



# Drivers' performances and their subjective feelings about their driving during a 40-min test on a circuit versus a dynamic simulator

Elise Gemonet, Clément Bougard, Vincent Honnet, Marion Poueyo, Stéphane Masfrand, Daniel Mestre

## ► To cite this version:

Elise Gemonet, Clément Bougard, Vincent Honnet, Marion Poueyo, Stéphane Masfrand, et al.. Drivers' performances and their subjective feelings about their driving during a 40-min test on a circuit versus a dynamic simulator. Transportation Research Part F: Traffic Psychology and Behaviour, 2021, 78, pp.466-479. 10.1016/j.trf.2021.03.001 . hal-03278072

**HAL Id: hal-03278072**

**<https://hal.science/hal-03278072>**

Submitted on 21 Oct 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

# Drivers' performances and their subjective feelings about their driving during a 40-min test on a circuit versus a dynamic simulator

Elise Gemonet<sup>1,2</sup>, Clément Bougard<sup>2</sup>, Vincent Honnet<sup>2</sup>, Marion Poueyo<sup>2</sup>, Stéphane Masfrand<sup>2</sup> & Daniel R. Mestre<sup>1</sup>

1 - Aix-Marseille University, CNRS, ISM Marseille, France.

2 - Groupe PSA, Centre technique de Vélizy, Vélizy-Villacoublay, Cedex, France

---

## Abstract

Car manufacturers expect driving simulators to be reliable **research and development** tools. Questions arise, however, as to whether drivers' behavior on simulators exactly matches that observed when they are driving real cars. Drivers' performances and their subjective feelings about their driving were compared between two groups during a 40-min driving test on the same circuit in a real car (n=20) and a high-fidelity dynamic simulator (n=27). Their speed and its variability, the braking force and the **engine revolutions per minute (rpm)** were recorded five times on a straight line and three times on a curve. The differences observed in these measurements between circuit driving (CD) and simulator driving (SD) from the 6th to 40th minute showed no significant changes during the drive. The drivers also completed the **NASA Raw Task Load Index (NASA RTLX)** questionnaire and the Simulator Sickness Questionnaire (SSQ) and estimated the ease and standard of their own driving performances. These subjective feelings differed significantly between the two groups throughout the experiment. The SD group's scores on the NASA RTLX and SSQ questionnaires increased with time and the CD group's perceived driving quality and ease increased with time, reaching non-significantly different levels from their usual car driving standards by the end of the drive. These findings show the existence of a fairly good match between real-life and simulated driving, which stabilized six minutes after the start of the test, regardless of whether the road was straight or curved. These objective findings and subjective assessments suggest possible ways of improving the match between drivers' performances on simulators and their real-life driving behavior.

---

**Keywords:** Simulated driving, Real car driving, Relative validity, Adaptation, Driving behavior, Subjective ratings

## 1. Introduction

Driving simulators are an essential part of car manufacturers' **research and development (R&D)** activities. They can be used for various purposes, including the development and validation of systems such as Advanced Driver Assistance Systems (ADAS), In-Vehicle Information systems (IVIS), including Lane Keeping Assist (Ishida & Gayko, 2004), Adaptive Cruise Control (ACC) and obstacle avoidance devices (Eichelberger & McCart, 2016). Driving simulators can also be used to investigate the impact of these systems on human factors and test drivers' performances under various circumstances (when they are

subject to stress, distraction, drowsiness, fatigue, etc.) (Jacobé De Naurois, Bourdin, Bougard, & Vercher, 2018; Paredes, Ordoñez, Ju, & Landay, 2018; Young & Regan, 2007). Studies on simulators have been conducted for drawing up driving quality and regulations, as well as in the context of product litigation and recalls (Weir, 2010). One of the main advantages of driving simulators is that the testing conditions can be repeated and adapted in order to obtain useful data without exposing the drivers to any serious risks (Weir, 2010).

Driving is a complex activity involving many factors (Galy & Mélan, 2015). Since driving performances depend on many technical and human factors, replicating these performances artificially is a considerable challenge. The question as to the validity of driving simulators has therefore persisted for several decades. Validity is usually thought to depend on an external component and an ecological component. External validity relates to causal relationships which can be taken to involve various factors (such as the persons involved, the settings, etc. ) (Calder, Phillips, & Tybout, 1983). Authors studying ecological validity have compared real-life and simulated driving conditions in order to model the gap between the two conditions (Araujo, Davids, & Passos, 2007). When modelling human factors in the field of virtual reality, ecological validity is a crucial point. The distinction has been made here between absolute and relative behavioral validity (Blaauw, 1982). Absolute validity is obtained when the users' behavior in terms of various driving parameters recorded during simulated driving is identical to that observed under real-life driving conditions. Relative validity is what occurs when similar tendencies are observed between the behavior observed under both simulated and real-life driving conditions (Godley, Triggs, & Fildes, 2002; Klüver, Herrigel, Heinrich, Schöner, & Hecht, 2016). Ecological validity can thus be qualified as the degree of individuals' ability to adopt similar responses in a simulated environment to those produced under real life conditions, and therefore as their propensity to respond appropriately to various stimuli. Ecological validity can be measured in terms of behavioral and psychological parameters, which are complementary. As far as driving behavior is concerned, simulated vehicles are usually controlled via three main devices: gas pedals, brake pedals and steering wheels (Sahami, Jenkins, & Sayed, 2009). Focusing on details of the longitudinal parameters helps to elucidate drivers' behavior more closely. As far as the psychological aspects are concerned, drivers' mental workload contributes importantly to the performance of complex tasks and has been recognized as a weighting factor contributing to the safety of driving (Costa, Costa, Pereira, & Arezes, 2019). This factor can be measured via the global score and the subscale scores (Galy, Paxion, & Berthelon, 2017) on the National Aeronautics and Space Administration Raw Task Load Index NASA questionnaire (RTLX), which is easier to apply and more sensitive than the original version, the NASA Task Load Index (NASA TLX) (Hart, 2006) questionnaire, which is widely used in the field of driving simulation research (Paxion, Galy, & Berthelon, 2014).

In most studies, the drivers are using a simulator for the first time as if it was a new vehicle, and their driving performances are liable to evolve with time. The driver is therefore allowed a short period of adaptation, which is sometimes called learning, training or familiarization phase, to become accustomed to this new vehicle. This initial phase can last anything from two minutes (Baas, Charlton, & Bastin, 2000) to two full days (O'Neill et al., 1999). The goal to incorporate an adaptation period is to provide participants with an opportunity to utilize their previously obtained driving skills (since in many driving simulator studies a prerequisite for participation is a valid driving license and some driving experience) to a specific driving simulator. However, this is part of an empirical approach and only a few

studies have examined these issues from a more rational point of view. The adaptation is a change in mental representations that can change over time and as they are not true copies of objective reality they may diverge from it considerably (Bellet, Bailly-Asuni, Mayenobe, & Banet, 2009). This clearly might be applied to the appropriation of a new vehicle (real as well as simulated). Wege et al. (2013) further developed this framework, including aspects of driving performance as well as information processing and/or mental workload. This type of adaptation involves concomitant processes, including (1) the transfer of motor and cognitive skills that are primarily needed to operate the simulator correctly; (2) the transfer of cognitive and mental abilities that are needed to maneuver correctly and respond to the simulated environment; and (3) the physiological transfer in which the individual must adapt to the physiological stimuli of the simulated system that may differ from what the driver is used to feel when driving a car in the real world. During the process of adaptation to a vehicle, drivers' performances tend to evolve, ending in the stabilization of various parameters (Ronen & Yair, 2013), involving their cognitive skills, cognitive abilities and physiological aptitudes. Previous authors have reported that the time required by participants to adapt to a driving simulator ranged between 5 seconds (Brandtner et al., 2019) and 15 minutes (Ronen & Yair, 2013). Others have observed that even during 15 minutes of driving, no adaptation occurred (Sahami et al., 2009). To the best of our knowledge, however, no studies have ever focused so far on longer periods of adaptation on a driving simulator, or made any direct comparisons with drivers' performances on real vehicles. Assessing behavioral adaptation in this context requires looking at the temporal pattern of the driver's appropriation of the new vehicle (a real car or a simulator).

In addition, previous studies have shown that in driving simulators, drivers adapt their behavior differently, depend on the type of road involved (urban, straight or curved road) (Ronen & Yair, 2013). These differences were also observed on real roads, circuits and simulators (Bobermin, Silva, Ferreira, Guedes, & Santos Baptista, 2019; Knapper, Christoph, Hagenzieker, & Brookhuis, 2015; Varotto, Farah, Bogenberger, van Arem, & Hoogendoorn, 2020). All these differences seem to point to the need to distinguish between drivers' performances on straight roads and curvilinear sections.

Lastly, one of the major issues that arise in driving simulator studies is simulator sickness symptoms, which can occur within quite a short time in some participants (Brooks et al., 2010). It has in fact been widely recognized that simulator sickness impairs driving performances on simulators (Klüver et al., 2016). In studies in this field, the impact of simulator sickness on drivers' behavior cannot therefore be neglected. Questionnaires such as the simulator sickness questionnaire (SSQ) (Kennedy, Lane, Berbaum, & Lilienthal, 1993) can be used in these studies in order to understand drivers' behavior more clearly. The aim of the present study was to observe and compare drivers' behavior while driving a real car on a secured circuit and driving on a dynamic simulator during a period of 40 minutes. Since driving parameters are known to be road type dependent, a specially designed circuit composed of straight and curved parts was used. This design was reproduced in the driving simulator tests. In order to determine drivers' perception of their own driving ease and driving ability, several questionnaires, such as visual analogical scales and the NASA RTLX questionnaire, were used to assess their subjective feelings. In addition, the SSQ was applied in order to assess the impact of simulator sickness on the participants' driving performances and their subjective feelings. With a view to determining the evolution of the parameters of interest during the test, several self-assessments were carried out during the drive. Based on

the literature, it can be hypothesized that objective and subjective driving data will show an initial gap between driving in a real vehicle and driving in a simulator that will decrease until it disappears. And that in a disruptive way, simulator sickness symptoms may interfere with the evolution of these driving parameters and drivers' subjective feelings. This questioning allowed us to go further in time than previous studies on driver adaptation to simulators with a direct comparison in a real vehicle.

## 2. Material and methods

### 2.1. Participants

Forty-seven drivers volunteered to participate in this study. They were recruited at an internationally reputed automotive company and participated for free. For safety reasons, employees with medical problems such as heart disease, epilepsy or pregnant women were not eligible to participate. All the participants had a normal or corrected no normal vision and hearing. The drivers had not taken any drugs, alcohol, or medications that could impair alertness or driving ability, at a minimum, of 24 hours prior to the study. They had passed their driving license more than 10 years ago and drove to work every day making around 1600 km per month. Their usual car was equipped with a manual gearbox and was not a powerful Peugeot 308 (used or simulated in the tests). Before participating in any tests, they signed an informed consent form and a safety questionnaire. They were informed that they could leave the experiment at any time. This study was approved by the local Ethics Committee and was conducted in keeping with the ethical standards defined in the 1964 Helsinki Declaration. The sample was divided into two independent groups: twenty participants (10 women and 10 men  $50 \pm 5$  years of age) were tested on the circuit in a real car and twenty-seven, on the simulator (7 women and 20 men  $38 \pm 13$  years of age). The drivers were given safety instructions before both driving sessions.

### 2.2. Vehicles and environments

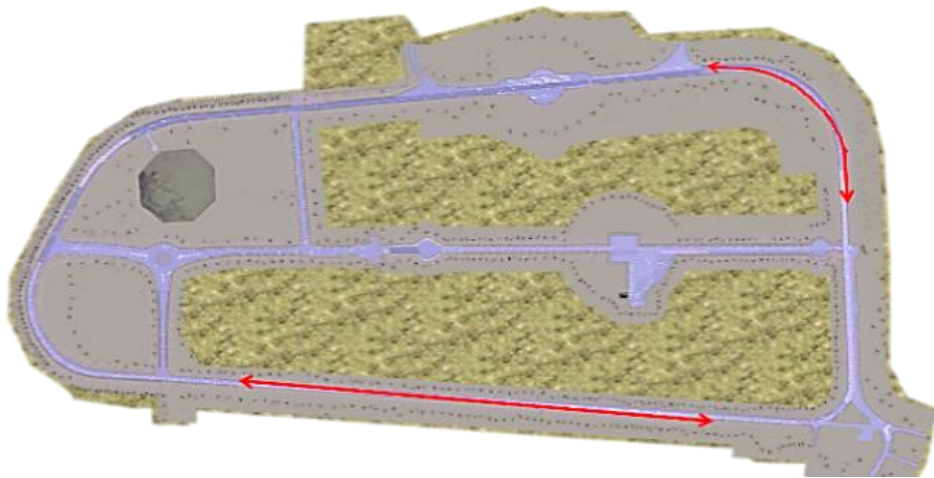
The real car used here was a Peugeot 308 GT Line, with a 130 horsepower engine, a manual gearbox and power steering. The driving simulator used was the Sherpa2 at the Groupe PSA's technical center in Vélizy (France). Its motion system is composed of a hexapod set on an X-Y platform with 8 degrees of freedom. Its cabin was a fully equipped half-cab with a manual gearbox Citroen C1 model with 3 screens covering a 180-degree wide field. The simulated car was based on the same dynamic model as the real car we used. This specific choice of real car and the corresponding dynamic model in the simulator was based on the need to place the participants in a novel situation. None of the participants had ever used a driving simulator before, nor a Peugeot 308.

The tests performed under real driving conditions were conducted in a closed arena, which was specially adapted to the requirements of the experiment. The simulated environment was a realistic 3-D model of the real circuit (Figure 1), and the itinerary (Figure 2) was the same for all the participants. The road included crossings, roundabouts and various signals such as traffic lights, stops, etc. To prevent boredom and anticipation of the path travelled, participants took a pre-defined randomized route prescribed during the journey by a GPS voice. They drove from one to five times along the same section of the road and covered about 16.5 km in all during the test, which lasted for approximately 40 minutes.





**Figure 1:** The start of the driving session: on the driving simulator in the picture on the left and on the circuit in the picture on the right. The vehicles used were the grey ones in the middle.



**Figure 2:** Map of the environment. The red arrows show the two parts of the road on which the drivers were tested (on a straight line and a curve).

### ***2.3. Procedure***

First, all the participants, whether they were tested under real driving or simulated driving conditions, filled in the Simulation Sickness Questionnaire (SSQ) and answered a question about how they perceived the quality of their own driving and another one about how easy they found driving when using their own vehicle in both cases.

They were then informed that they would have to drive for 40 minutes, obeying the traffic signs and taking the course specified by the GPS voice. In the real driving condition, an experimenter was present during the whole session on the passenger seat next to the driver, managing the recording instruments. In the simulated driving condition, the experimenter, who was in a nearby control room, kept in touch with the driver via an audio system. During the drive, the participants were asked seven times by a recorded voice how easy they found the task and how well they thought they were driving. These questions were always asked on the same part of the circuit, which differed from the parts on which the driving performances were analyzed.

After the driving test, all the participants filled in the SSQ once again, the NASA RTLX questionnaire, a question about whether they felt they drove well when using their own vehicle and another one about how easy they found it to drive their own vehicle.

## **2.4. Measurements and data processing**

### *2.4.1. Simulator Sickness Questionnaire*

The Simulator Sickness Questionnaire (SSQ) contains sixteen questions about personal feelings which commonly occur along with simulation sickness. We used the Canadian-French version of the questionnaire translated by Québec University's Cyberpsychology Laboratory in Outaouais, as suggested by Kennedy et al. (1993). The global score and the three subscale scores were calculated for each participant. The 'Nausea' score was based on answers to some of the questions about physiological discomfort, the 'Oculo-motor' score was based on answers to some of the questions about oculo-motor fatigue, and the 'Discomfort' score was based on answers to some of the questions about the participants' spatial perception of their body.

### *2.4.2. NASA Raw Task Load Index*

The NASA Raw Task Load Index (NASA-RTLX) is a six-dimensional questionnaire (mental demand, physical demand, temporal demand, performance, effort and satisfaction) used a posteriori to evaluate the subjective workload exerted by a task. Each rating scale ranges from 0 to 20. The mean scores obtained on the six questions and each question individually were analyzed.

### *2.4.3. Perceived driving quality*

To obtain an idea of how well the drivers evaluated they were performing during the test, the question "How well do you feel you have been driving?" was asked seven times by a recorded voice (see the original question in Complementary information 7.1). Participants had to answer on a scale ranging from 1 (very badly) to 10 (very well). Before and after the performing test, they were asked about driving their own vehicle, while during the test, they were asked about the ongoing drive, each time at the same point on the circuit. All these scores were subsequently analyzed.

### *2.4.4. Perceived ease of driving*

For the purpose of knowing how drivers felt during the test, the question "How easily do you feel you have been driving?" was asked seven times during the test by a recorded voice (see the original question in Complementary information 7.2), always at the same point on the circuit. Participants had to answer on a scale ranging from 1 "not at all easily" to 10 "very easily". Before and after the performing test, they were asked about driving their own vehicle, while during the test, they were asked about the ongoing drive, each time at the same point on the circuit. All these scores were subsequently analyzed.

### *2.4.5. Driving parameters*

In order to assess the drivers' performances in both conditions (on the circuit and on the driving simulator), several parameters involved in the longitudinal control of the vehicle were measured: vehicle longitudinal speed, longitudinal acceleration, number of rotation per minute of the engine (rpm), pressure in the brake pedal. The sampling frequency was set at 10 Hz in the case of the real car and 100 Hz in that of the driving simulator. The circuit on which

the drivers were tested included various portions such as straight lines of various lengths and curves with various radii of curvature. In order to standardize the data included in the analysis, however, only those recorded on two selected parts of the road were used: on a straight line traversed five times and a curve traversed three times (see red arrows in Figure 2). Using a specially designed Matlab® script, for each run on a portion, each participant's mean value of the vehicle longitudinal speed, the rpm and the pressure in the brake pedal were calculated. The longitudinal acceleration has been transformed to obtain the absolute longitudinal acceleration

In the simulated driving condition some of the participants ( $n = 13$ ) completed the whole driving test and were named SD group (Simulator Driving group). While some of the participants ( $n = 10$ ) did not complete the whole driving test. This group was named SDC (Simulator Drivers who did not Complete the test). In order to understand their reasons for abandoning, their answers to the SSQ were compared with the results obtained on the other groups. The data on the group SDC were not used for further analysis. In addition, since the data acquisition failed in the case of four of the participants in the SD group, their driving parameters were not included either in the statistical analysis. Only the driving parameters recorded on participants who completed the full driving session in the Circuit Driving condition (CD;  $n = 20$ ) and the Simulator Driving condition (SD;  $n = 13$ ) were therefore analyzed.

## ***2.5. Statistical analyses***

The answers to each questionnaire were discontinuous data and did not satisfy the normality condition. These data were therefore analyzed using non-parametric tests.

A Kruskal-Wallis test was applied in order to determine what differences existed in the SSQ scores between the 3 groups (the Circuit Driving (CD) group, the Simulator Driving (SD) group and the Simulator Driving group who did not complete the test (the SDC group). Whenever a significant difference was observed between two groups, a Mann-Whitney U-Test was conducted in order to examine this point more closely. In addition, a Wilcoxon test was performed in order to detect any differences between the pre and post-test values of the global score and the scores obtained on the three subscales by each group.

As with the answers to the NASA RTLX questionnaire, a Mann-Whitney U-test was applied in order to determine the existence of any significant differences between the two groups CD and SD.

In the analysis of perceived driving quality and perceived ease of driving, a Mann-Whitney U-Test was applied in order to detect the existence of any intergroup differences at each time point. In addition, a Friedman repeated measures analysis of variance was applied in order to determine how the scores evolved with time during the driving session. Lastly, a Wilcoxon test was applied in order to detect the existence of any differences between any two measurements recorded in each group during the test with a view to establishing whether the drivers' subjective feelings evolved during the drive.

A two-way repeated measures ANOVA [2 groups (the Circuit Driving and Simulated Driving groups)  $\times$  7 measurements performed during the driving session] was applied in order to establish whether the driving parameters differed between the two groups, how they



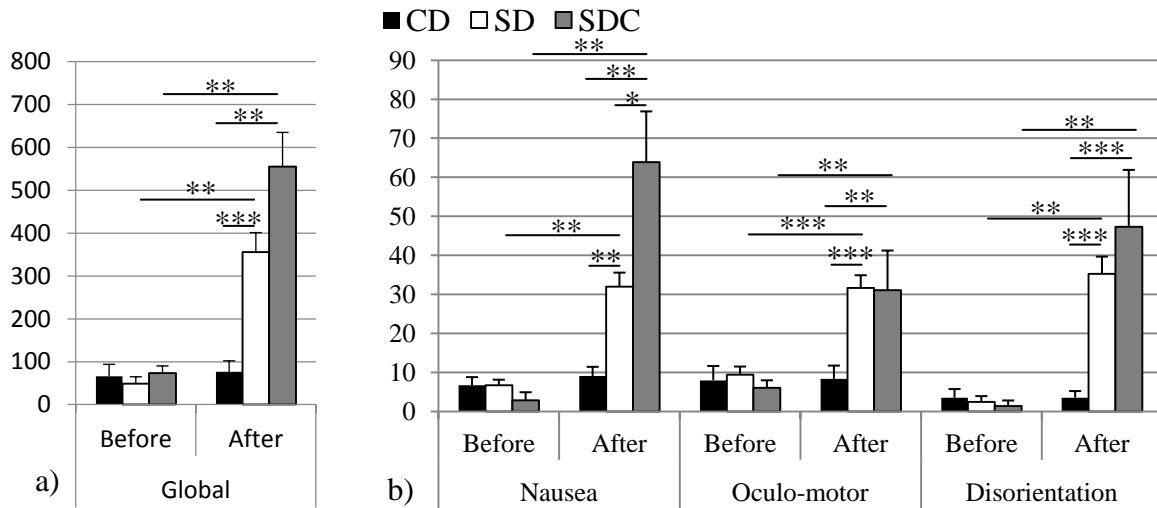
evolved during the driving session and whether any interactions were involved. A Bonferroni post-hoc analysis was applied whenever any significant differences were observed.

All the statistical analyses were conducted using Statistica® software (version 6.0). All statistical differences observed were taken to be significant at  $p < 0.05$ . Exact p-values were reported until they were over 0.001. Data were expressed as means  $\pm$  standard deviations.

### 3. Results

#### 3.1. Simulation Sickness Questionnaire

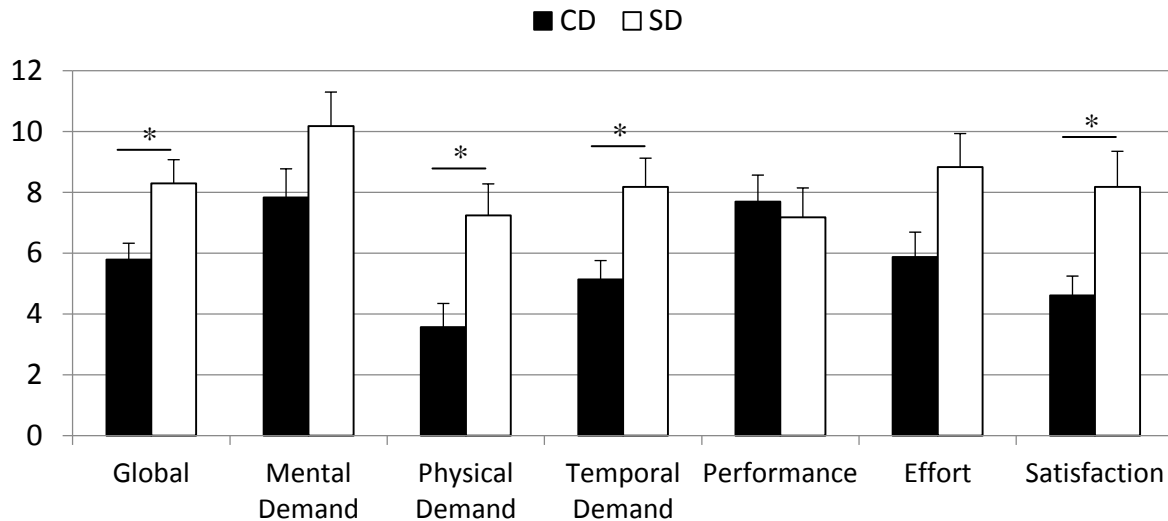
The Kruskal-Wallis test performed on the global scores and all the scores obtained on the subscales showed the existence of no significant differences between the three groups before the driving session (Figure 3), whereas significant differences were observed between the groups after the driving session. The Mann-Whitney U-test showed that after the driving session the SDC group had higher global scores than the CD group (76 vs 532,  $\times 7$ ,  $p = 0.002$ , Cohen's  $d = 1.593$ ), as well as higher scores on all the subscales [nausea (8 vs 63,  $\times 7$ ,  $p = 0.018$ , Cohen's  $d = 1.906$ ), oculo-motor (8 vs 31,  $\times 4$ ,  $p = 0.002$ , Cohen's  $d = 0.998$ ), disorientation (3 vs 47,  $\times 13$ ,  $p = 0.005$ , Cohen's  $d = 1.346$ )]. The SD group also had higher global scores than the CD group (76 vs 369,  $\times 4.5$ ,  $p = 0.003$ , Cohen's  $d = 1.700$ ), as well as higher scores on the various subscales [nausea (8 vs 31,  $\times 3.5$ ,  $p < 0.001$ , Cohen's  $d = 1.318$ ), oculo-motor (8 vs 31,  $\times 4$ ,  $p < 0.001$ , Cohen's  $d = 1.926$ ), disorientation (3 vs 35,  $\times 10$ ,  $p < 0.001$ , Cohen's  $d = 1.420$ )]. In the comparisons between the groups SDC and SD, a significant difference was observed in the case of nausea alone (31 vs 63,  $\times 2$  for SDC,  $p = 0.034$ , Cohen's  $d = 1.906$ ). The statistical analyses (Wilcoxon rank-sum tests) performed on the results obtained on each group showed the existence of no significant differences in the CD group's global and subscale scores on the SSQ questionnaire before versus after the driving session. In the groups SDC and SD, all the scores increased significantly between the pre- and post-tests (before versus after the driving session). The SD group's global scores increased (69 vs 369,  $\times 7$ ,  $p = 0.001$ ), as did the scores obtained by this group on the various subscales [nausea (6 vs 31,  $\times 5$ ,  $p = 0.002$ ), oculo-motor (9 vs 31,  $\times 3$ ,  $p < 0.001$ ), disorientation (2 vs 35,  $\times 14$ ,  $p = 0.003$ )]. The SDC group's global score increased (38 vs 532,  $\times 7.5$ ,  $p = 0.011$ ), as did the scores obtained by this group on the various subscales [nausea (2 vs 63,  $\times 31$ ,  $p = 0.011$ ), oculo-motor (6 vs 31,  $\times 5$ ,  $p = 0.011$ ), disorientation (1.4 vs 47,  $\times 34$ ,  $p = 0.011$ )].



**Figure 3:** a) The global SSQ score. b) The scores on the nausea, oculo-motor and disorientation subscales of the SSQ. Significant differences between groups CD and SD: \*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$ .

### 3.2. NASA Raw Task Load Index

The Mann-Whitney U-test indicated that the participants had higher global scores and some subscale scores in the NASA RTLX questionnaire after the simulated driving condition than in the real driving condition (Global: 5.7 vs 8.2,  $p = 0.049$ , Cohen's  $d = 0.733$ ; physical demand: 3.5 vs 7.2,  $p = 0.023$ , Cohen's  $d = 0.794$ ; temporal demand: 5.1 vs 8.1,  $p = 0.038$ , Cohen's  $d = 0.749$ ; satisfaction: 4.6 vs 8.1,  $p = 0.029$ , Cohen's  $d = 0.029$ ); whereas no significant differences were observed between the two groups in terms of the mental demand, performance and effort, or task difficulty scores (see Figure 4 for details).

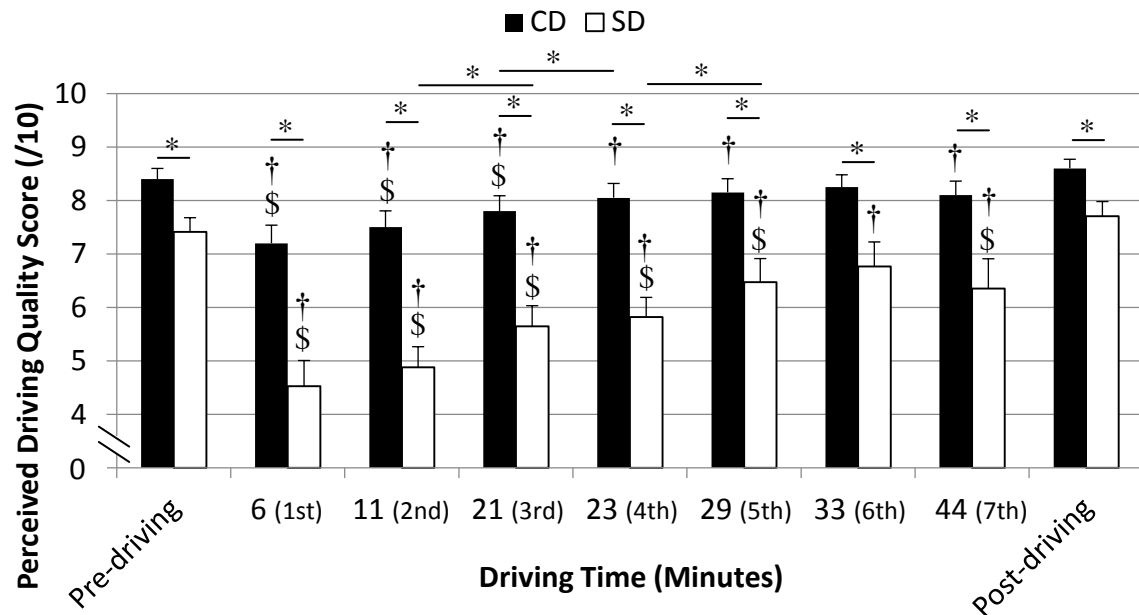


**Figure 4:** Mean values of the answers to the NASA RTLX depending on the driving conditions. Significant differences observed between groups CD and SD: \*  $p < 0.05$ .

### 3.3. Perceived quality of driving

The statistical analyses (Mann-Whitney tests) indicated that the scores obtained by groups CD and SD differed significantly before, during and after the driving session (Before:  $p =$

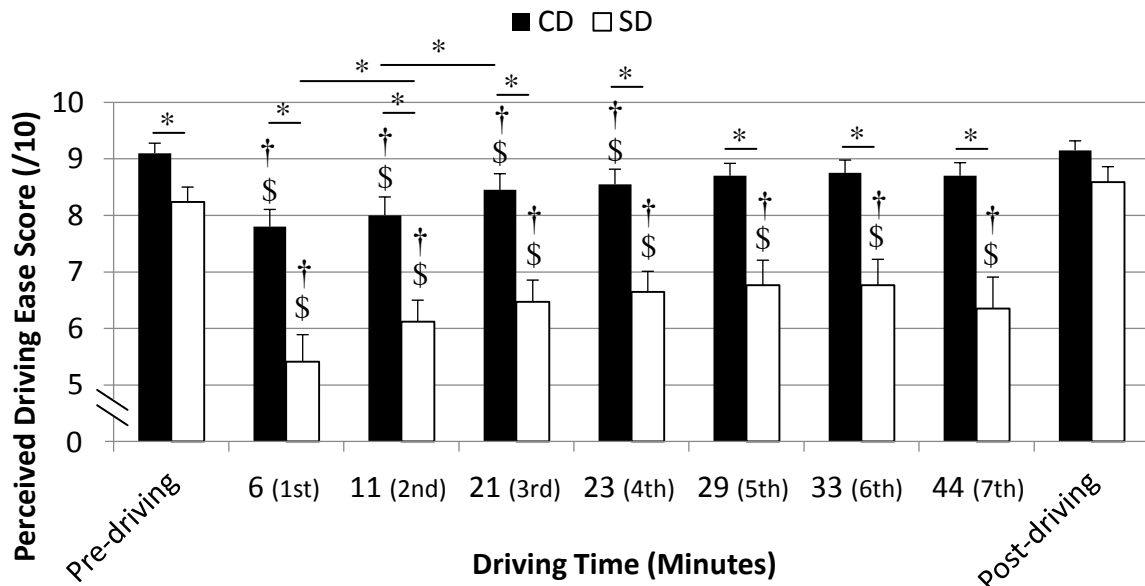
0.018, Cohen's  $d = 3.705$ ; 1st:  $p < 0.001$ , Cohen's  $d = 7.384$ ; 2nd:  $p < 0.001$ , Cohen's  $d = 8.246$ ; 3rd:  $p < 0.001$ , Cohen's  $d = 7.086$ ; 4th:  $p < 0.001$ , Cohen's  $d = 7.514$ ; 5th:  $p = 0.001$ , Cohen's  $d = 5.665$ ; 6th:  $p = 0.006$ , Cohen's  $d = 4.750$ ; 7th:  $p = 0.001$ , Cohen's  $d = 5.334$ ; After:  $p = 0.010$ , Cohen's  $d = 4.324$ ). A higher perceived quality of driving was always reported by group CD than by group SD (Figure 5). More specifically, during the driving session, the statistical analysis (Friedman repeated measures analysis of variance), showed that the CD group's perceived driving quality scores evolved with time ( $\chi^2_{(df=8)} = 50.7, p < 0.001$ ), as did those of the SD group ( $\chi^2_{(df=8)} = 81.4, p < 0.001$ ). The post hoc tests (Wilcoxon paired sign rank tests) indicated that the scores obtained by the group CD on the pre-driving question ( $8.4 \pm 0.1$ ) were higher than those obtained in response to the 1st ( $7.2 \pm 0.3, p = 0.002$ ), 2nd ( $7.5 \pm 0.3, p = 0.0049$ ) and 3rd questions ( $7.8 \pm 0.2, p = 0.025$ ). The scores increased significantly during the drive, between the 3rd and 4th ( $8.0 \pm 0.2, p = 0.043$ ) questions. The scores obtained by this group on the 1st, 2nd, 3rd, 4th ( $8.0 \pm 0.2$ ), 5th ( $8.1 \pm 0.2$ ) and 7th ( $8.1 \pm 0.2$ ) question were significantly lower than that obtained on the post driving question ( $8.6 \pm 0.1$ ; 1st:  $p < 0.001$ ; 2nd:  $p = 0.002$ ; 3rd:  $p = 0.005$ ; 4th:  $p = 0.016$ ; 5th:  $p = 0.040$ ; 7th:  $p = 0.032$ ). The pre-driving score obtained by the group SD ( $7.4 \pm 0.1$ ) was higher than those obtained in response to the 1st ( $4.5 \pm 0.3, p < 0.001$ ), 2nd ( $4.8 \pm 0.3, p < 0.001$ ), 3rd ( $5.6 \pm 0.3, p = 0.002$ ), 4th ( $5.8 \pm 0.3, p = 0.002$ ), 5th ( $6.4 \pm 0.3, p = 0.045$ ) and 7th questions ( $6.3 \pm 0.3, p = 0.023$ ). In addition, the SD group's scores increased significantly during the drive, between the 2nd and 3rd ( $p = 0.005$ ) questions and between the 4th and 5th ( $p = 0.003$ ) questions. All the scores obtained during the driving session were significantly lower than those obtained in response to the post-driving question ( $7.7 \pm 0.2$ ; 1st:  $p < 0.001$ ; 2nd:  $p < 0.001$ ; 3rd:  $p < 0.001$ ; 4th:  $p < 0.001$ ; 5th:  $p = 0.004$ ; 6th:  $p = 0.018$ ; 7th:  $p = 0.002$ ).



**Figure 5:** Mean value of the perceived driving quality scores. Significant differences between groups CD and SD: \*  $p < 0.05$ ; Significant differences between the pre-driving scores and the scores obtained on the questions asked at various times during the driving session: \$  $p < 0.05$ ; Significant differences between the post-driving scores obtained by each group and the scores obtained on the questions asked at various times during the driving session: †  $p < 0.05$ .

### 3.4. Perceived ease of driving

The results of the Mann-Whitney U-test indicate that the group CD obtained significantly higher scores than the group SD on almost all the questions: the pre-driving question ( $p = 0.016$ , Cohen's  $d = 0.908$ ), the 1st ( $p < 0.001$ , Cohen's  $d = 1.410$ ), 2nd ( $p = 0.001$ , Cohen's  $d = 1.242$ ), 3rd ( $p < 0.001$ , Cohen's  $d = 1.375$ ), 4th ( $p < 0.001$ , Cohen's  $d = 1.406$ ), 5th ( $p = 0.001$ , Cohen's  $d = 1.324$ ), 6th ( $p < 0.001$ , Cohen's  $d = 1.308$ ), and 7th question ( $p < 0.001$ , Cohen's  $d = 1.322$ ); whereas no difference was observed between the scores obtained by the two groups on the post-driving question (Figure 6). The statistical analyses (the Friedman repeated measures analysis of variance) indicated that the perceived quality of driving evolved during the driving session in the case of both group CD ( $\chi^2_{(df=8)} = 49.4$ ,  $p < 0.001$ ) and group SD ( $\chi^2_{(df=8)} = 58.9$ ,  $p < 0.001$ ). More specifically, the Wilcoxon paired sign rank tests showed that the group CD obtained a higher score on the pre-driving question ( $9.1 \pm 0.1$ ) than on the 1st ( $7.8 \pm 0.3$ ,  $p = 0.002$ ), 2nd ( $8.0 \pm 0.3$ ,  $p = 0.004$ ), 3rd ( $8.4 \pm 0.2$ ,  $p = 0.036$ ) and 4th ( $8.5 \pm 0.2$ ,  $p = 0.046$ ) question. In addition, the CD group's scores increased significantly during the drive between the 2nd and the 3rd questions ( $p = 0.011$ ). The scores obtained by this group on the 1st ( $p = 0.001$ ), 2nd ( $p = 0.003$ ), 3rd ( $p = 0.016$ ) and 4th ( $p = 0.021$ ) question were significantly lower than that obtained on the post driving question ( $9.1 \pm 0.1$ ). In the case of the group SD, the score obtained on the pre-driving question ( $8.2 \pm 0.2$ ) was higher than those obtained on 1st ( $5.4 \pm 0.4$ ,  $p < 0.001$ ), 2nd ( $6.1 \pm 0.3$ ,  $p < 0.001$ ), 3rd ( $6.4 \pm 0.3$ ,  $p = 0.004$ ), 4th ( $6.6 \pm 0.3$ ,  $p = 0.006$ ), 5th ( $6.7 \pm 0.4$ ,  $p = 0.018$ ), 6th ( $6.7 \pm 0.4$ ,  $p = 0.018$ ) and 7th ( $6.3 \pm 0.5$ ,  $p = 0.007$ ). In addition, the scores obtained during the drive by the group SD increased significantly between the 1st and 2nd question ( $p = 0.038$ ). The scores obtained by this group on all the questions asked during the driving session were significantly lower than those obtained on the post-driving question ( $8.5 \pm 0.2$ ; 1st:  $p < 0.001$ ; 2nd:  $p < 0.001$ ; 3rd:  $p < 0.001$ ; 4th:  $p = 0.001$ ; 5th:  $p = 0.002$ ; 6th:  $p = 0.002$ ; 7th:  $p = 0.001$ ).



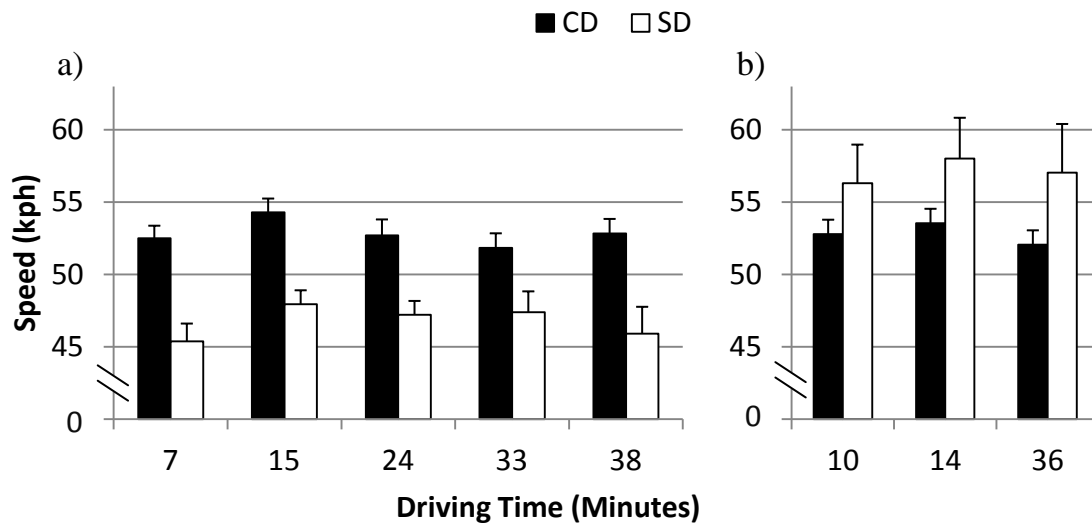
**Figure 6:** Mean value of the self-perceived driving quality. Significant differences between groups CD and SD: \*  $p < 0.05$ ; Significant differences between the pre-driving score and the scores obtained on the questions

answered during the driving session: \$  $p < 0.05$ ; Significant differences between the post-driving score and the scores obtained on the questions asked during the driving session: †  $p < 0.05$ .

### 3.5. Driving performances

#### 3.5.1. Mean speed

Based on the mean driving speed, the CD group drove significantly faster on the straight line (Figure 7a) than the SD group (52.8 kph vs 46.7 kph,  $\approx + 6$  kph,  $p < 0.001$ , Cohen's  $d = 1.345$ ), whereas no significant difference was observed between the two groups on the curved part of the road (Figure 7b). No significant effect of the driving time on this parameter was observed in either group.

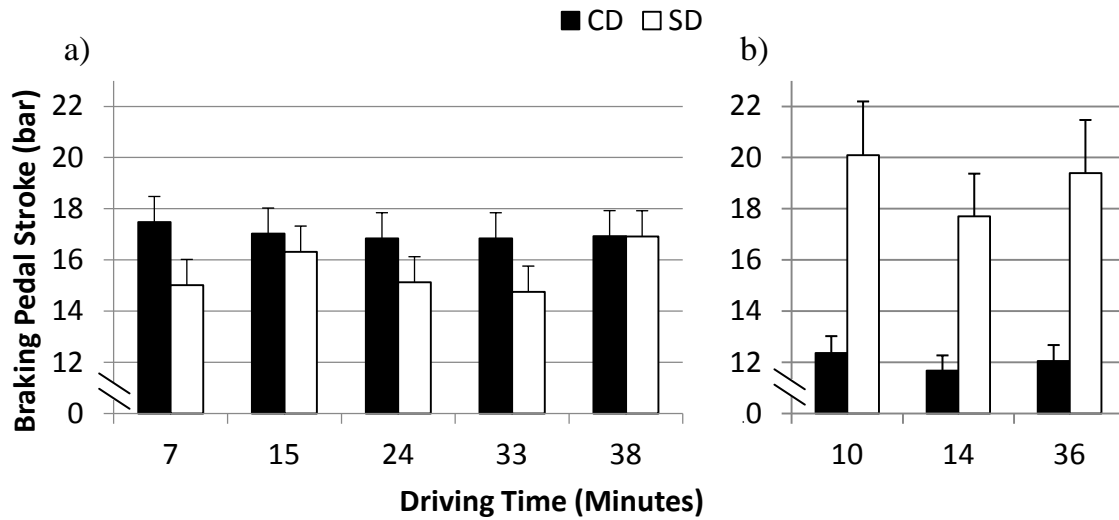


**Figure 7:** Mean speed during the driving session a) on a straight part of the drive, b) on a curved part of the drive.

#### 3.5.2. Mean brake pedal pressure

The braking pedal stroke did not differ significantly between the two groups on the straight line (Figure 8a). On the curved part of the road, however, the group SD activated the braking pedal about 37% more strongly than the group CD (12.0 bar vs 19.0 bar,  $p < 0.001$ , Cohen's  $d = 1.319$ ) (Figure 8b). No significant effects of the driving time on this parameter were observed in either group.

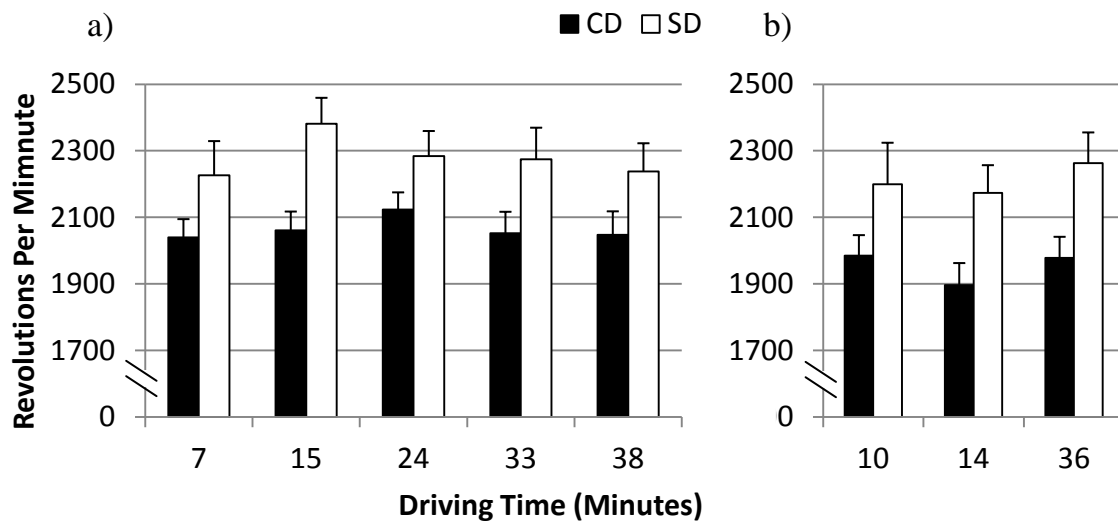




**Figure 8:** Mean braking pedal force exerted during the driving session a) on a straight part of the drive, b) on a curved part of the drive.

### 3.5.3. Mean rpm

The number of revolutions per minute (rpm) of the engine was significantly higher in the SD group, both on the straight line (2064 rpm vs 2281 rpm,  $\approx +200$  rpm,  $p = 0.007$ , Cohen's  $d = 0.746$ ), and on the curved part of the circuit (1865 rpm vs 2244 rpm,  $\approx +400$  rpm,  $p < 0.001$ , Cohen's  $d = 1.171$ ) (Figure 9). No significant effects of the driving time were observed in either group.

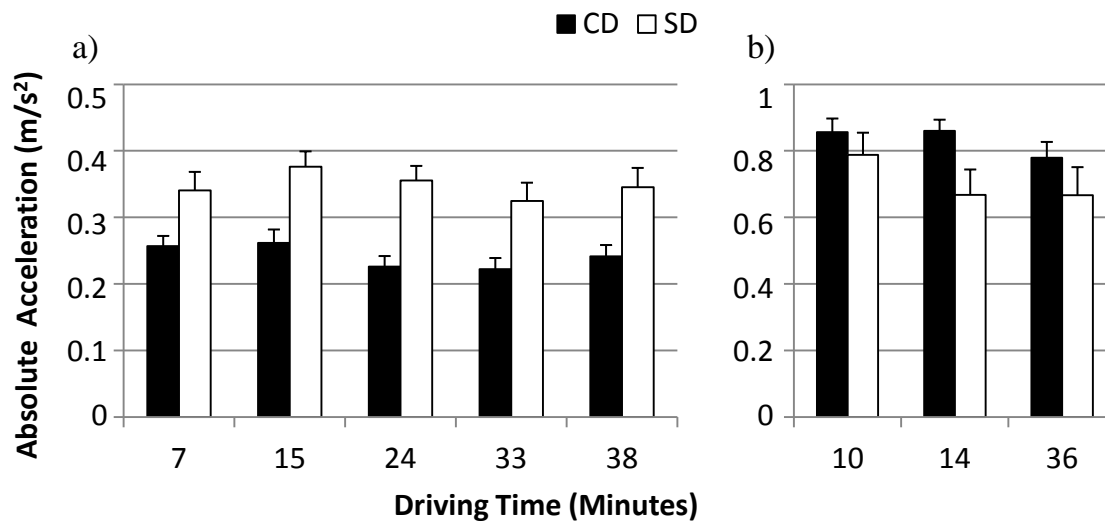


**Figure 9:** Mean rpm during the driving session a) on the left, on a straight part of the drive, b) on a curved part of the drive.

### 3.5.4. Mean Absolute Acceleration

The mean value of the absolute longitudinal acceleration on the straight line was significantly higher in the SD group than the CD group (0.24 m/s<sup>2</sup> vs 0.34 m/s<sup>2</sup>, +30%,  $p < 0.001$ , Cohen's  $d = 1.171$ ) (Figure 10).

0.001, Cohen's  $d = 1.258$ ) (Figure 10a), whereas no significant differences were observed between the two groups on the curved part of the road (Figure 10b). No significant effects of the driving time on this parameter were observed in either group.



**Figure 10:** Mean absolute acceleration during the driving session a) on the left, on a straight part of the drive, b) on a curved part of the drive.

#### 4. Discussion

The main aim of this study was to assess and compare the driving performances produced under real and simulated conditions, using for the first time a SHERPA<sup>2</sup> dynamic simulator and a new car for all the participants. Driving simulators are often used by car manufacturers for various purposes such as testing driving assistance systems, human machine interfaces, etc. The question still arises, however, as to whether drivers' performances are exactly the same on simulators as on real vehicles. A gap is frequently observed between the two driving conditions, the amplitude of which depends on the physical and psychological state of the drivers (sleepiness, distraction, stress, etc.), as well as on the characteristics of the simulator (Klüver et al., 2016). To test this hypothesis, the participants were asked to answer subjective questionnaires about how easy they found the drive and the quality of their driving, the NASA RTLX and simulator sickness questionnaires were also applied in order to determine the participants' feelings on these topics. Secondly, it was proposed to establish whether the participants adapted differently to the new vehicle, the real vehicle vs the simulator, during a fairly long drive (lasting approximately 40 minutes). Despite a large drop-out rate in the simulator group, the main findings obtained here show the existence of differences between the two groups who completed the experiment. Driving simulation caused a higher simulation sickness and an increase in declared workload. In line with the drivers' self-reported feelings, according to which the SD group, those who were tested on the simulator, felt less at ease and less confident in the quality of their driving than the group CD, those driving on a real circuit. In addition, it emerged that these differences remained unchanged throughout the 40 minutes driving test. Concerning the driving parameters, no evolution appeared with time. The driver's behavior showed identical data or a constant gap between the data.

Several authors have reported that driving in a dynamic simulator can induce symptoms of simulation sickness, which may impair the subjects' driving performances (Smyth, Jennings, Mouzakitis, & Birrell, 2018). Even though the sickest subjects were removed from the study, differences in the drivers' susceptibility to simulator sickness may have increased the variability of the driving performances observed in the SD group. In line with this possibility, some of the answers to the NASA RTLX questionnaire suggest that the group SD may have experienced greater physical demands than the CD group. On the one hand, this difference may be due to the simulation sickness experienced by some of the members of the SD group, as reported especially on the nausea subscale, which deals with symptoms such as sweating and stomach awareness (Kennedy et al., 1993). On the other hand, participants in the SD group may have had greater difficulty in controlling the vehicle's dynamics on the simulator and may therefore have had to make greater use of the pedals and steering wheel (Foy & Chapman, 2018). In addition, the present results also show that the CD group felt more aware of the temporal demands than the SD group. It was previously reported that learners driving under real-life conditions under the supervision of an instructor, as was the case in the present study, stated that they were aware of greater temporal demands than when driving alone (Milleville-Pennel & Charron, 2015). The SD group expressed greater satisfaction with their own performances than the CD group. This may have been due to the difficulty they had in handling the simulator at the beginning of the test. The authors of several studies have indeed reported that people who have had some difficulty initially in performing any task will be all the more satisfied when they eventually succeed (Guajardo, 2016). The answers obtained here to the questions about the ease of the drive and the quality of the participants' driving confirm these hypotheses. Although drivers have been previously found to assess their driving abilities accurately in comparison with the driving parameters observed (Hooft Van Huysduynen, Terken, & Eggen, 2018), the SD group consistently declared lower perceptions of ease and quality than the CD group during the 40 minute drive than the CD group.

In terms of the driving parameters, the straight part of the road and the curved part can be regarded as two distinct types of road. On the one hand, it turned out that the drivers drove faster on the straight line in the real car than on the driving simulator. This finding contrasts with those obtained in some previous studies, and with the well-established idea that drivers often drive faster in simulators (Kazemzadehazad, Monajjem, Larue, & King, 2018; Llopis-Castelló et al., 2016). However, some authors have also observed, as in the present study, that higher speeds can be recorded under real driving conditions than under simulated conditions (Llopis-Castelló et al., 2016; Santos, Merat, Mouta, Brookhuis, & de Waard, 2005). The tendency for drivers to show greater compliance with the speed limits (50 kilometers per hour in this study) in the simulator than in a normal vehicle might account for this fact. It was recently established that drivers who report that they are frequently guilty of speeding in their own vehicles do not replicate this behavior in a simulator (Tement et al., 2019). Although they were instructed to drive as naturally as possible, the fact of being constantly observed by the experimenter using a video camera and a micro may have increased participants' motivation to obey the driving rules. This compliance with the speed limits in driving simulators is consistent with the higher global and temporal demand scores obtained here on the NASA RTLX questionnaire. Some authors have previously reported that speed decreases with the mental workload (Törnros & Bolling, 2006). An increase in the absolute acceleration, and therefore in the variability of the speed, was also observed here. This finding is in agreement with previous studies in which greater variability of speed was observed under

simulated than real driving conditions (Santos et al., 2005). In agreement with the results based on the NASA RTLX questionnaire, greater speed variability may also be attributable to the increase in the mental workload associated with the use of driving simulators (Horberry, Anderson, Regan, Triggs, & Brown, 2006; Horrey & Wickens, 2004; Zheng, Tai, & McConkie, 2005). The answers obtained here about the drivers' ease and the quality of their driving also indicate that the greater variability of the speed observed in the driving simulator may be attributable to the fact that the participants had difficulties in maintaining a stable speed during the simulated drive on the straight part of the road. These findings may be linked to other driving parameters such as the acceleration or the braking possibilities on the simulator. As in the case of the force exerted on the braking pedal, no significant differences were detected in terms of the speed between the two groups on the straight line; whereas a higher average rpm was observed in the group tested on the simulator than in the CD group, which suggests that although the participants were driving more slowly than in real life, they maintained a lower gear ratio, possibly because they misjudged the rpm due to either the absence of vibration feedback (Tachiiri, Tanaka, & Sano, 2017), or the artificial sound of the engine emitted by the driving simulator (Pakdamanian, Feng, & Kim, 2018).

On the other hand, on the curved part of the road, no significant differences were observed between the CD group and the SD group as far as the speed was concerned. One possible explanation for this finding is that the participants were not limited by the speed regulations but by the physical constraints of the curve. In fact, the speed was limited to 80 kph on the curved part of the circuit. Driving as fast as that without going off the road was actually impossible. This limitation might also explain the fact that comparable values were recorded in both groups as far as the absolute acceleration, and hence the variability of the speed, was concerned. It has been established that under simulated driving conditions drivers may tend to drift around 6 kph beyond the speed limits, but that most of the time they reduce their speed on curves and then compensate by producing higher speeds on straight parts of the road (Comte & Jamson, 2000). In addition, the stronger braking pedal stroke recorded in the case of the SD group may be due to the fact that a later but more intense braking maneuver occurred on the curve in a simulator in comparison with the real-life driving performances. These findings are in agreement with those obtained in a recent study indicating that regardless of the presence of an oncoming intersection, drivers braked later but more intensively in a simulator than under field conditions in order to maintain the trajectory of the vehicle (Zöller, Abendroth, & Bruder, 2019). Drivers' misperception of the curvature radii in the simulator may also result in a stronger brake pedal stroke when they are negotiating a curve (Fildes & Triggs, 1984). In addition, it has been reported that the braking pressure increases with the workload (Horberry et al., 2006; Lansdown, Brook-carter, & Kersloot, 2004), which is in agreement with the results obtained with the NASA RTLX questionnaire in the present study. Lastly, the results obtained here show that a higher rpm also occurred during simulated driving than during circuit driving on the curved part of the circuit. As in the case of the straight line, this may be attributable to either the absence of vibration feedback (Tachiiri et al., 2017) or the drivers' misperception of the sound of the engine in the high fidelity driving simulator. In fact, some participants reported that they did not perceive the sound of the engine sufficiently clearly. This suggests that a scaling factor should therefore possibly be applied to the system of sound restitution adopted in the simulator, as done with the dynamic aspects (Chapron & Colinot, 2007).

Another important aim of this study was to describe the evolution of the discrepancies between real and simulated driving conditions with time, based on both objective and subjective parameters. The answers given by the CD group indicated that these participants experienced greater difficulty at the beginning of the drive than when driving their own vehicle, which decreased quite fast during the test (after driving for approximately 20 minutes); whereas this feeling of difficulty lasted longer in the SD group, as they obtained significantly lower scores in all the tests than in all the measurements performed before and after driving their own cars. This indicates that even after 40 minutes in the simulator, they did not feel driving to be as easy as with their own car and that the quality of their driving did not reach the level they reported with their own car. Several studies have focused on the self-assessment of the adaptation process while driving a simulator. Some discrepancies exist between the results obtained. Some authors reported that participants felt adapted to driving a simulator after approximately 11 minutes after the beginning of the drive. This finding is in agreement with the adaptation process measured objectively in terms of driving parameters such as the root mean square of the lane position, the root mean square of the longitudinal speed or the root mean square of the steering wheel deviations (Ronen & Yair, 2013); whereas other authors have mentioned that participants using a driving simulator tended to overestimate their driving performances in comparison with the stability of the objective driving parameters in the lateral or longitudinal control of the vehicle (Brandtner et al., 2019; Sahami & Sayed, 2010). A similar tendency has also been observed in drivers' self-declared assessments of their real driving quality, which they tended to overestimate due to a social desirability bias (Lajunen & Summala, 2003; Sundstro, 2008). **In view of the persistent devaluation gap in the SD group's driving assessment, this indicates that drivers are becoming aware of the discrepancy between driving in a real vehicle and driving in a simulator.**

No significant changes with time in the driving parameters were observed here during the 40 minutes driving tests. The differences initially observed between simulator and circuit driving persisted throughout the drive, in line with the answers obtained to the subjective questionnaires. Although no changes were observed in this study during the test, some adaptations may possibly have occurred before or after the time points on which the analysis was performed. Some authors have reported the occurrence of changes in the driving parameters when the participants were using a new real or partially virtual vehicle, below the 6 minutes threshold suggested in the present study (McGehee, Lee, Rizzo, Dawson, & Bateman, 2004). Beyond this threshold, the stability of the various driving parameters observed suggests that drivers did not change their way of driving, regardless of the duration of their initial period of adaptation, unless they possibly changed their approach after a much longer period of time.

This study has some limitations, which make it difficult to generalize the results obtained. **Both the CD group and the SD group were completely independent because of methodological constraints.** Different personality traits and driving styles may have been assigned to the one group or the other. Some studies have shown that the driving style may affect driving parameters such as the speed, the variability of the speed, and the force of the braking stroke (Amado, Arikan, Kaça, Koyuncu, & Turkan, 2014; Hooft Van Huysduynen et al., 2018). **In addition, in order to rule on external validity applicable to the real world, it would be necessary to extend the study to more road and traffic types.** Another aspect that should not be overlooked is that of simulator sickness, which prevents simulated driving from matching real driving conditions. Simulator sickness is known to affect drivers' performances



(Smyth et al., 2018). The gap observed here between drivers' subjective feelings about their ease and quality of driving and the scores they obtained on the NASA RTLX questionnaire may have been due to the fact that they experienced simulation sickness under simulated but not real driving conditions.

In conclusion, the present results show the existence of some discrepancies between the driving parameters of groups CD and SD, which depended on whether the road was straight or curved. These differences persisted throughout the whole driving session: the driving parameters remained stable from at least the 6th to the 40th minute in each of the two conditions. The lack of evolution observed confirms that the adaptation process either takes place during the first 6 minutes of driving, or takes much longer to develop. Another original feature of this study was the comparison made between the drivers' subjective feedback about how easy the test felt and how well they felt they had performed the test using a real car on either the circuit or a dynamic simulator in comparison with driving their own car. The results obtained here show that the gaps observed between participants' usual driving and simulator driving performances persisted during and even after the drive, whereas the gap between usual driving and real car driving on a circuit gradually decreased before disappearing after driving for 25 minutes. The results of this study confirm that driving simulators provide useful tools for studying driving behavior, providing the differences between real car driving and simulated driving performances are taken into account. In order to improve the validity of the results obtained on simulators and reduce these discrepancies, the sound and vibrations generated in simulators could be improved so as to provide drivers with more useful engine power and speed control feedback. Finding better ways of controlling simulators' dynamics in order to prevent simulator sickness would also be a major advantage. In order to further assess the ecological validity of driving simulators, future studies could focus on exploring drivers' physiological reactions during the drive. Further information of this kind in addition to the data obtained here on the driving parameters involved and drivers' subjective impressions would shed new light on drivers' processes of adaptation.

## 5. Bibliography

- Amado, S., Arikan, E., Kaça, G., Koyuncu, M., & Turkan, B. N. (2014). How accurately do drivers evaluate their own driving behavior? An on-road observational study. *Accident Analysis and Prevention*, 63, 65–73.
- Araujo, D., Davids, K., & Passos, P. (2007). Ecological Validity, Representative Design, and Correspondence Between Experimental Task Constraints and Behavioral Setting. *Ecological Psychology*, 19(1), 69–78.
- Baas, P. H., Charlton, S. G., & Bastin, G. T. (2000). Survey of New Zealand truck driver fatigue and fitness for duty. *Transportation Research Part F: Psychology and Behaviour*, 3(4), 185–193.
- Bellet, T., Bailly-Asuni, B., Mayenobe, P., & Banet, A. (2009). A theoretical and methodological framework for studying and modelling drivers' mental representations. *Safety Science*, 47(9), 1205–1221.
- Blaauw, G. J. (1982). Driving Experience and Task Demands in Simulator and Instrumented Car: a Validation Study. *Human Factors*, 24(4), 473–486.
- Bobermin, M. P., Silva, M. M., Ferreira, S., Guedes, J. C., & Santos Baptista, J. (2019).

- Driving simulator for evaluating the effects of road geometric design on driver behavior : protocol of a systematic review. *International Journal of Occupational and Environmental Safety*, 2, 46–51.
- Brandtner, A., Liebherr, M., Schweig, S., Maas, N., Schramm, D., & Brand, M. (2019). Subjectively estimated vs. objectively measured adaptation to driving simulators – Effects of age, driving experience, and previous simulator adaptation. *Transportation Research Part F: Psychology and Behaviour*, 64, 440–446.
- Brooks, J. O., Goodenough, R. R., Crisler, M. C., Klein, N. D., Alley, R. L., Koon, B. L., ... Wills, R. F. (2010). Simulator sickness during driving simulation studies. *Accident Analysis and Prevention*, 42(3), 788–796.
- Calder, B. J., Phillips, L. W., & Tybout, A. M. (1983). Beyond External Validity. *Journal of Consumer Research*, 10(1), 112–114.
- Chapron, T., & Colinot, J. (2007). The new PSA Peugeot-Citroën Advanced Driving Simulator Overall design and motion cue algorithm. *Driving Simulation Conference*, 42, 173.
- Comte, S. L., & Jamson, A. H. (2000). Traditional and innovative speed-reducing measures for curves: An investigation of driver behaviour using a driving simulator. *Safety Science*, 36(3), 137–150.
- Costa, N., Costa, S., Pereira, E., & Arezes, P. M. (2019). Workload Measures - Recent Trends in the Driving Context. *Occupational and Environmental Safety and Health*, 419–430.
- Eichelberger, A. H., & McCartt, A. T. (2016). Toyota drivers' experiences with Dynamic Radar Cruise Control, Pre-Collision System, and Lane-Keeping Assist. *Journal of Safety Research*, 56, 67–73.
- Fildes, B. N., & Triggs, T. J. (1984). The Effect of Road Curve Geometry on Curvature Matching Judgements. *Australian Road Research*, 12(5), 3.
- Foy, H. J., & Chapman, P. (2018). Mental workload is reflected in driver behaviour, physiology, eye movements and prefrontal cortex activation. *Applied Ergonomics*, 73, 90–99. <https://doi.org/10.1016/j.apergo.2018.06.006>
- Galy, E., & Mélan, C. (2015). Effects of Cognitive Appraisal and Mental Workload Factors on Performance in an Arithmetic Task. *Applied Psychophysiology Biofeedback*, 40(4), 313–325.
- Galy, E., Paxion, J., & Berthelon, C. (2017). Measuring mental workload with the NASA-TLX needs to examine each dimension rather than relying on the global score: An example with driving. *Ergonomics*, 61(4), 517–527.
- Godley, S. T., Triggs, T. J., & Fildes, B. N. (2002). Driving simulator validation for speed research. *Accident Analysis and Prevention*, 34(5), 589–600.
- Guajardo, M. G. P. (2016). *Examining context-based task characteristics: the effects of task characteristics on students' motivation and metacognitive experiences*. Universität Koblenz-Landau.
- Hart, S. G. (2006). Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(9), 904–908. Retrieved from <http://journals.sagepub.com/doi/10.1177/154193120605000909>

- Hooft Van Huysduynen, H., Terken, J., & Eggen, B. (2018). The relation between self-reported driving style and driving behaviour. A simulator study. *Transportation Research Part F: Psychology and Behaviour*, 56, 245–255.
- Horberry, T., Anderson, J., Regan, M. A., Triggs, T. J., & Brown, J. (2006). Driver distraction: The effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance. *Accident Analysis and Prevention*, 38(1), 185–191.
- Horrey, W. J., & Wickens, C. D. (2004). Driving and Side Task Performance : The Effects of Display Clutter, Separation, and Modality. *Human Factors*, 46(4), 611–624.
- Ishida, S., & Gayko, J. E. (2004). Development, evaluation and introduction of a lane keeping assistance system. *IEEE Intelligent Vehicles Symposium*, 943–944.
- Jacobé De Naurois, C., Bourdin, C., Bougard, C., & Vercher, J. (2018). Adapting artificial neural networks to a specific driver enhances detection and prediction of drowsiness. *Accident Analysis and Prevention*, 121(July), 118–128.
- Kazemzadehazad, S., Monajjem, S., Larue, G., & King, M. J. (2018). Driving simulator validation for speed research on curves of two-lane rural roads. *Proceedings of the Institution of Civil Engineers - Transport*, 1–6.
- Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator Sickness Questionnaire : An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology*, 3(3), 203–220.
- Klüver, M., Herrigel, C., Heinrich, C., Schöner, H. P., & Hecht, H. (2016). The behavioral validity of dual-task driving performance in fixed and moving base driving simulators. *Transportation Research Part F: Traffic Psychology and Behaviour*, 37, 78–96.
- Knapper, A., Christoph, M., Hagenzieker, M., & Brookhuis, K. (2015). Comparing a driving simulator to the real road regarding distracted driving speed. *European Journal of Transport and Infrastructure Research*, 15(2), 205–225.
- Lajunen, T., & Summala, H. (2003). Can we trust self-reports of driving? Effects of impression management on driver behaviour questionnaire responses. *Transportation Research Part F: Traffic Psychology and Behaviour*, 6(2), 97–107.
- Lansdown, T. C., Brook-carter, N., & Kersloot, T. (2004). Distraction from multiple in-vehicle secondary tasks : vehicle performance and mental workload implications. *Ergonomics*, 47(1), 91–104.
- Llopis-Castelló, D., Camacho-Torregrosa, F. J., Marín-Morales, J., Pérez-Zuriaga, A. M., García, A., & Dols, J. F. (2016). Validation of a low-cost driving simulator based on continuous speed profiles. *Transportation Research Record*, 2602(1), 104–114.
- McGehee, D. V., Lee, J. D., Rizzo, M., Dawson, J., & Bateman, K. (2004). Quantitative analysis of steering adaptation on a high performance fixed-base driving simulator. *Transportation Research Part F: Traffic Psychology and Behaviour*, 7(3), 181–196.
- Milleville-Pennel, I., & Charron, C. (2015). Driving for real or on a fixed-base simulator: is it so different? An explorative study. *Presence*, 24(1), 74–91.
- O'Neill, T. R., Krueger, G. P., Van Hemel, S. B., McGowan, A. L., Rogers, W. C., O'Neill, T. R., ... Rogers, W. C. (1999). Effects of cargo loading and unloading on truck driver alertness. *Transportation Research Record*, (1686), 42–48.

- Pakdamanian, E., Feng, L., & Kim, I. (2018). The effect of whole-body haptic feedback on driver's perception in negotiating a curve. *Proceedings of the Human Factors and Ergonomics Society*, 62(1), 19–23.
- Paredes, P. E., Ordoñez, F., Ju, W., & Landay, J. A. (2018). Fast & Furious : Detecting Stress with a Car Steering Wheel. *Conference on Human Factors in Computing Systems*, 1–12.
- Paxion, J., Galy, E., & Berthelon, C. (2014). Mental workload and driving. *Frontiers in Psychology*, 5(1344), 1–11.
- Ronen, A., & Yair, N. (2013). The adaptation period to a driving simulator. *Transportation Research Part F: Traffic Psychology and Behaviour*, 18, 94–106.
- Sahami, S., Jenkins, J., & Sayed, T. (2009). Methodology to Analyze Adaptation in Driving Simulators. *Transportation Research Record: Journal of the Transportation Research Board*, 2138(1), 94–101.
- Sahami, S., & Sayed, T. (2010). Insight into Steering Adaptation Patterns in a Driving Simulator. *Transportation Research Record: Journal of the Transportation Research Board*, 2185(1), 33–39.
- Santos, J., Merat, N., Mouta, S., Brookhuis, K., & de Waard, D. (2005). The interaction between driving and in-vehicle information systems: Comparison of results from laboratory, simulator and real-world studies. *Transportation Research Part F: Traffic Psychology and Behaviour*, 8(2), 135–146.
- Smyth, J., Jennings, P., Mouzakitis, A., & Birrell, S. (2018). Too Sick to Drive: How Motion Sickness Severity Impacts Human Performance. *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 1787–1793.
- Sundstro, A. (2008). Self-assessment of driving skill – A review from a measurement perspective. *Transportation Research Part F: Traffic Psychology and Behaviour*, 11(1), 1–9.
- Tachiiri, M., Tanaka, Y., & Sano, A. (2017). Vibrotactile Stimulation to Change Velocity Perception in Automobiles. *SICE Journal of Control, Measurement, and System Integration*, 10(3), 177–183.
- Tement, S., Plohl, N., Horvat, M., Musil, B., Jakus, G., & Sodnik, J. (2019). Is past behavior the best predictor of future behavior ? Comparing the behavioral validity of self-reported speeding and driving simulator data. *Age*, 30(10.16), 9–11.
- Törnros, J., & Bolling, A. (2006). Mobile phone use – effects of conversation on mental workload and driving speed in rural and urban environments. *Transportation Research Part F: Psychology and Behaviour*, 9, 298–306.
- Varotto, S. F., Farah, H., Bogenberger, K., van Arem, B., & Hoogendoorn, S. P. (2020). Adaptations in driver behaviour characteristics during control transitions from full-range Adaptive Cruise Control to manual driving: an on-road study. *Transportmetrica A: Transport Science*, 16(3), 776–806.
- Wege, C., Pereira, M.-S., Victor, T., & Krems, J. F. (2013). Behavioural adaptation in response to driving assistance technologies: A literature review. In *Driver Adaptation to Information and Assistance Systems* (pp. 13–44).
- Weir, D. H. (2010). Application of a driving simulator to the development of in-vehicle

human-machine-interfaces. *IATSS Research*, 34(1), 16–21.

Young, K., & Regan, M. (2007). Driver distraction : A review of the literature. *Distracted Driving*, 379–405.

Zheng, X. S., Tai, Y., & McConkie, G. W. (2005). Effects of Cognitive Tasks on Drivers' Eye Behavior and Performance. *Driving Assessment Conference*, 64–64.

Zöller, I., Abendroth, B., & Bruder, R. (2019). Driver behaviour validity in driving simulators – Analysis of the moment of initiation of braking at urban intersections. *Transportation Research Part F: Traffic Psychology and Behaviour*, 61, 120–130.

## 6. Acknowledgments

This research was supported by an agreement between Aix-Marseille University, CNRS and the Groupe PSA organized by the OpenLab “Automotive Motion Lab” and the Association Nationale de la Recherche Technologique n°2016/0502.

## 7. Complementary information

### 7.1. Perceived driving quality: the original version of the questions asked

Before and after the drive original French version: “Comment qualifieriez-vous la qualité de votre conduite **dans votre véhicule habituel**, sur une échelle allant de 1 très mauvaise à 10 excellente ?”; and its translation: “How would you rate your own standard of driving in your own car, on a scale ranging from 1 (very poor ) to 10 (excellent)?”.

During the drive original French version: “Comment qualifieriez-vous la qualité de votre conduite **en cet instant**, sur une échelle allant de 1 très mauvaise à 10 excellente ?”; and its translation: “How would you rate your own standard of driving at this very moment on a scale ranging from 1 ((very poor) to 10 (excellent)?”.

### 7.2. Perceived ease of driving: the original version of the questions asked

Before and after the drive original French version: “Comment vous sentez-vous **dans votre véhicule habituel**, sur une échelle allant de 1 très mal à l’aise à 10 parfaitement à l’aise ?”; and its translation: “How at ease do you feel when driving your own car, on a scale ranging from 1 (not at all at ease) to 10 (perfectly at ease)?”.

During the drive original French version: “Comment vous sentez-vous **en cet instant**, sur une échelle allant de 1 très mal à l’aise à 10 parfaitement à l’aise?”; and its translation: “How at ease do you feel at this very moment, on a scale ranging from 1 (not at all at ease) to 10 (perfectly at ease)?”.