Transportation Engineering 20 (2025) 100338

Contents lists available at ScienceDirect

Transportation Engineering

journal homepage: www.sciencedirect.com/journal/transportation-engineering

Real-time coaching programs for Manage-How-You-Drive insurance schemes: Analysis of retention after feedback removal

Federico Orsini [a,b,c,*], Giulia De Cet [a,b], Francesca Freuli [d], Leandro L. Di Stasi [e], Mariaelena Tagliabue [c,b,a], Massimiliano Gastaldi [a,b,c], Riccardo Rossi [a,b]

[a] Department of Civil, Environmental and Architectural Engineering, University of Padua, Padua, Italy
[b] Mobility and Behavior Research Center – MoBe, University of Padua, Padua, Italy
[c] Department of General Psychology, University of Padua, Padua, Italy
[d] Faculty of Humanities and Social Sciences, Helmut Schmidt University, Hamburg, Germany
[e] Mind, Brain, and Behavior Research Centre, University of Granada, Granada, Spain

ARTICLE INFO

Keywords:
Road safety
Real-time feedback
Long-term
In-vehicle system
Human-machine interface
ADAS
Smart mobility

ABSTRACT

Manage-How-You-Drive (MHYD) is an innovative usage-based insurance scheme where drivers are charged premiums based on their safety performance, incorporating real-time coaching programs to provide drivers with contingent feedback, nudging them to drive more safely. As limited research exists on these novel schemes, this study aims to confirm their effectiveness, by expanding the sample size and the scope of analysis from a previous study by the authors, and to specifically focus on the retention of improved behavior and the impact of driver characteristics and feedback types on retention.

A driving simulator experiment involving 100 drivers was used to test four feedback systems, with different modality (auditory vs. visual) and valence (i.e., pleasantness of the feedback: positive vs. negative), based on the occurrence of Elevated Gravitational-Force Events (EGFEs, i.e., harsh acceleration/deceleration events). Drivers completed three trials, spaced four weeks apart. The first trial served as a baseline without any feedback, in the second trial one of the feedback systems was presented, and the third trial had no feedback. Program effectiveness and retention were assessed based on EGFE occurrences and mean acceleration/deceleration. Its indirect influence on speeding, tailgating, and lateral control was investigated to assess potential additional enduring effects on safety performance.

Drivers, especially those identified as "aggressive" during the baseline trial, not only significantly benefited from using the coaching program, but were also able to at least partially retain such benefits in terms of acceleration/deceleration, speeding and tailgating, irrespective of feedback type. These findings highlight the potential practical advantages of MHYD real-time coaching systems for road safety.

1. Introduction

1.1. Usage-based insurance schemes

Usage-based insurance (UBI) schemes are innovative insurance programs that calculate premiums based on individual car usage or driving behavior rather than relying solely on traditional risk factors. There are two main types of UBI schemes: the Pay-As-You-Drive (PAYD), in which the premium depends on drivers' mileage, and the Pay-How-You-Drive (PHYD), which considers actual on-road behavior of the users [1,2]. Recently, a third type of scheme has been introduced, the Manage-How-You-Drive (MHYD), in which the premium is calculated similarly to the PHYD, but users receive real-time alerts and suggestions to improve their safety performance [3,4]. It should be noted that there is some ambiguity in published literature with some MHYD simply referred to as "PHYD with real-time feedback" (see Section 1.2 for more details).

To assess risk PHYD/MHYD schemes use real-time data collected from a vehicle's onboard telematics system, instead of general demographics (age, gender, location) and historical data. Drivers who exhibit safer driving habits, such as maintaining steady speeds, gentle braking, and obeying traffic laws, may be eligible for lower insurance

* Corresponding author at: Department of Civil, Environmental and Architectural Engineering, University of Padua, Padua, Italy.
E-mail address: federico.orsini@dicea.unipd.it (F. Orsini).

https://doi.org/10.1016/j.treng.2025.100338
Received 23 December 2024; Received in revised form 18 April 2025; Accepted 28 April 2025
Available online 29 April 2025

premiums [5–8]. In addition to driving pattern monitoring, MHYD can also include fatigue, drowsiness and driving distraction monitoring, with the use of cameras or wearable sensors [3].

Although these schemes are promoted by insurance companies mainly because they have a direct financial interest in reducing insurance claims caused by their clients, in general, PHYD/MHYD schemes have been found to offer several potential benefits to users and the general community [2,9]. First, they provide increased fairness, since premiums are more accurately tailored to individual driving behavior, rewarding safe drivers and encouraging improved habits. Secondly, they are an incentive for safer driving, as users have a financial motivation to drive more responsibly, leading to improved road safety and reduced accidents [10]. Lastly, they can provide environmental benefits, as some PHYD/MHYD programs encourage eco-friendly driving behavior (e.g., reducing harsh accelerations or limiting speed), contributing to reduced fuel consumption and lower emissions [11,12].

In terms of acceptance, these programs hold promise. Fafoutellis et al. [13] investigated the perception of a PHYD system among drivers in Athens, Greece. The proposed system monitored drivers' behavior and offered incentives to those who adopted eco-friendly driving habits. The study found that most drivers would be willing to change their driving behavior in exchange for discounts, despite the potential increase in travel time, with eco-active and eco-aware drivers, as well as regular commuters, being more likely to accept the system.

1.2. Effectiveness and retention of PHYD/MHYD schemes

Within PHYD schemes, users receive a feedback on their performance, which consists of an "after-drive" report [2,12,14]. These schemes have shown to be effective, as outlined in the review by Ziakopoulos et al. [15]; for example a smartphone application providing disaggregate post-trip feedback was able to produce a 14.5 % decrease in speeding among motorcyclist [16] and of 10~12 % in harsh braking/accelerations among car drivers [17]. Some studies have investigated the retention of such systems, revealing that after-drive feedback was ineffective in producing long-term effects on users [18,19].

On the other hand, the MHYD systems, which provide real-time feedback, have emerged and have been found to be potentially more effective than those relying on delayed feedback (i.e., PHYD). In the study by Dijksterhuis et al. [20], 60 participants were involved in a driving simulator study, in which they were either presented after-drive or real-time feedback, in addition to a monetary incentive, showing an overall better performance for the real-time feedback group. This feedback was visually provided through an in-car device, informing drivers about harsh acceleration, deceleration, and cornering events, as well as speed limit violations. Drivers underwent four trials: an initial baseline trial without feedback, two trials with feedback and incentives, and a final baseline trial without feedback nor incentives. However, participants did not retain their safety improvement in the final trial, reverting to their baseline performance. No factors regarding participants' characteristics were included in the analyses.

While not necessarily explicitly referring to MHYD schemes, other studies have investigated the effectiveness and retention of real-time feedback and monetary incentive systems primarily aimed at improving road safety.

Reagan et al. [21] investigated the effects of real-time feedback and monetary incentives on driving behavior with a naturalistic study involving 50 participants. The system considered only speeding violations. Participants were divided into three groups: a control group, and two experimental groups (with and without monetary incentives). The experiment lasted four weeks. The first was a baseline week without any feedback; successively, in either the second or the third week, participants received real-time feedback, which consisted in a combination of visual and auditory cues upon speeding violations. The final week was again a baseline without any incentive/feedback. As in the study by Dijksterhuis et al. [20], the impact of driver characteristics was not investigated. Their findings indicated that the incentive system resulted in significant reductions in driving faster than the posted limit, while the feedback system led to modest changes in speeding. However, in general, behavioral changes were not retained upon removing the incentives. Similar evidence of the effectiveness of monetary incentives was observed in the subsequent study by Mullen et al. [22], which however did not investigate retention.

A study by Merrikhpour et al. [23] presented findings from a field operational trial conducted to evaluate the effects of a feedback-reward system on speed limit and headway compliance. The study involved 37 participants, and consisted of a 2-week baseline, a 12-week intervention, and a 2-week post-intervention phase. The feedback was delivered visually and included both positive and negative valence. Results showed that the system was effective in improving both speed limit and headway compliance, especially for drivers who were less compliant during the baseline. These improvements decreased slightly in the post-intervention phase, while still remaining significant compared to the baseline.

A similar design has been recently applied in the naturalistic study by Chen and Donmez [24] which investigated long-term impacts of real-time feedback and financial incentives on driving behavior; their study involved 58 drivers, with a 4-week baseline period without feedback, a 10-week intervention period in which several combinations of feedback/incentives were tested, and a 2-week post-intervention period without any feedback/incentives. Their program provided negative visual feedback upon speeding violations. A short-term effectiveness, which did not sustain after the removal of the feedback, was observed. Investigation of the impact of driver characteristics was not included in their study.

The effectiveness on safety performance of a MHYD scheme available in the Italian insurance market has been previously investigated in our laboratory through a driving simulator study involving 43 participants [25]. Participants completed two trials: a baseline trial where they were asked to drive as they normally would, and a second trial featuring a real-time coaching program. Participants were divided into two groups, balanced by gender and driving style observed during the first trial (two driving style groups were identified through a cluster analysis on several kinematic driving variables). In the second trial, one group received contingent negative visual feedback based on the occurrence of harsh acceleration/deceleration events, while the other group received contingent positive visual feedback based on smooth acceleration/deceleration events. The findings demonstrated a significant reduction in harsh events (referred to here as elevated gravitational-force events, or EGFEs) among aggressive drivers, regardless of feedback type.

Subsequently, we explored the system's effectiveness on other behaviors, such as overtaking a cyclist [26] and exiting a highway [27], using an extended sample that included two additional groups to examine differences in feedback modality (visual vs. auditory). These studies revealed that, although the coaching program provided only basic feedback about the occurrence of harsh accelerations/decelerations, it effectively encouraged safer driving behavior, particularly among aggressive drivers.

1.3. Study motivation, contribution, and objectives

As evidenced by the studies reviewed in Section 1.2, despite the increasing popularity and diffusion of MHYD schemes, there is relatively sparse literature on investigating their effectiveness, with some contrasting results especially concerning their retention over time. In addition, the impact of the design of the system delivering real-time feedback has been largely overlooked, and the potential modulating effect of driving characteristics such as gender or driving style has only been included in the analyses of Merrikhpour et al. [23] and in our previous studies described in Section 1.2.

The present study aims to address these gaps in the literature, by extending our previous study [25]. A driving simulator experiment involving 100 participants was conducted to investigate the effectiveness and retention of an MHYD scheme. Participants underwent three sessions separated by 4 weeks: a baseline trial with no feedback, a second trial with the feedback system active, and a final trial again without feedback.

This study holds practical relevance, as the coaching program developed and tested is based on an existing real-time feedback system available in the Italian UBI market. Investigating retention after feedback removal is of particular interest because, in real-world scenarios, drivers may choose to deactivate such systems, a behavior observed with other driving assistance technologies (e.g., lane-keeping assist), which drivers sometimes disable due to perceived inconvenience [28]. Furthermore, the tested coaching program could serve as an offline training tool within PHYD schemes or for other driver education purposes.

Compared to the our preliminary reference study [25], the present work significantly expands both the sample size and the scope of analysis. In particular:

- This study expanded to a total of 100 drivers, compared to the previous 43.
- It included an additional third trial, conducted without any feedback, to assess the long-term retention of driving behaviors acquired during the second trial (i.e., the trial with active feedback).
- It featured four experimental groups, expanding from two, allowing for an investigation not only into feedback emotional valence (negative vs. positive) but also feedback type (visual vs. auditory).
- In evaluating the effectiveness of the coaching program, we assessed not only the reduction of EGFEs but also changes in mean acceleration and deceleration.
- Building on our previous findings regarding cyclist overtaking behavior and highway exiting [26,27], we examined not only the program's effectiveness in reducing harsh acceleration/deceleration but also its indirect impact on other related driving behaviors such as speeding, tailgating, and lateral control.

In contrast to some of the studies reviewed in Section 1.2, here we focus on the feedback system rather than the monetary incentive. As expressed above, various feedback systems were evaluated, differing in modalities and valence, since shedding light on the importance of feedback design is crucial for real-world implementation. This study did include a monetary incentive, as it was necessary to mimic a real-world MHYD scheme and motivate participants to try improving their safety performance, with the limitations discussed in Section 5.3.

Similarly to the approach followed in [25], the four experimental groups were balanced for gender and driving style, allowing exploration of whether these factors modulate program effectiveness and retention.

In summary, this study aims to:

1. Investigate in a detailed way the effectiveness of the coaching program when the real-time feedback is active and evaluate the retention after its removal. The research hypothesis was that the system, consistent with our preliminary study [25], would improve acceleration/deceleration behavior. Additionally, we expected this effectiveness to be at least partially retained, since the tested feedback aligned with precision teaching protocols [29], as further discussed in Section 4.1.
2. Assess the indirect impact of the coaching program on other driving behaviors, such as speeding, tailgating, and lateral control. We hypothesized that the system, when active, could provide a general improvement in safety behavior, consistent with our previous studies [26,27]. In addition, since driving at lower speed and maintaining a safe distance from the vehicle in front is associated with smoother longitudinal control, we expected that any retention in the improvement of acceleration/deceleration behavior would likely extend to speeding and tailgating.
3. Investigate the impact of some driver characteristics such as gender and driving style. A significant modulating effect of the driving style was expected, given previous literature findings [23,25], whereas no specific evidence supported the expectation of a gender-based difference.
4. Evaluate the impact of the modality and valence of the feedback delivery. Although existing literature supports positive valence over negative, our previous studies on the same system found no significant effects of either valence or modality [25–27]. Therefore, our research hypothesis was that no significant effects would be observed.

2. Methodology

2.1. Driving simulator experiment

2.1.1. Participants

One hundred active drivers, comprising 51 men and 49 women aged between 20 and 33, were recruited for this experiment, expanding the sample of our preliminary study [25]. To be eligible for participation, drivers needed to have normal or corrected-to-normal eyesight, no prior experience with a driving simulator, travel at least 1000 km annually, and possess a minimum of one year of driving experience. We specifically focused on young-adult drivers as they are more inclined to choose UBI schemes [30], and previous research indicates that they can significantly benefit from such systems [18].

Each participant was asked to complete three driving trials. Five participants experienced simulator sickness during the first trial and were unable to finish it; seventeen participants could not attend all trials, and four others failed to complete at least one of them. Consequently, the final sample comprised 74 drivers (37 males and 37 females, aged 20–33; mean age = 24, SD = 2.80).

Following the first trial, participants were assigned to four experimental groups, balanced for gender and observed driving style (further details in Section 2.1.3). Each group was treated with a specific feedback system (please refer to Section 2.1.4). The participants' characteristics for each group are outlined in Table 1. Note that any observed imbalance in the groups resulted from participants dropping out of the experiment.

All participants were unaware of the experiment's purpose and received compensation for their involvement. This study was conducted following the Code of Ethics of the World Medical Association (Declaration of Helsinki) [31]. The University of Padua's Ethical Committee for Psychological Research approved the study (IRB N 3024 06/06/2019), and all participants completed an informed consent form before participating.

2.1.2. Apparatus

This study was carried out with the dynamic driving simulator

Table 1
Participants' characteristics for each of the four experimental groups. Mean values for continuous variables with standard deviation in round brackets.

	Experimental group			
	Auditory-Negative	Auditory-Positive	Visual-Negative	Visual-Positive
N	14	19	21	20
Gender [F/M]	8/6	8/11	11/10	10/10
Driving style [Defensive/Aggressive]	8/6	9/10	9/12	10/10
Age [years]	24.9 (2.6)	24.5 (2.5)	23.7 (2.6)	24.4 (3.4)
Driving experience [years]	6.4 (2.5)	5.4 (2.5)	5.2 (2.6)	5.2 (2.2)
Annual mileage [km]	9214 (8250)	7692 (5999)	11,690 (10,409)	7400 (6010)

located at the University of Padua's Transportation Laboratory (Fig. 1), which has been previously validated [32].

The simulator was manufactured by STSoftware®; it featured a cockpit, a movable car seat, and an interactive force feedback steering wheel, along with clutch, brake, and gas pedals. To create an immersive experience, the simulator employed five 60-inch full high-definition displays, providing participants with a wide 330° horizontal by 45° vertical field of view. To further enhance realism, the simulator integrated a simulated surround sound system. Throughout the experiments, the simulator recorded 31 vehicle kinematic variables at a sampling frequency of 50 Hz.

2.1.3. Experimental design and procedure

The experiment consisted of three trials, carried out over three separate days (Fig. 2). The design is similar to that reported in our reference preliminary study [25], but with the significant addition of the third trial, which enabled the investigation of retention.

The simulated scenario encompassed a route that combined rural and urban roads, various types of intersections (priority, signalized, roundabout), and a highway section, spanning a total length of 11.5 km, without any elevation change. Throughout all three trials, the road's geometry remained constant, while the 3D objects surrounding the road varied for each trial. This approach ensured consistent analysis across trials, while participants were unaware that they were driving on the same route in all three trials, thus mitigating any potential learning effect.

On the first day, participants underwent a brief training session, after which they completed the initial trial. No feedback was provided during this trial.

Following the first set of trials, a cluster analysis was employed to categorize participants into two groups: defensive and aggressive drivers (the reader is referred to [25], for further details). These groups were then further divided into four equal subgroups based on driving style and gender.

Four weeks after the initial experiment, participants returned for a second trial. The four-week spacing was necessary to complete the first trial for all participants and to perform the cluster analysis on the trial 1 data, which was required before conducting trial 2. During trial 2 driving task, participants received one out of four distinct types of feedback (Section 2.1.4), depending on their group. Before the experiment, all participants received detailed information regarding the functioning of the real-time coaching program. They were made aware that the real-time feedback (either negative or positive, depending on the group) related to the occurrence (or not occurrence) of harsh accelerations/decelerations.

The third and final test was conducted four weeks after the second session, without any feedback being given to the participants.

To simulate the monetary incentive of a PHYD/MHYD insurance scheme, participants were informed, after the baseline trial, that their retribution for completing the experiments would depend on their driving behavior, as monitored by the simulator system, ranging between a minimum and maximum value. Regardless of their actual performance, all participants received the maximum possible reward at the end of the experiment.

2.1.4. Real-time coaching program

The real-time coaching program was designed to provide negative or positive feedback based, respectively, on the occurrence of severe or smooth driving events during the driving task. These events were defined as follows:

- A harsh event was a period longer than one second that exceeded a deceleration threshold of −0.4 g or an acceleration threshold of 0.3 g. These events were called EGFE [33].
- A smooth event was a period longer than one second that exceeded a minimal deceleration/acceleration threshold (± 0.075 g) without surpassing the −0.4 g deceleration threshold or the 0.3 g acceleration threshold.

These specific thresholds were chosen through sensitivity analysis in pilot tests, with support from the findings of previous studies [33,34].

Feedback was provided during the second trial based on the participant's subgroup, related to whether a severe or smooth driving event occurred. This resulted in the creation of four feedback systems, each differing in valence (positive or negative) and modality (auditory or visual). Whenever a harsh or smooth event was recorded, a 4-second visual or auditory signal was delivered.

For auditory cues, two sounds from the International Affective Digitized Sounds (IADS) library were chosen [35]. Sound #712 "Buzzer" was selected for negative feedback, as it was characterized by a low pleasure rating (2.35) on the normative 9-point rating scale for IADS sounds. Sound #717 "SlotMachine2" was chosen for positive feedback, having a high pleasure rating (7.32).

The visual signals consisted of a bright circle on a dark backdrop. Negative feedback was represented by a purple circle, while positive feedback was represented by a white circle (Fig. 2). This circle, simulating a device in the upper portion of the windscreen, had a diameter of 3 cm and was placed within 13 degrees of the participant's field of vision.

The negative visual feedback system was designed to closely resemble a real-time coaching tool already available in the Italian UBI market. The other systems were tested to assess potential improvements resulting from variations in feedback valence or modality.

2.2. Variables analyzed

2.2.1. Dependent variables

In order to evaluate the effectiveness of the real-time coaching programs, several variables were analyzed:

- *EGFE*, the total number of EGFEs observed during the driving task
- *Acc*, the mean acceleration during the driving task (in m/s^2)
- *Dec*, the mean deceleration during the driving task (in m/s^2)
- *Speed*, the mean speed during the driving task (in km/h)

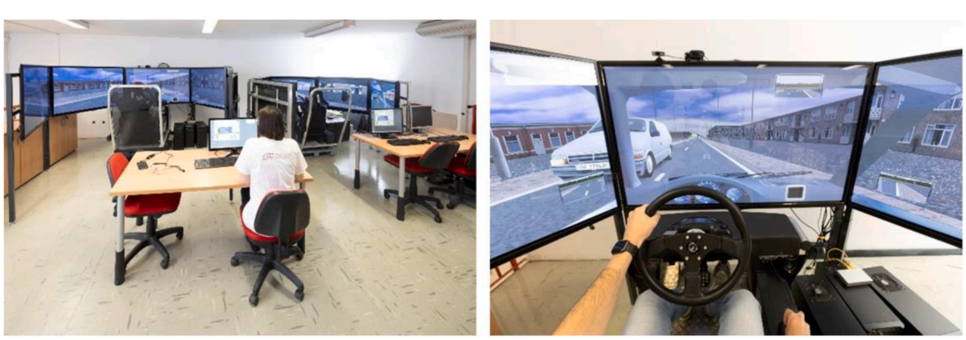

Fig. 1. Driving simulator at the Transportation Laboratory, University of Padua.

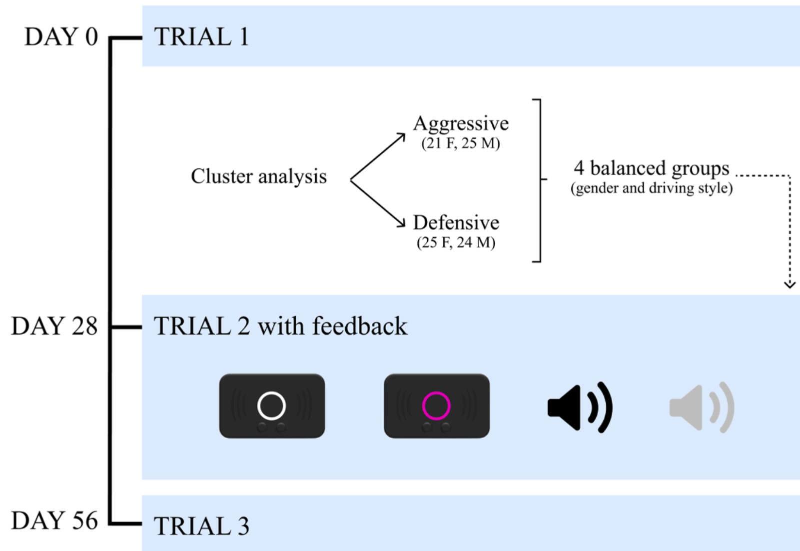

Fig. 2. Experiment design diagram.

- *Speed_Viol*, the ratio of time violating the speed limit, over the total time spent driving (in %)
- *THW*, the mean time headway to a leading vehicle (in s)
- *THW_Viol*, the ratio of time tailgating a leading vehicle, over the total time spent following a leading vehicle (in %)
- *SDLP*, the standard deviation of lateral position during the driving task (in m)
- *LatAcc*, the mean lateral acceleration during the driving task (in m/s^2)

EGFE, Acc, and *Dec* allowed for a direct evaluation of the system, as these variables are directly related to the occurrence of harsh braking/accelerations. The other variables were analyzed to investigate whether the coaching program had any other indirect effect on participants driving behavior, in terms of speeding, tailgating, and lateral control. *Speed* and *Speed_Viol* were computed considering only when the participant's vehicle had a speed higher than zero. *THW* was computed whenever the participant was at least 5 s behind another vehicle; this threshold was set to exclude all the situations in which the participants were reasonably too far away from a leading vehicle to be influenced by it [36,37]. For the *THW_Viol* variable, a participant was considered tailgating a leading vehicle, if the headway was less than 2 s, adhering to the "two seconds" rule-of-thumb that has been applied in several countries regulations [38]. As regards lateral control, we analyzed both the variability in lateral position (*SDLP*) and the mean lateral accelerations recorded during the driving task (*LatAcc*); both were computed only when the participant's vehicle had a speed higher than zero.

2.2.2. Independent variables

Based on the objectives of the study declared in Section 1.3, the following independent variables were considered in the analyses:

- *Trial*, a factor with three levels ("trial 1", "trial 2", "trial 3"). This factor allowed us to evaluate if the proposed real-time coaching program was effective when active (trial 2 vs. trial 1), and if its effectiveness was retained, at least to some extent, over time (trial 3 vs. trial 2; trial 3 vs. trial 1).
- *Cluster*, a factor with two levels, i.e. the two categories in which participants were divided based on the observed driving style in trial 1 ("defensive" vs. "aggressive")
- *Gender*, a factor with two levels ("female" vs. "male"; all participants identified themselves within these two groups, although other options were available)
- *Feedback Modality*, a factor with two levels, depending on the feedback system of trial 2 ("auditory" vs. "visual")
- *Feedback Valence*, a factor with two levels, depending on the feedback system of trial 2 ("positive" vs. "negative")

2.3. Statistical analyses

Several statistical analyses were used to evaluate the real-time coaching program effectiveness and retention.

The same cluster analysis adopted in Rossi et al. [25], was carried out to divide participants into two groups based on their driving style observed during trial 1. A total of 31 driving variables, recorded by the simulator in trial 1, were considered. The aim was to evenly distribute participants with various driving styles among the four feedback systems administered during trial 2. Ward's approach to hierarchical clustering with squared Euclidean distance measures was applied when determining the number of clusters [39,40]. Z-scores were created by standardizing the grouping factors [41–43]; in order to find the best cluster solution for the sample, a K-means cluster analysis was performed.

As regards the analysis on program effectiveness and retention, they were conducted separately for each dependent variable, using generalized linear mixed models (GLMMs), and assuming an alpha level of 0.05. These models were chosen to account for the panel structure of our data, which included repeated measurements on individual subjects, and the non-normal distributions of the dependent variables [44].

Since *EGFE* variable consisted of count data, a Poisson distribution family with logarithmic link function was used. For the *Acc, Dec, Speed, THW, SDLP*, and *LatAcc* variables, which were continuous non-negative, a

Table 2
Estimated marginal means and standard errors (in round brackets) of all dependent variables, at different levels of factors Trial and Cluster.

Variable	Cluster	Trial			
		1	2	3	Total
EGFE	Defensive	4.40 (0.389)	3.38 (0.333)	2.73 (0.294)	3.44 (0.227)
	Aggressive	8.38 (0.574)	4.21 (0.372)	4.03 (0.363)	5.22 (0.301)
	Total	6.07 (0.342)	3.77 (0.253)	3.32 (0.235)	
Acc [m/s^2]	Defensive	0.51 (0.017)	0.40 (0.014)	0.46 (0.015)	0.46 (0.010)
	Aggressive	0.71 (0.023)	0.48 (0.016)	0.52 (0.017)	0.56 (0.012)
	Total	0.60 (0.014)	0.44 (0.010)	0.49 (0.011)	
Dec [m/s^2]	Defensive	0.59 (0.018)	0.50 (0.015)	0.53 (0.016)	0.54 (0.012)
	Aggressive	0.79 (0.023)	0.61 (0.018)	0.62 (0.018)	0.67 (0.014)
	Total	0.68 (0.014)	0.55 (0.011)	0.57 (0.012)	
Speed [km/h]	Defensive	53.7 (0.566)	51.3 (0.540)	53.8 (0.566)	53.0 (0.406)
	Aggressive	63.8 (0.658)	57.9 (0.596)	58.5 (0.602)	60.0 (0.450)
	Total	58.6 (0.431)	54.5 (0.402)	56.1 (0.413)	
Speed_Viol [%]	Defensive	7.73 (1.069)	6.21 (0.892)	8.35 (1.102)	7.38 (0.748)
	Aggressive	30.44 (2.373)	16.14 (1.662)	16.45 (1.691)	20.32 (1.493)
	Total	16.1 (1.282)	10.1 (0.907)	11.8 (0.994)	
THW [s]	Defensive	2.98 (0.066)	2.89 (0.064)	2.92 (0.065)	2.93 (0.043)
	Aggressive	2.32 (0.050)	2.61 (0.057)	2.71 (0.059)	2.54 (0.037)
	Total	2.63 (0.041)	2.75 (0.043)	2.81 (0.044)	
THW_Viol [%]	Defensive	22.1 (1.72)	25.9 (1.87)	22.5 (1.74)	23.4 (1.19)
	Aggressive	43.1 (2.16)	32.4 (2.00)	27.9 (1.90)	34.2 (1.39)
	Total	31.6 (1.45)	29.0 (1.38)	25.1 (1.30)	
SDLP [m]	Defensive	0.28 (0.006)	0.28 (0.006)	0.29 (0.006)	0.28 (0.005)
	Aggressive	0.30 (0.006)	0.28 (0.005)	0.30 (0.006)	0.29 (0.005)
	Total	0.29 (0.004)	0.28 (0.004)	0.29 (0.004)	
LatAcc [m/s^2]	Defensive	0.28 (0.011)	0.24 (0.009)	0.32 (0.012)	0.28 (0.008)
	Aggressive	0.38 (0.014)	0.28 (0.010)	0.34 (0.012)	0.33 (0.009)
	Total	0.33 (0.009)	0.26 (0.007)	0.33 (0.009)	

Gamma distribution family with logarithmic link function was selected; for the *Speed_Viol* and *THW_Viol* variables, which are expressed in percentage, a Beta distribution family with logit link function was used. For each chosen distribution family and its corresponding link function in the GLMMs, focused diagnostics were conducted to assess their suitability for modeling the data. Diagnostics included plotting the residuals to investigate any abnormal pattern and creating QQplots to detect overall deviations from the expected distribution [45]. Full specification of the models and respective diagnostics are reported in Supplemental Materials.

Several fixed effect factors were included in the models: *Trial, Cluster, Feedback Modality, Feedback Valence, Gender*, as detailed in Section 2.2.2. The interaction between *Trial* and the other factors was also included, enabling the investigation of any relationships between program effectiveness/retention and driver/feedback characteristics. Participant ID was considered as a random grouping factor.

The appropriateness of the mixed-effects analyses was confirmed through a set of likelihood ratio tests comparing each GLMM with the corresponding GLM (i.e., without the random factor). In addition, to quantify how much additional variance is explained by the random factor, for each GLMM we reported marginal and conditional R-squared values. Marginal R-squared (R^2_m) quantifies the variance in the dependent variable explained solely by the fixed effects, while conditional R-squared (R^2_c) includes both fixed and random effects [46].

Given the relatively small number of factors and the need to investigate all of them to address the study's research questions, no variable selection procedure was applied. In other words, a hypothesis-driven (rather than data-driven) approach was adopted for model selection, avoiding potential biases introduced by automatic selection procedures [47].

Given the presence of factors with several levels and interactions between factors, interpreting the impact of the fixed effects parameter estimates can be challenging [48]. For this reason, in Section 3, the reported inference on the fixed factors was performed with Wald chi-squared tests [49], supported by several figures to facilitate visual interpretation of the results. Model parameter estimates, for brevity, are provided in the Supplemental Materials.

In the post-hoc analysis, pairwise comparisons were conducted using z statistics derived from Wald tests, to detect specific group differences following the significant omnibus Wald chi-squared test in the GLMM. Corrections were applied to control the familywise error rate, with p-values adjusted using the Tukey method.

For estimation and inference on GLMMs, R packages "glmmTMB" [50] and "car" [49] were used, model diagnostics was carried out with "DHARMa" [45] and "MuMIn" [51], and post-hoc analyses were performed using the "emmeans" package [52].

3. Results

This Section provides a concise report of the results. Detailed specifications of all GLMMs and respective goodness-of-fit diagnostics are reported in Supplemental Materials. An extensive discussion of the results is provided in Section 4.

Section 3.1 briefly presents the cluster analysis. Sections 3.2 – 3.5

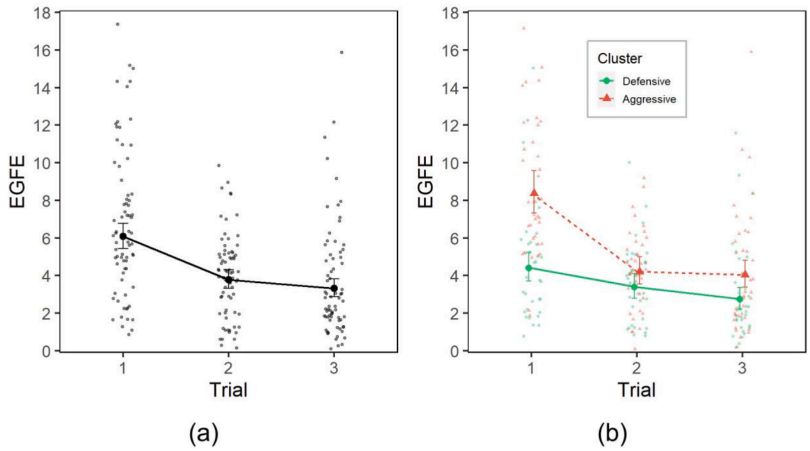

Fig. 3. Visualization of the effects of: (a) factor Trial on EGFE; (b) factors Trial and Cluster on EGFE. Solid dots represent marginal means, bars are 95 % confidence intervals of marginal means, background dots are individual observations.

presents all the GLMM results; estimated marginal means for all dependent variables are all reported in Table 2.

3.1. Cluster analysis

The cluster analysis was carried out considering the data of all 95 participants which completed the first trial. K-means clustering results are presented in detail in Rossi et al. [26].

A two-cluster optimal solution was obtained. Compared to Cluster 2, participants in Cluster 1 exhibited a smoother driving style, characterized by fewer and more consistent acceleration and speed peaks. Additionally, participants in Cluster 1 showed higher mean scores and lower SD scores for time to collision, as well as lower scores (both mean and SD) for engine RPM, accelerator pedal position, brake pedal position, and steering errors. Moreover, they had lower mean scores for deviation from lateral position and lateral speed.

Based on these findings, participants in Cluster 1 were identified as "defensive" drivers (49 participants, 25 females), while participants in Cluster 2 were classified as "aggressive" drivers (46 participants, 21 females).

3.2. Acceleration/deceleration behavior

3.2.1. EGFE

To evaluate the effectiveness of the real-time coaching program, which is directly linked to the occurrence or non-occurrence of EGFEs, we conducted an analysis with GLMMs. The objective was to examine how the number of EGFEs changed across *Trials* and whether it was influenced by participant characteristics (*Cluster, Gender*) or feedback factors (*Modality* and *Valence*).

Trial factor demonstrated a significant effect, $\chi^2{}_2 = 30.0$, $p < .001$. As reported in Table 2 and illustrated in Fig. 3, the number of EGFEs tended to decrease between trials 1 and 2, with a relatively stable level observed in the third trial. Post-hoc tests confirmed this trend, revealing statistically significant differences between trials 1 and 2, $z = 6.16$, $p = < .001$, as well as between trials 1 and 3, $z = 7.49$, $p = < .001$, yet not between trials 2 and 3. Overall, the coaching program was effective in reducing the number of EGFEs, and this effectiveness persisted even after the removal of the feedback.

Cluster also exhibited significance, $\chi^2{}_1 = 33.5$, $p < .001$, revealing that aggressive drivers were more prone to EGFEs (estimated marginal mean EMM = 5.22) compared to defensive drivers (EMM = 3.44), as anticipated.

The interaction between *Trial* and *Cluster* was also found to be significant, $\chi^2{}_2 = 8.1, p = 0.018$. Fig. 3b depicts an interesting interpretation of this interaction. For aggressive drivers, the coaching program proved highly effective in reducing EGFEs between trials 1 and 2, as the marginal means nearly halved, decreasing from 8.38 to 4.21, $z = 7.11$, $p = < .001$, and this effect was sustained in trial 3. In contrast, the program had a modest positive impact on the behavior of defensive drivers, with no significant difference observed between trials 1 and 2, or between trial 2 and 3. Nevertheless, an overall significant reduction was reported between trials 1 and 3, $z = 3.78$, $p = .002$ (4.40 vs. 2.73).

Neither *Gender* nor feedback *Modality/Valence* had any significant impact on *EGFE*.

3.2.2. Mean acceleration and deceleration

Further analyses with GLMMs investigated dependent variables *Acc* and *Dec*.

As can be observed in Fig. 4a and Fig. 4c, *Trial* had a significant effect on both *Acc*, $\chi^2{}_2 = 40.3, p < .001$, and *Dec*, $\chi^2{}_2 = 33.6, p < .001$. Mean acceleration significantly decreased in trial 2, $z = 10.15$, $p < .001$, and then increased back in trial 3, $z = -3.21$, $p = .004$, while remaining significantly lower than the baseline, $z = 6.94$, $p < .001$. Conversely, mean deceleration significantly decreased in trial 2, $z = 8.53$, $p < .001$, but then the improvement in deceleration behavior was maintained in trial 3.

Cluster was also significant on both variables ($\chi^2{}_1 = 50.1, p < .001$ on *Acc*, $\chi^2{}_1 = 50.8$, $p < .001$ on *Dec*), as well as the interaction between *Trial* and *Cluster* ($\chi^2{}_2 = 13.7$, $p < .001$ on *Acc*, $\chi^2{}_1 = 9.5$, $p = .009$ on *Dec*). Fig. 4b and Fig. 4d show that aggressive drivers tended to overall show harsher mean acceleration and deceleration. Interestingly, while both clusters significantly reduced *Acc* and *Dec* in trial 2, the effectiveness of the coaching program in trial 3 was different: aggressive drivers were able to maintain the same level of *Acc* and *Dec* also in trial 3, while defensive drivers worsened their performance in trial 3, returning to their baseline value. All pairwise comparisons are reported in Tables S8 and S12 in Supplemental Material.

Again, neither *Gender* nor feedback *Modality/Valence* had any significant impact on acceleration/deceleration behavior.

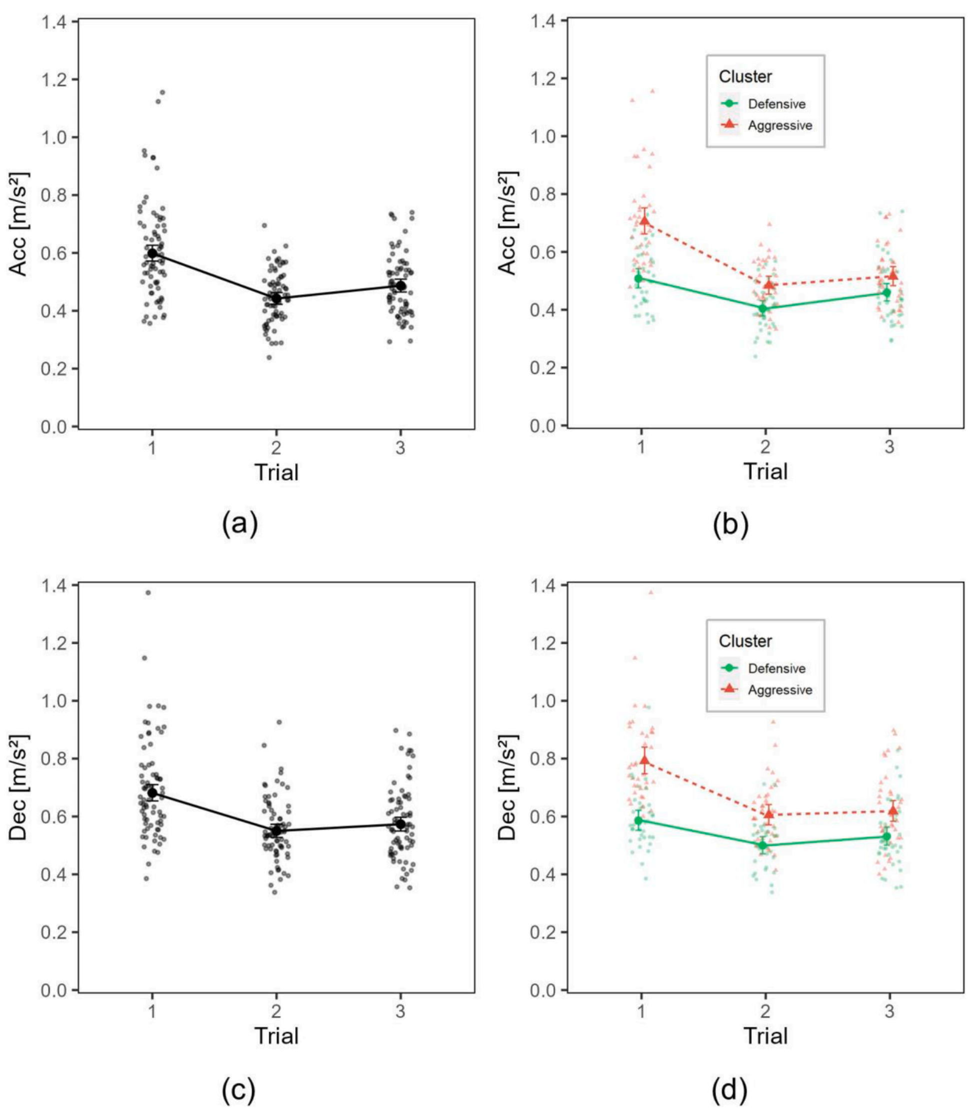

Fig. 4. Visualization of the effects of: (a) factor Trial on Acc; (b) factors Trial and Cluster on Acc; (c) factor Trial on Dec; (d) factors Trial and Cluster on Dec. Solid dots represent marginal means, bars are 95 % confidence intervals of marginal means, background dots are individual observations.

3.3. Speeding behavior

Speeding behavior was investigated considering *Speed*, i.e., the mean average speed recorded during the driving task,and *Speed_Viol*, i.e., the percentage of time spent above the speed limit.

Factor *Trial* had a significant effect on *Speed,* $\chi^2{}_2 = 39.6$, $p < .001$, with participants significantly, reducing their average speed in trial 2, $z = 8.18$, $p < .001$, and increasing it back in trial 3, $z = -3.26$, $p = .003$, albeit maintaining a significantly lower value compared to trial 1, $z = 4.93$, $p < .001$ (Fig. 5a). *Trial* has also a significant effect on *Speed_Viol,* $\chi^2{}_2 = 21.4$, $p < .001$; in this case however, participants, after improving their safety performance in trial 2, $z = 4.79$, $p < .001$, were able to maintain it in trial 3 (Fig. 5c).

Cluster had a significant effect on both *Speed,* $\chi^2{}_1 = 136.5$, $p < .001$, and *Speed_Viol,* $\chi^2{}_1 = 80.6$, $p < .001$, aggressive participants keeping on average higher speed (60.0 km/h vs. 53.0 km/h) and being more prone to speeding violations (20.3 % vs. 7.4 %).

The interaction between *Trial* and *Cluster* was again significant for both variables ($\chi^2{}_2 = 26.5$, $p < .001$ for *Speed,* $\chi^2{}_2 = 17.5$, $p < .001$ for *Speed_Viol*). As regards *Speed,* aggressive drivers reduced their average speed in trial 2 and maintained this reduction in trial 3, whereas defensive drivers were also able to reduce it in trial 2 but returned to their baseline level in trial 3 (Fig. 5b). As regards *Speed_Viol*, aggressive drivers again improved their behavior in trial 2 and maintained it in trial 3, while no significant change was observed for defensive drivers in the three trials (Fig. 5d). All pairwise comparisons are reported in Tables S16 and S20 in Supplemental Material.

Factor *Gender* had a significant on *Speed,* $\chi^2{}_1 = 6.5$, $p = .011$. Regardless of the trial, and therefore of the coaching system, male drivers (EMM = 57.8 km/h, SE = 0.439 km/h) tended to drive on average faster than females (EMM = 55.0 km/h, SE = 0.416 km/h). Feedback *Modality* and *Valence* did not have any significant impact on speeding behavior.

3.4. Tailgating behavior

As regards tailgating behavior, *THW*, i.e., the average time headway to any leading vehicle encountered during the driving task, and *THW_Viol*, i.e., the percentage of time spent violating the "two seconds rule" were investigated.

Factor *Trial* had a significant effect on *THW,* $\chi^2{}_2 = 13.1$, $p < .001$, and *THW_Viol,* $\chi^2{}_2 = 8.0$, $p = .018$. Participants had a slight tendency to maintain larger headways in trial 2, albeit the difference was not significant; in trial 3 this tendency continued, resulting in a significant improvement with respect to trial 1, $z = -3.27$, $p = .003$ (Fig. 6a). In terms of violations, no significant effect was observed in trial 2, while in

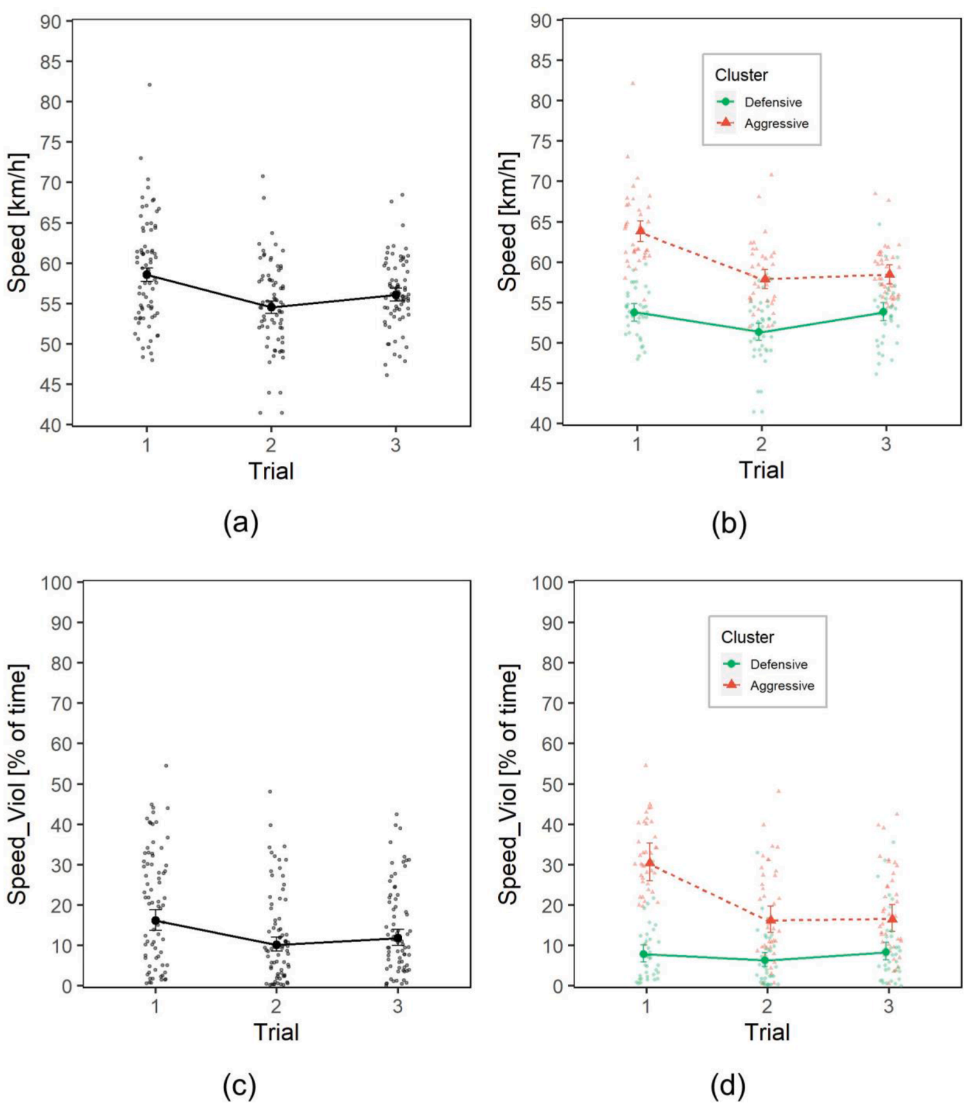

Fig. 5. Visualization of the effects of: (a) factor Trial on Speed; (b) factors Trial and Cluster on Speed; (c) factor Trial on Speed_Viol; (d) factors Trial and Cluster on Speed_Viol. Solid dots represent marginal means, bars are 95 % confidence intervals of marginal means, background dots are individual observations.

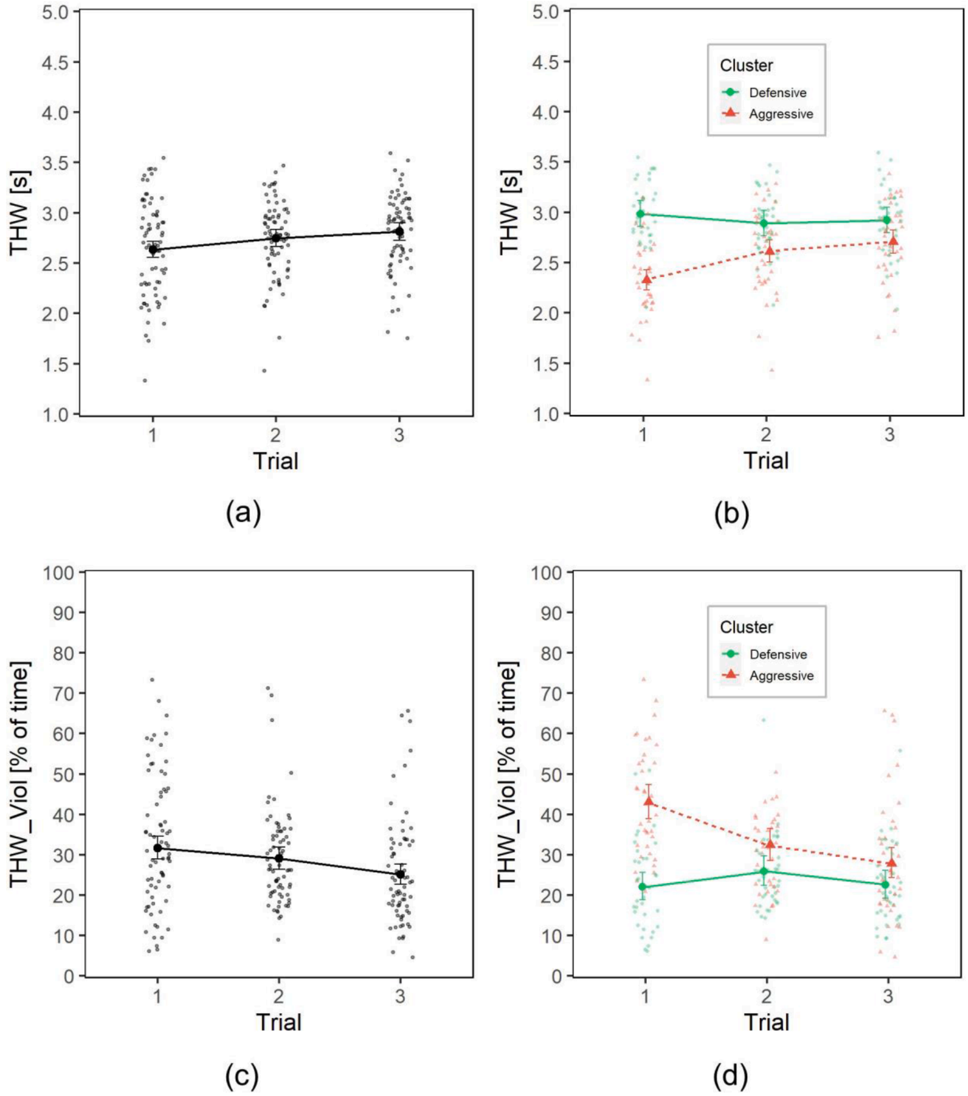

Fig. 6. Visualization of the effects of: (a) factor Trial on THW; (b) factors Trial and Cluster on THW; (c) factor Trial on THW_Viol; (d) factors Trial and Cluster on THW_Viol. Solid dots represent marginal means, bars are 95 % confidence intervals of marginal means, background dots are individual observations.

trial 3 participants tailgated for significantly less time when compared to trial 1, $z = 3.72$, $p < .001$ (Fig. 6c).

Aggressive drivers showed a riskier behavior than defensive drivers, as evidenced by factor *Cluster*, which had a significant effect on both *THW*, $\chi^2{}_1 = 65.4$, $p < .001$, and *THW_Viol*, $\chi^2{}_1 = 55.0$, $p < .001$.

The interaction between *Trial* and *Cluster* was also significant on both variables ($\chi^2{}_2 = 22.2$, $p < .001$ for *THW*, $\chi^2{}_2 = 21.1$, $p < .001$ for *THW_Viol*). Defensive drivers did not significantly modify their behavior across the three trials, whereas aggressive drivers adopted safer behavior, both in terms of *THW* and *THW_Viol*, in trial 2 and maintained it in trial 3. This is evidenced in both Fig. 6b and Fig. 6d Pairwise comparisons are reported in Tables S24 and S28 in Supplemental Material.

Factors *Gender*, feedback *Modality* and feedback *Valence* did not show any significant impact on tailgating behavior.

3.5. Lateral control

In terms of lateral control, factor *Trial* had a significant effect on *SDLP*, $\chi^2{}_2 = 6.8$, $p = .033$. Participants significantly improved their performance, by reducing *SDLP* in trial 2, $z = 2.42$, $p = .041$, but they increased it back in trial 3, $z = -3.70$, $p = .006$, reverting to their baseline level (difference between trial 1 and 3 was non-significant) (Fig. 7a).

Factor *Cluster* was also significant, $\chi^2{}_1 = 4.2$, $p = .042$, with defensive drivers generally demonstrating better lateral control. Interaction between *Trial* and *Cluster* was nonsignificant, as were all other investigated variables (Fig. 7b).

Both *Trial* and *Cluster* factors also significantly influenced another lateral control variable, *LatAcc*, $\chi^2{}_2 = 27.1$, $p < .001$ and $\chi^2{}_1 = 29.2$, $p < .001$, respectively. Participants again significantly improved lateral control in trial 2, $z = 7.78$, $p < .001$, followed by a reversal in trial 3, $z = -7.94$, $p < .001$ (Fig. 7c). Defensive drivers showed an overall lower *LatAcc* compared to aggressive drivers. The interaction term was also significant, $\chi^2{}_1 = 15.6$, $p < .001$, indicating a more pronounced improvement among aggressive drivers between trials 1 and 2. Both clusters significantly improved performance in trial 2, returning to baseline levels in trial 3 (Fig. 7d).

Factor *Gender* was significant, $\chi^2{}_1 = 18.7$, $p < .001$, with female drivers (EMM = 0.33 m/s^2, SE = 0.009 m/s^2) displaying lower lateral

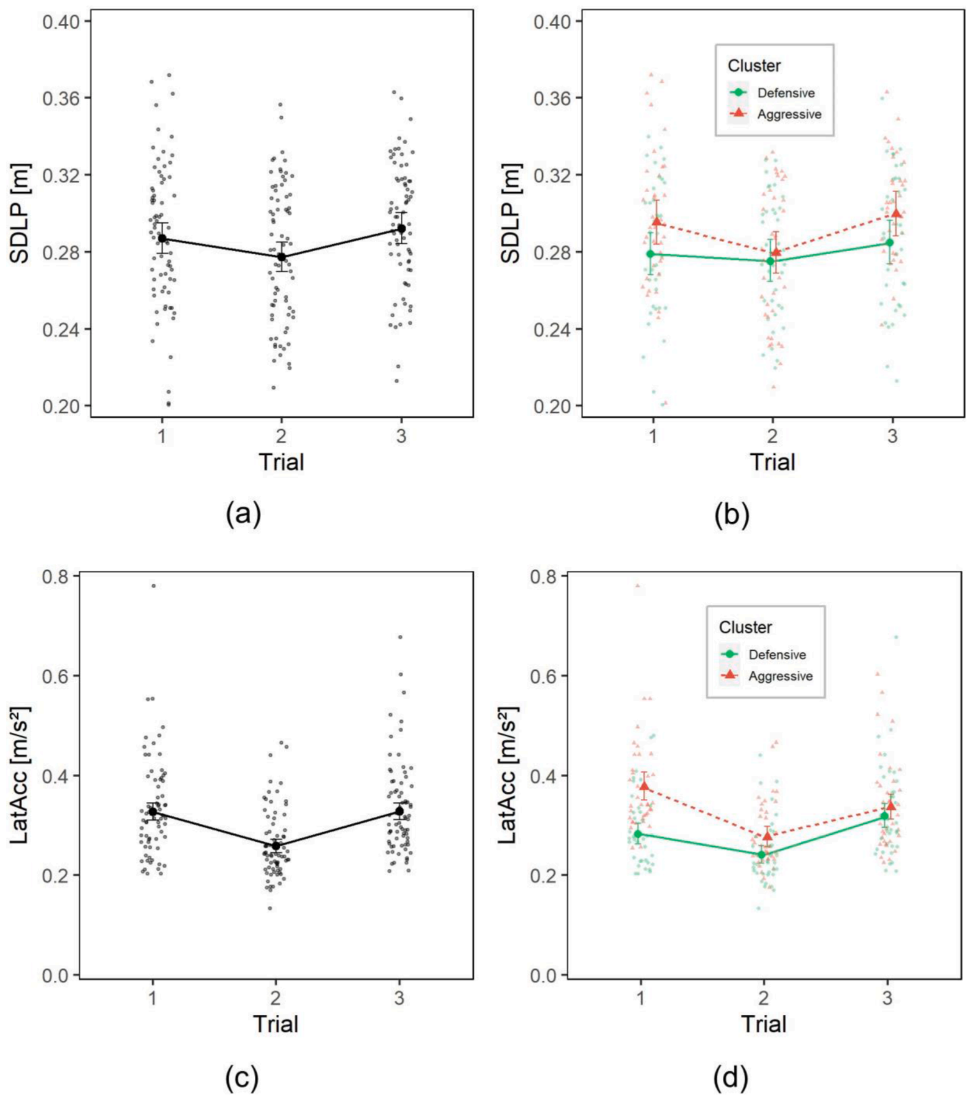

Fig. 7. Visualization of the effects of: (a) factor Trial on SDLP; (b) factors Trial and Cluster on SDLP; (c) factor Trial on LatAcc; (d) factors Trial and Cluster on LatAcc. Solid dots represent marginal means, bars are 95 % confidence intervals of marginal means, background dots are individual observations.

control than males (EMM = 0.28 m/s^2, SE = 0.008 m/s^2). However, no interaction with *Trial* was reported, meaning that the evolution of *LatAcc* was similar regardless of gender. Feedback *Modality* and feedback *Valence* did not show any significant impact on lateral control.

4. Discussion

In this section the findings from the analysis reported in Section 3 are discussed, with reference to the four objectives stated in the Introduction.

4.1. Effectiveness and retention of the program

The real-time coaching program evaluated in this study effectively reduced the number of EGFEs, with participants showing an overall 38 % reduction in EGFEs when driving with the coaching program active. These findings are consistent with previous research on this specific system [25], even after extending the sample size and testing an alternative feedback modality. Crucially, a novel finding of this study is that, in terms of program retention, the analyses demonstrated that participants maintained the improved performance they acquired even four weeks later while driving without any feedback.

The participants' ability to reduce the occurrence of EGFEs is confirmed by analyses on mean acceleration/deceleration. These analyses were introduced in this study to provide more robust evidence on the program effectiveness regarding acceleration and deceleration patterns, since they do not directly depend on the specific threshold chosen to define EGFEs. In trial 2, the mean acceleration and deceleration decreased by 27 % and 19 %, respectively. While the overall performance in mean acceleration slightly declined in trial 3 (though still maintaining an overall 18 % improvement compared to the baseline), the reduction in mean deceleration was fully retained.

These significant and enduring effects can be understood within the context of precision teaching techniques, which aims to enhance task fluency for learners [29,53–56]. This educational approach seeks to automatize learner responses, allowing them to generate appropriate patterns of responses to specific stimuli while minimizing the involvement of conscious control processes. It is particularly suited for structured tasks with clear stimulus-response connections. The coaching program in this study fits this definition, as it trained participants to avoid harsh accelerations and decelerations, providing contingent feedback on their performance via auditory or visual cues. Prior research on the application of precision teaching in the field of transportation has demonstrated the effectiveness of this technique and its ability to sustain long-lasting effects, even after the removal of the feedback system [57].

This study confirms the effectiveness and retention observed in the system tested by Merrikhpour et al. [23]; however, several other previous studies did not report significant retention, as discussed in Section 1.2.

In the case of Dijksterhuis et al. [20] it should be noted that their feedback system comprised a relatively complex visual interface. This interface included: three visual scales providing feedback on harsh cornering, braking, and acceleration; an indication of the current speed limit that turned red in case of violation; a real-time display of the monetary payout expressed in euros. Such complexity sets their system apart from ours, deviating from the “one stimulus – one response” paradigm of precision teaching. In the study by Dijksterhuis et al. [20], it is possible that participants may have been overwhelmed by information, hindering the assimilation of a long-lasting change in behavior.

The present findings contradicts also the results reported by Chen and Donmez [24] and Reagan et al. [21]. Although the feedback systems in those studies were generally similar to the one tested here, there was one major distinction: it provided feedback on speeding violations instead of harsh accelerations/decelerations. One possible explanation for the difference in outcomes is that speeding is a voluntary behavior, while performing harsh accelerations or decelerations involves a combination of voluntary and skill-based elements. Therefore, it could be hypothesized that the system tested in our study retained its effects because it not only signaled incorrect behavior but also implicitly taught drivers how to accelerate and decelerate smoothly.

4.2. Effects on speeding, tailgating behaviors, and lateral control

Despite being exclusively related to the occurrence of harsh acceleration/deceleration events, the system demonstrated significant and enduring indirect effects on speeding and tailgating behavior. When the feedback was active, participants reduced their average speed by 7 % and spent 37 % less time above the speed limit. Additionally, they maintained, on average, 5 % higher time headway from the leading vehicle and spent 8 % less time within two seconds of the vehicle in front, although these improvements reached statistical significance only for aggressive drivers (the modulating effect of driving style is further discussed in Section 4.3).

Concerning speeding behavior, the effectiveness was partially retained in trial 3, while improvements in tailgating behavior persisted and even increased, resulting in an overall 7 % higher headway and 21 % less time within two seconds of the vehicle in front compared to the baseline.

These findings highlight that the real-time coaching program tested here not only effectively reduced the occurrence of harsh acceleration/deceleration events but also prompted a more cautious driving behavior with respect to speeding and tailgating, which was retained even after the removal of the feedback. Participants seemed to understand that, to prevent EGFEs, they needed to drive at lower speeds and maintain larger safety distances from the vehicle in front. The delayed improvement in tailgating behavior might be attributed to its less immediate connection to EGFEs: participants likely became aware throughout trial 2 that maintaining larger headways could help avoid harsh acceleration/deceleration events and consolidated this awareness in the final trial.

The system also showed an indirect positive effect on lateral control in trial 2, with participants reducing both *SDLP* and *LatAcc*. It should be noted that, contrary to speeding and tailgating, these variables were substantially unrelated with the feedback provided. These indirect effects of the system align with previous observations regarding safety implications during cyclist overtaking [26] and highway exits [27], highlighting additional promising safety performance improvements when the system is active.

However, it should also be underlined that participants were not able to retain these lateral control improvements in trial 3. This further stress the role of the coaching program: since lateral control was largely unrelated to the real-time feedback, participants did not acquire any new knowledge on how to correctly maintain lateral control of the vehicle and, therefore, upon the withdrawal of the feedback, they reverted to their baseline level.

4.3. The modulating effect of driving style on program effectiveness and retention

A significant result emerging from the present study, confirming what observed in Rossi et al. [25] and aligning with the findings of Merrikhpour et al. [23], is that the program had varying impacts on drivers based on their driving styles. Aggressive drivers showed a significant reduction in EGFEs (−50 %) between the first two trials, whereas defensive drivers displayed a more modest improvement (−23 %). Similar trends were observed in the analysis of impacts on mean acceleration and deceleration, revealing lower program effectiveness in trial 2 for defensive drivers.

Interestingly the present study provides novel evidence on how driving style modulates retention. Unlike aggressive drivers, defensive drivers were unable to sustain their improvements in trial 3. Additionally, while aggressive drivers improved and retained positive changes in both speeding and tailgating behavior, defensive drivers did not exhibit any significant behavioral changes across the three trials.

The explanation for this modulating effect could be twofold. Firstly, there was likely some kind of ceiling effect, since aggressive drivers had more room for improvement than defensive drivers. Secondly, aggressive drivers might have been initially less aware of their incorrect behaviors, making it easier for the system to guide them towards improvement. In contrast, defensive drivers were already conscious of how to avoid harsh acceleration/decelerations, leaving them with less to learn; this is even more confirmed by the fact that defensive drivers did not modify at all their speeding/tailgating behavior. This observation aligns with the earlier comments in Section 4.1 on the retention of our system compared to that of Chen and Donmez [24] or Reagan et al. [21]: it suggests that the effect was sustained primarily by those participants who learned valuable skills from the coaching program. In other words, the retention of positive outcomes was linked to the acquisition of new knowledge and behaviors.

The other driver characteristics investigated, *Gender*, did not show any significant modulating effect, suggesting that the tested system is equally effective on the population, regardless of their gender.

4.4. Feedback valence and modality

The presentation of feedback itself did not significantly affect the program's effectiveness and retention, as both factors *Valence* and *Modality* turned out to be nonsignificant, in line with our research hypothesis.

The fact that feedback valence had no effect was contrary to some literature evidence suggesting that positive feedback might have a more pronounced impact, particularly in young drivers [58]. The sounds utilized as feedback in this study—a slot machine sound indicating 'a gain' and a buzzing sound indicating that 'something wrong' is happening—were chosen for their symbolic meanings. It is possible that these symbolic associations may have induced a ceiling effect, wherein the meaning of the sounds could have magnified their impact, potentially masking the differences in their outcomes.

Conversely, regarding feedback modality, this study aligns with others providing overall inconclusive findings on whether auditory-only or visual-only is more effective [59–61], which partly originally motivated testing both modalities in our system.

5. Conclusion

5.1. Main findings

The aim of this study was to confirm the effectiveness and investigate the retention of a real-time coaching program developed within the context of MHYD insurance schemes. This was accomplished through a driving simulator experiment, involving three trials spaced four weeks apart: a baseline trial without feedback, a trial with the coaching program, and a third trial without the program. The program's effectiveness was evaluated based on the number of harsh acceleration/deceleration events (EGFEs) recorded in the trials, and on the mean acceleration/deceleration observed during the driving task. In addition, indirect effects on speeding, tailgating, and lateral control were also investigated. The main findings, which answer to the research questions declared in the Introduction, can be summarized as follows:

1. The presence of the coaching program led to a significant reduction in EGFEs (−38 %), mean acceleration and mean deceleration compared to the baseline trial. The effectiveness in reducing EGFEs and mean deceleration was maintained four weeks after using the coaching program and, while the mean acceleration slightly increased, it remained significantly lower than in the baseline trial.
2. Positive safety impacts were also observed in terms of speeding and tailgating behavior, which were at least partially retained even four weeks after using the program. Lateral control also improved in trial 2, but reverted to baseline levels in trial 3.
3. Aggressive drivers benefited and retained more from the program than defensive drivers, not only in terms of harsh braking/acceleration events, but also of speeding and tailgating behavior.
4. The modality and valence of the feedback did not impact the program's effectiveness.

The application of this coaching program shows promise, as it effectively improves road safety, even in the long-term, through a simple and non-invasive feedback system, with flexibility regarding modality and valence.

5.2. Practical implications

Several practical implications can be derived from the present study.

Firstly, this study showcased the effectiveness of a very simple real-time coaching program, which is already available in the PHYD/MHYD insurance market, confirming results from our preliminary study. The reduction of EGFEs was quite drastic, producing an overall significant improvement on road safety (previously demonstrated in [26,27]). This tool can not only motivate drivers to adopt smoother acceleration/deceleration behavior, but also teach them how to effectively implement it. The improvement in safety provides benefits to insurance companies, ensuring the financial sustainability of such schemes.

Secondly, the study contributes evidence to the retention and long-term effectiveness of precision teaching techniques. In practical terms, this provides very interesting consequences, because it demonstrates that drivers acquire new skills from the system, enabling them to implement such skills without the need for continuous training. This aspect gains relevance when considering that some real-world drivers might decide to turn off the feedback system for various reasons, yet still retain its positive impacts. In addition, while not necessarily directly related to MHYD insurance schemes, it highlights the potential use of these techniques for offline education purposes. For instance, they could be applied as occasional training sessions within PHYD schemes, for the re-education of drivers who had their license revoked or for eco-driving teaching campaigns. By employing an appropriate feedback system during a short simulated or naturalistic session, drivers could gain long-term real-world benefits.

Thirdly, the tested system demonstrated greater effectiveness with aggressive drivers, holding practical relevance for safety. Although it has been reported that high-risk drivers are less inclined to participate in PHYD schemes [62], it is worth noting that reckless drivers may not be fully aware of their behavior, often overestimating their skills [63]. Regardless, it is advisable to allocate particular attention and implement specific campaigns targeting aggressive drivers, in order to attract them toward these schemes.

Lastly, the study revealed that different feedback systems (positive vs. negative, auditory vs. visual) exhibited similar effectiveness. This finding provides enhanced flexibility in choosing the appropriate system, optimizing other aspects such as avoiding interference with other in-vehicle warning signals.

5.3. Limitations and future research

While the present study offers valuable insights into the effectiveness and retention of real-time coaching programs, for a comprehensive generalization of these results, replication on different scenarios, possibly with naturalistic experiments, is recommended. This should involve testing alternative feedback systems (e.g., multi-modal) on cohorts of drivers with diverse sociodemographic characteristics.

The study involved young adult participants, who might be particularly receptive to acquiring new driving skills as a result of the real-time coaching program. Previous literature shows that even older drivers can significantly benefit from receiving feedback on their driving [64], so it is possible that the findings of the present study may be generalizable to the wider population, although this should be tested by replicating the study with an older cohort of participants.

Regarding the chosen experimental design, the potential for an order effect on the results cannot be entirely ruled out, although previous works on other real-time coaching programs in comparable scenarios did not exhibit any order effect [29]. It is important to highlight that the scenario was purposedly designed to mitigate any such effect, positioning 3D objects in different positions in each trial, preventing participants from recognizing the repetition of the planimetric route in all three trials.

It should be noted that the monetary incentive in this study may not fully represent that of MHYD schemes, as it was a short-term reward with a higher value per unit of time. While this might have potentially influenced the results, existing evidence supports our choice. Dijksterhuis et al. [65] argued that even with a high value per unit of time, if the total absolute value is small (as in this case), it should not significantly impact the outcome. Additionally, Mortimer et al. [66] demonstrated that increases in monetary incentives do not improve their overall effectiveness, and Vaezipour et al. [67] observed minimal effects of the addition of monetary incentives to an in-vehicle interface.

Another potential limitation of the experimental design employed is the inability to fully disentangle the effects of the coaching system and the monetary incentives on retention. Nonetheless, our findings indicate that the coaching program had a central role. This is partly evidenced by the absence of indirect effects on lateral control in trial 3: given that lateral control was independent of the feedback systems, participants did not gain new knowledge, thus failing to sustain the beneficial effects, with the monetary incentive alone unable to provide any effect in this regard.

In addition to addressing these limitations through new experiments, future research will also focus on evaluating additional indirect effects of the current real-time coaching program—specifically, how pollutant emissions are influenced with and without the feedback system. The reductions in mean acceleration and speed reported in this paper are likely indicators of reduced emissions, which need to be quantified to comprehensively assess and discuss implications in terms of eco-driving.

Further research will also examine how long the acquired skills are retained and whether it would be beneficial to incorporate regular driving sessions with the real-time feedback system active, as well as the optimal frequency of such sessions. Lastly, potential negative effects related to distraction during real-time system usage were not investigated in this study and cannot be ruled out, warranting further investigation.

Funding

The research was conducted as part of the "use-inspired basic research" initiative, for which the Department of General Psychology at the University of Padua was awarded the "Dipartimento di Eccellenza" designation by the Ministry of University and Research (MIUR). In addition, POR FSE 2014 - 2020 (Project ID: 2105–56–11–2018) and Generali Italia S.p.A. supported the project.

CRediT authorship contribution statement

Federico Orsini: Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Giulia De Cet:** Writing – review & editing, Visualization, Methodology, Investigation, Data curation, Conceptualization. **Francesca Freuli:** Writing – review & editing, Methodology, Investigation, Data curation, Conceptualization. **Leandro L. Di Stasi:** Writing – review & editing, Methodology, Conceptualization. **Mariaelena Tagliabue:** Writing – review & editing, Supervision, Resources, Methodology, Funding acquisition, Conceptualization. **Massimiliano Gastaldi:** Writing – review & editing, Supervision, Resources, Methodology, Funding acquisition, Conceptualization. **Riccardo Rossi:** Writing – review & editing, Supervision, Resources, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors 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 authors would like to thank Giulio Vidotto, Alberto Sarto and Giulia Gaita for their assistance with the experiment's design, as well as Rosa Rita Parisi, Elisabetta Dalrì, and Giovanni Nascimben for their assistance with data collection.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.treng.2025.100338.

Data availability

Data will be made available on request.

References

[1] M.F. Carfora, F. Martinelli, F. Mercaldo, V. Nardone, A. Orlando, A. Santone, et al., A "pay-how-you-drive" car insurance approach through cluster analysis, Soft Comput. 23 (2019) 2863–2875, https://doi.org/10.1007/s00500-018-3274-y.

[2] D.I. Tselentis, G. Yannis, E.I. Vlahogianni, Innovative motor insurance schemes: a review of current practices and emerging challenges, Accid. Anal. Prev. 98 (2017) 139–148, https://doi.org/10.1016/j.aap.2016.10.006.

[3] S. Arumugam, R. Bhargavi, A survey on driving behavior analysis in usage based insurance using big data, J. Big Data 6 (2019) 1–21, https://doi.org/10.1186/s40537-019-0249-5.

[4] H. Singh, A. Kathuria, Analyzing driver behavior under naturalistic driving conditions: a review, Accid. Anal. Prev. 150 (2021) 105908, https://doi.org/10.1016/j.aap.2020.105908.

[5] L. Boquete, J.M. Rodríguez-Ascariz, R. Barea, J. Cantos, J.M. Miguel-Jiménez, S. Ortega, Data acquisition, analysis and transmission platform for a Pay-As-You-Drive system, Sensors 10 (2010) 5395–5408, https://doi.org/10.3390/s100605395.

[6] P. Handel, I. Skog, J. Wahlstrom, F. Bonawiede, R. Welch, J. Ohlsson, et al., Insurance telematics: opportunities and challenges with the smartphone solution, IEEE Intell. Transp. Syst. Mag. 6 (2014) 57–70, https://doi.org/10.1109/MITS.2014.2343262.

[7] X. Hu, X. Zhu, Y.L. Ma, Y.C. Chiu, Q. Tang, Advancing usage-based insurance - A contextual driving risk modelling and analysis approach, IET Intell. Transp. Syst. 13 (2019) 453–460, https://doi.org/10.1049/iet-its.2018.5194.

[8] O. Mendels, G. Bertental, T. Kamara, Unsupervised driver safety estimation at scale, a collaboration with Pointer Telocation. https://devblogs.microsoft.com/cse/2018/07/30/unsupervised-driver-safety-estimation-at-scale/, 2018.

[9] T. Litman, Pay-As-You-Drive vehicle insurance: converting vehicle insurance premiums into use-based charges, Victoria Victoria Transp. Policy Inst. (2010).

[10] I. Reimers, B.R. Shiller, The impacts of telematics on competition and consumer behavior in insurance, J. Law Econ. 62 (2019) 613–632, https://doi.org/10.1086/705119.

[11] F. Cavallaro, F. Giaretta, S. Nocera, The potential of road pricing schemes to reduce carbon emissions, Transp. Policy 67 (2018) 85–92, https://doi.org/10.1016/j.tranpol.2017.03.006.

[12] G. Toledo, Y. Shiftan, Can feedback from in-vehicle data recorders improve driver behavior and reduce fuel consumption? Transp. Res. Part A Policy Pract. 94 (2016) 194–204, https://doi.org/10.1016/j.tra.2016.09.001.

[13] P. Fafoutellis, E.G. Mantouka, E.I. Vlahogianni, Acceptance of a Pay-How-You-Drive pricing scheme for city traffic: the case of Athens, Transp. Res. Part A Policy Pract. 156 (2022) 270–284, https://doi.org/10.1016/j.tra.2022.01.009.

[14] F. Tarlochan, S. Mohammed, Intelligent Transportation System: application of Telematics Data for Road Safety, in: Proc - 2023 9th Int Conf Inf Manag ICIM 2023, 2023, pp. 66–71, https://doi.org/10.1109/ICIM58774.2023.00018.

[15] A. Ziakopoulos, V. Petraki, A. Kontaxi, G. Yannis, The transformation of the insurance industry and road safety by driver safety behaviour telematics, Case Stud. Transp. Policy 10 (2022) 2271–2279, https://doi.org/10.1016/j.cstp.2022.10.011.

[16] A. Kontaxi, A. Ziakopoulos, G. Yannis, Investigation of the speeding behavior of motorcyclists through an innovative smartphone application, Traffic Inj. Prev. 22 (2021) 460–466, https://doi.org/10.1080/15389588.2021.1927002.

[17] A. Kontaxi, A. Ziakopoulos, G. Yannis, Trip characteristics impact on the frequency of harsh events recorded via smartphone sensors, IATSS Res. 45 (2021) 574–583, https://doi.org/10.1016/j.iatssr.2021.07.004.

[18] J.W. Bolderdijk, J. Knockaert, E.M. Steg, E.T. Verhoef, Effects of Pay-As-You-Drive vehicle insurance on young drivers' speed choice: results of a Dutch field experiment, Accid. Anal. Prev. 43 (2011) 1181–1186, https://doi.org/10.1016/j.aap.2010.12.032.

[19] T. Toledo, T. Lotan, In-Vehicle Data Recorder for Evaluation of Driving Behavior and Safety, Transp. Res. Rec. J. Transp. Res. Board 1953 (2006) 112–119, https://doi.org/10.1177/0361198106195300113.

[20] C. Dijksterhuis, B. Lewis-Evans, B. Jelijs, D. De Waard, K. Brookhuis, O. Tucha, The impact of immediate or delayed feedback on driving behaviour in a simulated Pay-As-You-Drive system, Accid. Anal. Prev. 75 (2015) 93–104, https://doi.org/10.1016/j.aap.2014.11.017.

[21] I.J. Reagan, J.P. Bliss, R. Van Houten, B.W. Hilton, The effects of external motivation and real-time automated feedback on speeding behavior in a naturalistic setting, Hum. Factors 55 (2013) 218–230, https://doi.org/10.1177/0018720812447812.

[22] N.W. Mullen, H. Maxwell, M. Bédard, Decreasing driver speeding with feedback and a token economy, Transp. Res. Part F Traffic. Psychol. Behav. 28 (2015) 77–85, https://doi.org/10.1016/j.trf.2014.11.008.

[23] M. Merrikhpour, B. Donmez, V. Battista, A field operational trial evaluating a feedback-reward system on speeding and tailgating behaviors, Transp. Res. Part F Traffic Psychol. Behav. 27 (2014) 56–68, https://doi.org/10.1016/j.trf.2014.09.002.

[24] W. Chen, B. Donmez, A naturalistic driving study of feedback timing and financial incentives in promoting speed limit compliance, IEEE Trans. Human-Mach. Syst. 52 (2022) 64–73, https://doi.org/10.1109/THMS.2021.3117234.

[25] R. Rossi, M. Tagliabue, M. Gastaldi, G. De Cet, F. Freuli, F. Orsini, et al., Reducing elevated gravitational-force events through visual feedback: a simulator study, Transp. Res. Procedia 52 (2021) 115–122, https://doi.org/10.1016/j.trpro.2021.01.013.

[26] R. Rossi, F. Orsini, M. Tagliabue, L.L. Di Stasi, G. De Cet, M. Gastaldi, Evaluating the impact of real-time coaching programs on drivers overtaking cyclists, Transp. Res. Part F Traffic Psychol. Behav. 78 (2021) 74–90, https://doi.org/10.1016/j.trf.2021.01.014.

[27] F. Orsini, M. Tagliabue, G. De Cet, M. Gastaldi, R. Rossi, Highway deceleration lane safety: effects of real-time coaching programs on driving behavior, Sustain 13 (2021) 9089, https://doi.org/10.3390/su13169089.

[28] I.J. Reagan, J.B. Cicchino, L.B. Kerfoot, R.A. Weast, Crash avoidance and driver assistance technologies – are they used? Transp. Res. Part F Traffic Psychol. Behav. 52 (2018) 176–190, https://doi.org/10.1016/j.trf.2017.11.015.

[29] F.N. Biondi, R. Rossi, M. Gastaldi, F. Orsini, C. Mulatti, Precision teaching to improve drivers' lane maintenance, J. Safety Res. 72 (2020) 225–229, https://doi.org/10.1016/j.jsr.2019.12.020.

[30] D.I. Tselentis, A. Theofilatos, G. Yannis, M. Konstantinopoulos, Public opinion on usage-based motor insurance schemes: a stated preference approach, Travel. Behav. Soc. 11 (2018) 111–118, https://doi.org/10.1016/j.tbs.2018.02.003.

[31] World Medical Association, World medical association declaration of Helsinki: ethical principles for medical research involving human subjects, JAMA - J. Am. Med. Assoc. 310 (2013) 2191–2194, https://doi.org/10.1001/jama.2013.281053.

[32] R. Rossi, M. Gastaldi, G. Gecchele, F. Biondi, C. Mulatti, Traffic-calming measures affecting perceived speed in approaching bends: on-field validated virtual environment, Transp. Res. Rec. 2434 (2014) 35–43, https://doi.org/10.3141/2434-05.

[33] B.G. Simons-Morton, Z. Zhang, J.C. Jackson, P.S. Albert, Do elevated gravitational-force events while driving predict crashes and near crashes? Am. J. Epidemiol. 175 (2012) 1075–1079, https://doi.org/10.1093/aje/kwr440.
[34] A.E. af Wåhlberg, Changes in driver celeration behaviour over time: do drivers learn from collisions? Transp. Res. Part F Traffic Psychol. Behav. 15 (2012) 471–479, https://doi.org/10.1016/j.trf.2012.04.002.
[35] M.M. Bradley, P. Lang, Affective norms for English words (ANEW): stimuli, instruction manual and affective ratings (Technical Report No. C-1). NIMH Cent Res Psychophysiology, Univ Florida, Gainesville, FL, 1999.
[36] K. Vogel, A comparison of headway and time to collision as safety indicators, Accid. Anal. Prev. 35 (2003) 427–433, https://doi.org/10.1016/S0001-4575(02)00022-2.
[37] X. Zhang, G.H. Bham, Estimation of driver reaction time from detailed vehicle trajectory data, in: Proc. IASTED Int. Conf. Model. Simul, ACTA Press, USA, 2007, pp. 574–579.
[38] TG road safety, Safe Distance between Vehicles (CEDR Report 2009/10.1), 2010.
[39] F. Lucidi, A.M. Giannini, R. Sgalla, L. Mallia, A. Devoto, S. Reichmann, Young novice driver subtypes: relationship to driving violations, errors and lapses, Accid. Anal. Prev. 42 (2010) 1689–1696, https://doi.org/10.1016/j.aap.2010.04.008.
[40] D. Marengo, M. Settanni, G. Vidotto, Drivers' subtypes in a sample of Italian adolescents: relationship between personality measures and driving behaviors, Transp. Res. Part F Traffic Psychol. Behav. 15 (2012) 480–490, https://doi.org/10.1016/j.trf.2012.04.001.
[41] E. Gianfranchi, A. Spoto, M. Tagliabue, Risk profiles in novice road users: relation between moped riding simulator performance, on-road aberrant behaviors and dangerous driving, Transp. Res. Part F Traffic Psychol. Behav. 49 (2017) 132–144, https://doi.org/10.1016/j.trf.2017.06.016.
[42] E. Gianfranchi, M. Tagliabue, A. Spoto, G. Vidotto, Sensation seeking, non-contextual decision making, and driving abilities as measured through a moped simulator, Front. Psychol. 8 (2017) 2126, https://doi.org/10.3389/fpsyg.2017.02126.
[43] E. Gianfranchi, M. Tagliabue, G. Vidotto, Personality traits and beliefs about peers' on-road behaviors as predictors of adolescents' moped-riding profiles, Front. Psychol. 9 (2018) 2483, https://doi.org/10.3389/fpsyg.2018.02483.
[44] C.E. McCulloch, S.R. Searle, J.M. Neuhaus, Generalized Linear Mixed Models (2008) 424.
[45] F. Hartig, DHARMa: residual diagnostics for hierarchical (Multi-Level /Mixed) regression models. https://doi.org/10.32614/CRAN.package.DHARMa.
[46] S. Nakagawa, P.C.D. Johnson, H. Schielzeth, The coefficient of determination R2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded, J. R. Soc. Interface 14 (2017) 20170213, https://doi.org/10.1098/rsif.2017.0213.
[47] B. Donmez, A.D. McDonald, J.D. Lee, L.N. Boyle, Road user behavior: describing, inferring, predicting and beyond, Transp. Res. Interdiscip. Perspect. 22 (2023) 100932, https://doi.org/10.1016/j.trip.2023.100932.
[48] H. Singmann, D. Kellen, An introduction to mixed models for experimental psychology, Methods Cogn. Psychol. (2019) 4–31, https://doi.org/10.4324/9780429318405-2.
[49] J. Fox, S. Weisberg, An R Companion to Applied Regression: Appendices, 3rd ed., SAGE Publications Inc, 2019.
[50] M.E. Brooks, K. Kristensen, K.J. van Benthem, A. Magnusson, C.W. Berg, A. Nielsen, et al., glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling, R J 9 (2017) 378–400, https://doi.org/10.32614/rj-2017-066.
[51] K. Barton, MuMIn: multi-model inference. https://doi.org/10.32614/CRAN.package.MuMIn.
[52] R. Lenth, H. Singmann, J. Love, P. Buerkner, M. Herve, emmeans: estimated marginal means. R package version 1.4. 4, Am. Stat 34 (2020) 206–221.
[53] O.R. Lindsley, Precision teaching's unique legacy from B. F. Skinner, J. Behav. Educ. 1 (1991) 253–266, https://doi.org/10.1007/BF00957007.
[54] C. Binder, Behavioral fluency: evolution of a new paradigm, Behav. Anal. 19 (1996) 163–197.
[55] E. Gallagher, R. Bones, J. Lombe, Precision teaching and education: is fluency the missing link between success and failure? Irish Educ. Stud. 25 (2006) 93–105, https://doi.org/10.1080/03323310600597642.
[56] M. Tagliabue, G. De Cet, F. Orsini, M. Gastaldi, R. Rossi, A comparative analysis of feedback delivery modality within a precision teaching protocol to enhance drivers' lane maintenance, Transp. Res. Interdiscip. Perspect. 28 (2024) 101266, https://doi.org/10.1016/j.trip.2024.101266.
[57] R. Rossi, G. De Cet, E. Gianfranchi, F. Orsini, M. Gastaldi, How precision teaching can shape drivers' lateral control over time, Transp. Res. Procedia. 62 (2022) 565–572, https://doi.org/10.1016/j.trpro.2022.02.070.
[58] E.L. Harbeck, A.I. Glendon, T.J. Hine, Reward versus punishment: reinforcement sensitivity theory, young novice drivers' perceived risk, and risky driving, Transp. Res. Part F Traffic Psychol. Behav. 47 (2017) 13–22, https://doi.org/10.1016/j.trf.2017.04.001.
[59] E. Adell, A. Várhelyi, M. Alonso, J. Plaza, Developing human-machine interaction components for a driver assistance system for safe speed and safe distance, IET Intell. Transp. Syst. 2 (2008) 1–14, https://doi.org/10.1049/iet-its:20070009.
[60] J.P. Cabral, G.B. Remijn, Auditory icons: design and physical characteristics, Appl. Ergon. 78 (2019) 224–239, https://doi.org/10.1016/j.apergo.2019.02.008.
[61] T. Horberry, C. Mulvihill, M. Fitzharris, B. Lawrence, M. Lenne, J. Kuo, et al., Human-centered design for an in-vehicle truck driver fatigue and distraction warning system, IEEE Trans. Intell. Transp. Syst. 23 (2022) 5350–5359, https://doi.org/10.1109/TITS.2021.3053096.
[62] S. Sahebi, H. Nassiri, Assessing public acceptance of connected vehicle systems in a new scheme of usage-based insurance, Transp. Res. Rec. 2625 (2017) 62–69, https://doi.org/10.3141/2625-07.
[63] B. Palat, G. Saint Pierre, P. Delhomme, Evaluating individual risk proneness with vehicle dynamics and self-report data – toward the efficient detection of At-risk drivers, Accid. Anal. Prev. 123 (2019) 140–149, https://doi.org/10.1016/j.aap.2018.11.016.
[64] R. Sangrar, L.E. Griffith, L. Letts, J. Ma, M.M. Porter, B. Vrkljan, Refreshing older adults' driving skills (ROADSkills): a randomized controlled trial examining the effect of video feedback, Transp. Res. Part F Traffic Psychol. Behav. 90 (2022) 35–46, https://doi.org/10.1016/j.trf.2022.07.019.
[65] C. Dijksterhuis, B. Lewis-Evans, B. Jelijs, O. Tucha, D. de Waard, K. Brookhuis, In-car usage-based insurance feedback strategies. A comparative driving simulator study, Ergonomics 59 (2016) 1158–1170, https://doi.org/10.1080/00140139.2015.1127428.
[66] D. Mortimer, J.S. Wijnands, A. Harris, A. Tapp, M. Stevenson, The effect of 'smart' financial incentives on driving behaviour of novice drivers, Accid. Anal. Prev. 119 (2018) 68–79, https://doi.org/10.1016/j.aap.2018.06.014.
[67] A. Vaezipour, A. Rakotonirainy, N. Haworth, P. Delhomme, A simulator study of the effect of incentive on adoption and effectiveness of an in-vehicle human machine interface, Transp. Res. Part F Traffic Psychol. Behav. 60 (2019) 383–398, https://doi.org/10.1016/j.trf.2018.10.030.