including intersection training and vehicle control exercises. As a result, exposure to these fundamental training components was comparable across participants. The participating driving schools were distributed across the Netherlands, supporting the sample's geographical diversity, although one driving school contributed a relatively large share of participants. The respondent sample was relatively highly educated and included a larger proportion of women than typically observed in the general driver population. To assess potential bias, baseline simulator performance and demographic characteristics were compared between the Hazard Awareness Training (HAT) group and the control group. These comparisons indicated that the groups were comparable at the start of the study.

The sample exhibited notable distributional skews: females were overrepresented relative to the general learner-driver population, participants tended to have higher education levels, and a disproportionate share of the data came from a single driving school. Education level was excluded from the analyses entirely because its strongly skewed distribution precluded meaningful subgroup comparisons. These compositional features may limit the generalisability of the findings to broader and more diverse populations of novice drivers.

The primary behavioural outcomes were measured using an in-house questionnaire that has independently validated. Unlike the DBQ or similar instruments, the psychometric properties of this instrument are unknown beyond the internal consistency estimates reported here and by Kuipers, De Winter and Mulder (2023). Several key outcomes — including kilometres driven, traffic violations, and driving errors — also relied on self-report. Self-reported driving behaviour has been disputed as a reliable measure (Bailey & Wundersitz, 2019). The extended interval between driving education and inventory completion further reduces memory accuracy, particularly for on-road training hours and early post-licensing kilometres. Deviations in reported kilometres were considerable, which undermines the reliability of accident risk calculations. A more frequent, prospective measurement approach — for example, monthly inventories — would improve the accuracy of these measures.

The statistical approach carries several limitations. Group differences were analysed using independent samples t-tests and one-way ANOVA; within-person change over time was examined with paired samples t-tests; and associations between categorical variables were assessed with chi-square tests. Where distributional assumptions were violated, non-parametric equivalents were applied (Mann-Whitney U, Kruskal-Wallis). However, count-based outcome variables such as violation and error scores are inherently right-skewed, meaning parametric tests were not always the most appropriate choice even in larger subgroups. Furthermore, the study computes numerous correlations across large matrices (Tables 2-7) without correction for multiple comparisons (e.g., Bonferroni or false discovery rate). This substantially increases the risk of Type I errors, and correlations near the significance threshold should therefore be interpreted with caution. Additionally, the composite scores for violations ($\alpha = 0.55$) and errors ($\alpha = 0.44$) showed moderate internal consistency, which is expected for behavioural frequency scales measuring distinct risky acts. However, this level of reliability introduces measurement error that attenuates observed correlations with other variables. Findings involving these composites — particularly small or non-significant effects — may therefore underestimate true associations rather than reflect a genuine absence of effect.

Beanland et al. (2013) identified structural obstacles to demonstrating safety effects: limited suitable datasets, data noise, subjectivity in outcome measures, and weak research designs. This study faces several such obstacles. Accident numbers were low, and self-assessed accident responsibility is inherently uncertain, as it is questionable whether drivers can correctly assess their own culpability. Research using police-documented accidents could reduce these biases and improve the significance and validity of safety-related findings. The absence of a randomised controlled trial — as in the classical DeKalb study (Lund et al., 1986; Stock et al., 1983) — further limits causal inference; such designs are now exceedingly difficult to implement due to ethical and regulatory constraints.

## 6. Conclusions

This study examined the long-term effects of a simulator hazard awareness training program (HAT) on supervised and unsupervised driving. Hypotheses were formulated assuming positive learning effects of HAT for three phases of driving: (1) in the simulator, supervised by AI (2) on the road, supervised by instructors and examiners, and (3) on the road, unsupervised. Furthermore, it was postulated that (4) emotion and motivation affect hazard awareness.

Results show that HAT had a significant positive learning effect for supervised driving in the simulator and on the road, but this effect disappeared when driving unsupervised after licensing. The asymmetry between supervised and unsupervised outcomes is itself a substantive finding: it confirms Vlakveld's (2011) theoretical distinction between trainable hazard awareness and dispositional risk awareness, and it establishes that formal driving education operates primarily on the former while the latter remains largely stable across the training period. In contrast, factors related to emotion and motivation affected hazard awareness across all three phases of driving. This study confirms earlier findings by Kuipers, De Winter, and Mulder (2023) that personal characteristics explain progress in education and in unsupervised driving.

Further research should explore whether repeating HAT across the curriculum can yield more durable effects on hazard detection and anticipation skills.

The results support the recommendations of the Dutch National Curriculum Driver Training and the European Commission to use driving simulators for hazard awareness training. HAT demonstrably accelerates competence development during supervised driving and reduces the resources needed to reach licensing readiness. Nevertheless, the findings consistently show that intrinsic personal characteristics — gender, licensing age, fear of driving, and self-assessed competence — have a more substantial and lasting influence on post-licensing safety behaviour than the type of training. The safety benefits of HAT are real but bounded: apparent through the driving exam, they do not persist once drivers operate unsupervised. For practitioners, this implies that reducing accident risk in novice drivers requires addressing motivational and emotional factors alongside skill training. Explicitly teaching and reinforcing higher safety margins throughout driving education — rather than rewarding driving fluency — may produce more durable safety benefits. Further research should investigate whether repeated HAT, application of higher safety margins in education, or interventions targeting self-calibration and risk perception can extend the benefits of hazard awareness training into the unsupervised driving phase.

## CRediT authorship contribution statement

**Jorrit Kuipers:** Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization. **Max Mulder:** Writing – review & editing, Supervision. **Maran Voskes:** Writing – review & editing.

## Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Kuipers reports financial support and statistical analysis were provided by Green Dino BV. Kuipers reports a relationship with Green DIno BV that includes: board membership, employment, and non-financial support. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper..

## Acknowledgements

The Dutch Ministry of Economic Affairs and Green Dino BV supported this study.

## Appendix 1

### *Driving simulators*

Between 2008 and 2023, Dutch driving schools used four types of driving simulators (Fig. 1). Whereas considerable differences existed in image projection methods (projectors “Classic” or screens”Drive Master B”) and control mechanisms (car parts or Logitech products), all simulators ran identical software applications and curricula.

The Green Dino simulator provides automated adaptive instruction using an AI (virtual) driving instructor, which supplies automated performance feedback (Weevers et al., 2003; Fikkert et al., 2006). The simulator substitutes both the training vehicle and the human driving instructor. The thirty-four 30-minute curriculum lessons address vehicle handling, intersection navigation, highway driving, and manoeuvring. The simulator is used early in driver training, typically before students start lessons in an on-road vehicle.

The research presented in this paper utilised scores from one of the summary reports (Fig. 2), which is frequently used to brief students, their parents, and instructors about their progress in learning driving skills. The report includes composite scores categorised under “driving skills score” and “safety score”. In short, the driving skill score is a composite score derived from combining various task scores. These include excessive braking or collisions, improper use of turn signals, swerving or deviating from the centre of the road, and operating at inconsistent engine rpm within the simulator. The safety score is based on the degree to which the student driver exceeded speed limits or maintained an insufficient distance from the vehicle ahead in the simulator, amongst other safety-related behaviours. The scores are explained in more detail in Appendix II.

### *Hazard Detection Training and Testing (HDTT)*

For the training and assessment of hazard detection, eight fields of view are differentiated (Fig. 4): straight forward, left forward, right forward, left, right, straight backwards (using the interior mirror), left backwards (using the left-side mirror), and right backwards (using the right-side mirror). These fields are related to the Dutch national car driving procedures for crossing lanes. HDTT instructs where to look depending on the active driving task. It doesn’t instruct on what to see or how to anticipate, because that is too cognitively demanding.

Fig. 3 shows an example of a detection assessment for turning right; here, the driver looked at the green fields 1, 3, and 6 and did not (or not for long enough) view the red fields 2, 5, and 8. Fields 4 and 7 are not needed as the task involved turning right. The white area represents the learner driver's current field of view. In the graphical user interface of the driving simulator, red regions are used, along with icons, text, and audio, to inform the learner driver about any procedural detection faults (Fig. 5).

The automated detection instruction trains the learner to release the gas pedal long before entering an intersection and scan specific areas of the surroundings related to the activated procedure. The training partly assesses the processing of the information visible in the target field. Feedback about the appropriate approaching speed, following distance, stopping distance and time to give right of way is provided. However, there is no direct connection between detection errors and these procedural faults because the detection of scanning errors was implemented later and not fully integrated. When approaching an intersection, feedback on the car’s velocity directs the learner driver to slow down and create more time to attend to the required areas of view in the correct sequence.

The post-evaluation safety report presents scores for approaching speed and viewing procedures, integrating these into scores for driving skill, safe driving, risk avoidance, and eco driving (Fig. 3).

### *Hazard Anticipation Training and Testing (HATT)*

In 2013, Green Dino released hazard anticipation error training and testing based on Vlakveld (2011). In the HATT, five hazard scenarios are simulated which provoke crashes with other cars. The five scenarios are: 1) overtaking on an 80 km/h road, 2) covering a situation at an intersection (Fig. 6), 3) covering a situation at a T-junction, 4) passing a truck on a too-narrow road, (5) two confrontations with an ambulance with alarm signals. After crashing, an explanation follows, including correct detection and anticipation instructions. Fig. 5 shows a scenario where the red truck blocks the grey car driver’s view of a green approaching car. The driver entered the crossing and tried to look behind the truck. A crash with the upcoming car is unavoidable. During the post-evaluation in the simulator, the driver receives an overview of the hazardous situation, the crash and the hazard precursor. The driver is instructed to increase safety margins by stopping before the intersection and waiting until the truck has left. All detection and anticipation steps are mentioned and visualised. Then, in the second trial, the driver can try again, slow down and stop before entering the intersection and experience the positive effects of using higher safety margins. At the end of the Hazard Anticipation training, the learner driver follows a 6.5-minute Driving Style Test (De Winter & Kuipers, 2017) and receives safety scores related to the shown driving style. The scores support the instructor in making an education plan. Some driving schools use the Driving Style Test to assess whether a student is ready to drive unsupervised.

HDTT and HATT reinforce each other. HDTT instructs on the speed and viewing procedure for approaching and overtaking hazardous situations, such as intersections. It teaches students how to detect dangerous situations: slowing down long before entering a complex traffic situation by releasing the gas pedal (instead of braking just before entering). HATT increases knowledge about the potential consequences of errors in executing the procedures (smooth driving) when approaching potentially hazardous situations. It targets the internal motivation of the learner driver to be more careful and slow down or even stop before entering a complex situation to allow oneself more time to detect and anticipate (potential) hazards.

## Appendix II

### *Simulator assessment measures*

For driving style assessment, Green Dino released the Safety Risk Framework© in 2007. This framework aims to provide both quantitative and qualitative ways of describing safety risks. The time when the safety risk is possible is taken into account for calculating safety risk values. By applying ‘smooth penalties’, the safety risk values are modulated proportionally to the extent of exceeding the minimal and maximal thresholds. The driver is penalised with a proportional value starting at 10, dropping to 0. Note that the different scores contain similar measures and, therefore, are not independent. The names of the measurements are derived from the Safety Risk Framework.

*Risk mitigation score (0–10)*

The risk mitigation score is created to quantify risk awareness. This score is a combined score based on scores for (1) approaching intersection too fast, (2) crossing intersection too fast, (3) driving too fast, (4) approaching intersection too slow, (5) crossing intersection too slow, (6) driving too slow, (7) ignoring right of way, (8) ignoring traffic rules, (9) inappropriate distance, (10) bad engine rpm, (11) inappropriate speed, and (12) inappropriate steering. The mean of all scores, except driving too slowly, lowers the avoiding risk score. The driving too slow score is inverted, meaning a higher score (i.e., when a student drives too slowly) decreases the risk mitigation score less. The share for driving too slow in the risk mitigation score is 50%.

The scores for driving too slowly are calculated with the optimal speed, current speed, acceleration, and distance from the intersection. For going straight through an intersection, the optimal speed is 40 km/h; for turning right, the optimal speed is 20 km/h; and for turning left, the optimal speed is 30 km/h. The smooth penalties are calculated between 5 km/h and 15 km/h below the optimum speed.

Note that the 'too low' scores were designed to identify cautious drivers who did not drive smoothly, like anxious and elderly drivers. This study classifies them as the more risk-aware drivers, who use higher safety margins.

*Safety score (0–10)*

The safety score is created to quantify anticipation skills.

This score is a combined score based on scores for (1) bad engine rpm, (2) ignoring traffic rules, (3) inappropriate distance, (4) inappropriate speed, (5) inappropriate steering, and (6) traffic accidents (collisions). The mean of all scores lowers the safety score, including driving too slow. The score for inappropriate speed combines driving too fast and driving too slowly. In contrast to the risk mitigation score, the share for driving too slow in the safety score is 8.7%. Huizinga et al. (2019) concluded, after clinical testing, that the driving safety score provided by the Safety Risk Framework reflects a rich composite of driving behavioural factors for assessing driving under the influence of medication and drugs.

*Driving skill score (0–10)*

The driving skill score is created to quantify vehicle handling skills.

This score combines scores for (1) bad vehicle control, (2) ignoring traffic rules, and (3) traffic accidents. Bad vehicle control is a combined score of bad rpm and inappropriate steering. Ignoring traffic rules is a combined score that includes ignoring the right-of-way, ignoring traffic lights, incorrect indicator usage, and incorrect indicator usage on roundabouts. Traffic accidents is a combined score of braking too hard, dynamic or static collision, driving off-road entirely or partially.

*Eco driving score (0–10)*

The eco-driving score is designed to quantify fuel-efficiency skills. Eco-driving is also considered a safety-related factor in driving education. Drivers prevent braking by anticipating speed changes early. They enhance detection and anticipation skills, such as with hazard awareness.

This score is a combined score based on scores for (1) engine rpm too high and (2) large braking energy losses. Engine rpm too high is the combined highest and lowest max rpm score. The lowest rpm is multiplied by 1.2.

*Field of view score (0–10)*

The field of view score is created to quantify detection skills. This score is a combination of scores for (1) inadequate scanning behaviour and (2) bad viewing behaviour for intersections. Inadequate scanning is the combined score of not observing the windscreen adequately and not checking the inner rearview mirror before making a speed adjustment. Bad viewing behaviour at intersections is the combined score for not checking the windscreen, mirrors, and left and right windows while performing the activated driving task.

Scanning behaviour is assessed by measuring the driver's face orientation using a webcam and a head-tracking application based on OpenCV and Dlib (shape_predictor_68_face_landmarks.dat), which synchronises the driver's head movements with the traffic simulation.

*Self-assessment measures*

*Driving Anxiety score (1–5)*

The driving anxiety score is based on the respondents' answers to the question "Were you scared to drive a car when you started your driving lessons?" (not at all (1), barely (2), a little (3), quite (4), very much (5)).

*Violations*

The violation score is the sum of responses to five statements: How often did you? (1) drive more than 10 km/h faster than the speed limit inside city limits, (2) use your mobile phone to read/send a text, (3) (intentionally) crossed a red light, (4) drive after drinking alcohol, (5) drive without using the safety belt. Respondents could choose between seven answers: (1) never (score = 0), (2) rarely – less than one time a month (score = 1), (3) sometimes – approximately once a month (score = 2), (4) frequently – approximately once a week (score = 3), (5) very frequently – multiple times a week (score = 4), (6) almost every time I drive (score = 5), and (7) I cannot remember. The score is calculated for the first six months, the second six months, the first year and the last twelve months of driving. Questions regarding the last twelve months were included only in the 2023 sample questionnaire, so there are no violation scores for the last twelve months in the 2015 sample. Violation scores were not generated for respondents who answered 'I cannot remember' to one of the statements, as their scores would not be representative.

The violation scores are likely to be low. It is unlikely that drivers exhibit all these risky driving behaviours every time they get into a car. There will be drivers who sometimes drive after consuming alcohol, but they very improbably drink every time they drive. Nevertheless, differences in scores are interesting as they could demonstrate changes in risk-taking behaviour.

*Errors*

The error score is the sum of responses to three statements: How often did you? (1) have to make an emergency stop because you were driving too close to the car in front, (2) get off-road, for example, in the verge, against the sidewalk, or deviate into the wrong lane, and (3) not give right of way where you should have. Respondents could choose between the seven options, similar to the choices for violations. The score is calculated for the first six months, the second six months, the first year and the last twelve months of driving. Again, questions regarding the last twelve months were only distributed to the 2023 survey, so there are no error scores for the last twelve months in the 2015 sample. Error scores were not generated for respondents who answered with 'I cannot remember' to one of the statements since their scores would no longer be representative.

Again, these scores are not likely to be very high. It is not plausible that drivers make these errors every time they get into the car. For example, drivers are unlikely to get off the road for all their rides. Nevertheless, differences in scores are interesting because they indicate changes in driving proficiency.

*Subjective driving skill competence*

Two statements were used to measure subjective driving skill competence: (1) "I think I am a more skilled driver than other drivers", and (2) "I sometimes experience difficulties driving". Respondents could answer these two questions with "yes" or "no". These statements were added to the 2023 questionnaire; therefore, this information is only available for drivers in the 2023 sample. The two statements will be labelled "Comparison driving skills" and "Subjective driving difficulty".

*Accidents*

Accident involvement is a binary variable with ‘no’ and ‘yes’ as answers. The accident risk was calculated to correct for exposure. The following formula was used: Accident risk per 1,000 km = (Number of accidents/distance driven in km) × 1,000.

## Data availability

The data that has been used is confidential.

## References

Anderson, J.R., 1982. Acquisition of cognitive skill. Psychol. Rev. 89 (4), 369–406. https://doi.org/10.1037/0033-295X.89.4.369.

Bailey, T., & Wundersitz, L. (2019). The relationship between self-reported and actual driving-related behaviours: A literature review. University of Adelaide. Retrieved from https://casr.adelaide.edu.au/casrpubfile/2280/CASR162.pdf.

Beanland, V., Goode, N., Salmon, P.M., Lenné, M.G., 2013. Is there a case for driver training? a review of the efficacy of pre- and post-licence driver training. Saf. Sci. 51, 127–137. https://doi.org/10.1016/j.ssci.2012.06.021.

Damasio, A.R., 1996. The somatic marker hypothesis and the possible functions of the prefrontal cortex [and discussion]. Philos. Trans.: Biol. Sci. 351 (1346), 1413–1420.

De Craen, S. (2010). The X-factor: A longitudinal study of calibration in young novice drivers (Doctoral dissertation). SWOV, Leidschendam.

De Winter, J., & Kuipers, J. (2017). Relationships between years of licensure and driving style measured with a short simulator-based test (N=650). In Stanton, N., Landry, S., Di Bucchianico, G., & Vallicelli, A. (Eds.), Advances in Human Aspects of Transportation (pp. 487–494). Springer. https://doi.org/10.1007/978-3-319-41682-3_54.

ECMT. (2006). Young Drivers: The Road to Safety. OECD/ECMT Transport Research Centre, Paris.

European Commission. (2023). Revision of the Driving Licence Directive. Retrieved from https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/749788/EPRS_BRI(2023)749788_EN.pdf.

Fikkert, W., Heylen, D., Van Dijk, B., Nijholt, A., Kuipers, J., & Brugman, A. (2006). Estimating the gaze point of a student in a driving simulator. In Proceedings of the Sixth IEEE International Conference on Advanced Learning Technologies (pp. 497–501). IEEE. doi: 10.1109/ICALT.2006.1652484.

Fitts, P. M., & Posner, M. I. (1967). Human performance. Brooks/Cole.

Huizinga, C.R., Zuiker, R.G., de Kam, M.L., et al., 2019. Evaluation of simulated driving in comparison to laboratory-based tests to assess the pharmacodynamics of alprazolam and alcohol. Journal of Psychopharmacology 33 (7), 791–800. https://doi.org/10.1177/0269881119836198.

Koppel, S., Charlton, J.L., Kopinathan, C., Stephens, A.N., 2022. Drivers' self-perceptions of driving ability: a replication and extension of Svenson's (1981) study. Transportation Research Part f: Traffic Psychology and Behaviour 89, 1–11. https://doi.org/10.1016/j.trf.2022.07.001.

Kruger, J., Dunning, D., 1999. Unskilled and unaware of it: how difficulties in recognising one's own incompetence lead to inflated self-assessments. Journal of Personality and Social Psychology 77 (6), 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121.

Kuipers, J., De Winter, J., Mulder, M., 2023. From fear to forecast: the role of simulators, accompanied driving, age, gender, and information-processing style in driver training and beyond. Transportation Research Part f: Traffic Psychology and Behaviour 99, 389–407. https://doi.org/10.1016/j.trf.2023.10.003.

Lund, A.K., Williams, A.F., Zador, P., 1986. High school driver education: further evaluation of the DeKalb County study. Accident Analysis & Prevention 18 (5), 349–357. https://doi.org/10.1016/0001-4575(86)90048-5.

McDonald, C.C., Goodwin, A.H., Pradhan, A.K., Romoser, M.R., Williams, A.F., 2015. A review of hazard anticipation training programs for young drivers. Journal of Adolescent Health 57 (1 Suppl), S15–S23. https://doi.org/10.1016/j.jadohealth.2015.02.013.

McKnight, A.J., McKnight, A.S., 2003. Young novice drivers: Careless or clueless? Accident Analysis & Prevention 35 (6), 921–925. https://doi.org/10.1016/S0001-4575(02)00100-8.

Omran, Y.H., Sadeghi-Bazargani, H., Yarmohammadian, M.H., Atighechian, G., 2023. Driving hazard perception tests: a systematic review. Bulletin of Emergency and Trauma 11 (2), 51–68. https://doi.org/10.30476/beat.2023.95777.1370.

Prabhakharan, P., Bennett, J.M., Hurden, A., Crundall, D., 2024. The efficacy of hazard perception training and education: a systematic review and meta-analysis. Accident Analysis and Prevention 202, 107554. https://doi.org/10.1016/j.aap.2024.107554.

Roemer, E. (2021). Van rijles naar rijonderwijs: Advies verbetering autorijscholenbranche. Ministry of Infrastructure and Water Management. Retrieved from https://www.rijksoverheid.nl/documenten/rapporten/2021/04/14/bijlage-1-van-rijles-naar-rijonderwijs-advies-verbeteren-autorijscholenbranche.

Roelofs, E., Vissers, J., & Tsapi, A. (2023). Kerndoelen voor het nationaal leerplan rijopleiding B. Ministry of Infrastructure and Water Management. Retrieved from https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/rapporten/2023/10/20/bijlage-3-kerndoelen-nationaal-leerplan-rijopleidingen/bijlage-3-kerndoelen-nationaal-leerplan-rijopleidingen.pdf.

Stock, J. R., Weaver, J. K., Ray, H. W., Brink, J. R., & Sadof, M. G. (1983). Evaluation of Safe Performance Secondary School Driver Education Curriculum Demonstration Project. U.S. Department of Transportation, National Highway Traffic Safety Administration. Report No. HS-806 568.

Svenson, O., 1981. Are we all less risky and more skilful than our fellow drivers? Acta Psychologica 47 (2), 143–148. https://doi.org/10.1016/0001-6918(81)90005-6.

SWOV. (2019). Driver training and driving tests [Fact sheet]. SWOV, The Hague. Retrieved from https://swov.nl/sites/default/files/bestanden/downloads/FS%20Driver%20training.pdf.

SWOV. (2021). Young drivers [Fact sheet]. SWOV, The Hague. Retrieved from https://swov.nl/sites/default/files/bestanden/downloads/FS%20Young%20drivers_2.pdf.

Vlakveld, W., 2011. Hazard anticipation of young novice drivers: Assessing and enhancing the capabilities of young novice drivers to anticipate latent hazards in road and traffic situations. University of Groningen. Doctoral dissertation.

Weevers, I., Kuipers, J., Brugman, A. O., Zwiers, J., Van Dijk, E. M., & Nijholt, A. (2003). The virtual driving instructor: Creating awareness in a multiagent system. In 16th Conference of the Canadian Society for Computational Studies of Intelligence, AI 2003, Halifax, Canada, June 11–13, 2003, Proceedings (pp. 556–567). Springer. https://doi.org/10.1007/3-540-44886-1_56.