



**HAL**  
open science

## Behavioral and workload measures in real and simulated driving: Do they tell us the same thing about the validity of driving simulation?

Régis Lobjois, Vérane Faure, Lara Désiré, Nicolas Benguigui

### ► To cite this version:

Régis Lobjois, Vérane Faure, Lara Désiré, Nicolas Benguigui. Behavioral and workload measures in real and simulated driving: Do they tell us the same thing about the validity of driving simulation?. Safety Science, 2021, 134, pp.105046. 10.1016/j.ssci.2020.105046 . hal-03007202

**HAL Id: hal-03007202**

**<https://hal.science/hal-03007202>**

Submitted on 24 Oct 2022

**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 4.0 International License

## **Behavioral and workload measures in real and simulated driving:**

### **Do they tell us the same thing about the validity of driving simulation?**

Régis Lobjois<sup>1</sup>, Vérane Faure<sup>1</sup>, Lara Désiré<sup>2</sup>, Nicolas Benguigui<sup>3</sup>

<sup>1</sup> COSYS-PICS-L, Univ Gustave Eiffel, IFSTTAR, F-77454 Marne-la-Vallée, France  
regis.lobjois@ifsttar.fr  
verane.faure06@gmail.com

<sup>2</sup> Centre for Studies and Expertise on Risks, Environment, Mobility, and Urban and Country Planning (Cerema), Project-team STI, 5 rue Jules Vallès, F-22015, Saint-Brieuc, France  
lara.desire@cerema.fr

<sup>3</sup> Normandie Univ, UNICAEN, CESAMS, 14000 Caen, France  
nicolas.benguigui@unicaen.fr

#### **Acknowledgments**

This research was supported by a grant from the French National Institute of Sciences and Technology for Transport, Development and Networks (IFSTTAR) to the second author when she was with the Laboratory for Road Operations, Perception, Simulators and Simulation. The authors would like to thank Jacky Robuant and Fabrice Vienne (IFSTTAR) for designing the simulated roads, and Dominique Bill and Renaldo Gritti (CEREMA) for their assistance in recruiting participants and data collection. The authors are also grateful to Steve Elcock for English-language editing.

#### **Disclosure statement**

The authors declare no conflict of interest.

Corresponding author:

Régis Lobjois

Laboratoire Perceptions, Interactions, Comportements & Simulations des usagers de la route et de la rue

Institut Français des Sciences et Technologies des Transports, de l'Aménagement et des Réseaux

Université Gustave Eiffel

14-20 Boulevard Newton, Cité Descartes

Champs sur Marne, F-77447 Marne la Vallée Cedex 2, France

Phone: +33 1 81 66 83 46.

76                   **Behavioral and workload measures in real and simulated driving:**  
77                   **Do they tell us the same thing about the validity of driving simulation?**

78   **1. Introduction**

79           For the sake of controllability, repeatability, and safety, driving simulators are  
80   extensively used for driver behavior research. However, they must have appropriate validity  
81   to be useful human factor research tools. Since Blaauw (1982), few studies have compared  
82   on-road and simulator driver behavior. While such comparative studies are not easy to carry  
83   out (Carsten & Jamson, 2011; Carsten et al., 2013), they remain the most appropriate means  
84   to evaluate and improve driving simulator validity. Most of them relied on behavioral  
85   measures to capture the differences between on-road and simulated driving (Blana, 1996;  
86   Mullen et al., 2011; Shechtman, 2010). More recently, few studies assessed the level of  
87   mental workload when driving in a real vehicle or in a simulator and showed contradictory  
88   results (e.g., Diels et al., 2011; Johnson et al., 2011; Milleville-Pennel & Charron, 2015). On  
89   another note, it has been found that mental workload measures were more sensitive to task  
90   demands than behavioral measures (Mehler et al., 2009, 2012; Yang et al., 2013). Therefore  
91   the present study was aimed at assessing behavioral and mental workload measures to better  
92   understand the differences between real and simulated driving.

93           Blaauw (1982) proposed to assess driving simulator validity based on physical and  
94   behavioral criteria. Physical validity corresponds to the extent to which software and  
95   hardware components of the simulator reproduce the physical reality of driving. Behavioral  
96   validity has been defined as the correspondence between driver behavior in a simulator and on  
97   the road. While some studies have paid attention to hardware components of the simulator to  
98   improve physical fidelity (e.g., Chatziastros et al., 1999; Jamson & Jamson, 2010; Kappé et  
99   al., 1999; Klüver et al., 2016), it is relatively well established that behavioral validity is a  
100   more important criterion than physical validity. This can be explained by the fact that the

101 strict duplication of reality is impossible to obtain in simulated driving (Pinto et al., 2008). In  
102 addition, a high level of physical fidelity is not always required to achieve good levels of  
103 behavioral validity (Blana, 1996; Klüver et al., 2016; Pinto et al., 2008). Blaauw (1982) also  
104 made a distinction between relative and absolute behavioral validity. Absolute validity is  
105 established with the sameness of numerical values in the behavioral measures in the two  
106 different environments. Relative validity is concluded when the differences due to the  
107 experimental manipulation are of the same order and direction and have a similar or identical  
108 magnitude in real and simulated driving conditions (for a discussion, see Klüver et al., 2016).

109       Longitudinal and lateral control have been mainly used to assess the behavioral validity  
110 of driving simulators, with speed being undoubtedly the most widely used (see  
111 Supplementary Table 1 for an overview of results obtained in studies that have compared real  
112 and simulated driving under free driving conditions. For a recent review of driving simulator  
113 validation studies, see Wynne et al., 2019). For lateral position, strong absolute differences  
114 between simulated and real driving have been found in the mean and standard deviation of  
115 lateral position. Relative validity of simulated driving is however relatively well preserved  
116 (see Supplementary Table 1). For speed, 11 out of the 20 studies which examined vehicle  
117 speed when driving freely concluded absolute validity while 7 concluded relative validity  
118 (two studies showed significant differences between the two systems). When relative validity  
119 of the experimental treatment ~~is was~~ found, driving speed ~~is was~~ higher in simulators than on  
120 roads (see Supplementary Table 1). Whether for lateral control or speed measures, these  
121 results greatly depended on the type of simulator (for a recent review, see Klüver et al., 2016)  
122 and driving context, such as the type of road, road geometry (Bella, 2008; Bittner et al., 2002;  
123 Branzi et al., 2017), presence of oncoming traffic (Alm, 1995) or traffic in the lane (Wang et  
124 al., 2010), and both posted and traffic speed (Wang et al., 2010).

125           While most of the studies evaluated behavioral validity, Blaauw (1982) initially  
126 proposed not only to assess behavioral measures but also mental workload measures. For  
127 Malaterre and Fréchaux (2001), this corresponds to the psychological validity of driving  
128 simulators defined as the sameness of the psychological processes underlying the driving  
129 activity, including mental workload. Focusing specifically on mental workload, which is the  
130 result of the interaction between task demands and the attentional resource capacity of the  
131 operator (Borghini et al., 2014; O'Donnell & Eggemeier, 1986), this distinction finds model-  
132 based and empirical evidence. De Waard (1996) offered a model which relates the operator's  
133 performance and workload to task demands. When the task demands are low, the operator is  
134 weakly engaged in the task and the level of mental workload is high while performance is  
135 low. When the task demands outmatch the operator's attentional resources, the latter is  
136 overloaded and performance is dramatically low. In-between these two extrema, two states  
137 are of interest here: i) when task demands are high enough to engage the operator whose  
138 attentional resources meet the task requirements, performance is high while mental workload  
139 level is low; ii) when the task demands increase, the operator has to engage more resources to  
140 maintain his performance and mental workload level is consequently high. This last state has  
141 been highlighted in studies showing that when task demands increased, physiological or  
142 visual attention measures of mental workload were sensitive to these changes while driving  
143 performance measures were not (e.g., Mehler et al., 2009, 2012; Yang et al., 2013).

144           Based on these considerations, when driving performance, namely vehicle handling, is  
145 maintained in a simulator compared to on-road driving, the mental workload level may then  
146 be either the same in both environments or different. If the mental workload level is the same,  
147 then the simulator does not place more demands on vehicle handling. The control of the  
148 vehicle, which belongs to the operational level of driving activity and requires little  
149 attentional resources, is not hampered when driving a simulator. If the mental workload level

150 is higher in the simulator, this suggests that more attentional resources have to be allocated to  
151 maintain driving performance. This increase in task demands could come from the lack of  
152 vestibular and proprioceptive cues (e.g., Bittner et al., 2002; Boer et al., 2000; Klee et al.,  
153 1999). Lastly, if the mental workload is higher when driving on a road, the difference could  
154 be due to a higher perception of risk and consequences of collisions that are not present in the  
155 simulator (Bella, 2008; Deniaud et al., 2015; Engström et al., 2005; Mueller, 2015).

156 Surprisingly, the assessment of mental workload when driving on a road and ~~on~~ in a  
157 simulator has given rise to a small pool of studies and to contradictory findings. Alm (1995)  
158 and Blaauw (1982) first showed that, while longitudinal control was well preserved whether  
159 the simulator was static or dynamic, driving in a simulator was quoted as being more difficult  
160 and more effort-demanding than driving for real. Similarly, Diels et al. (2011) found a very  
161 close correspondence in speed between the car and the simulator but subjective ratings of  
162 mental workload tended to be moderately higher in the driving simulator. Engström et al.  
163 (2005) compared vehicle-handling measures and physiological measures of mental workload  
164 (heart rate and skin conductance) in a static simulator, a dynamic simulator and an  
165 instrumented vehicle. Driving measures in the simulators were consistent with those obtained  
166 in real driving conditions. However, physiological measures of workload were higher in the  
167 field, suggesting an increased effort due to higher actual risk in real traffic. Other studies  
168 confirmed that physiological signals of mental workload were higher when driving on the  
169 road than driving in a simulator; however, no speed or lateral-control measures were reported  
170 (Johnson et al., 2011; Milleville-Pennel & Charron, 2015; Reimer & Mehler, 2011). At the  
171 very least, Milleville-Pennel and Charron (2015) reported that the self-rated driving  
172 performance was judged to be lower when driving in the simulator than in the field. Using a  
173 cognitive secondary task, Reimer and Mehler (2011) and Mehler et al. (2009) showed in two  
174 different studies sharing the same manipulation that when the secondary task difficulty

175 increased, the performance decreased by a greater amount in the simulator (Mehler et al.,  
176 2009) than in real driving conditions (Reimer & Mehler, 2011). This is in keeping with the  
177 idea that mental demand coming from simulated driving is higher. A similar result was found  
178 by Chen et al. (2015). In this study, participants had to perform a reaching task to virtual and  
179 real targets while verbally responding to an auditory tone. Results showed that the response  
180 time to the auditory stimulus was significantly longer for the virtual than for the real target  
181 condition, confirming that a VR-based task was more attention-demanding than its  
182 conventional real-world counterpart.

183 In sum, it appears that measures of behavioral validity may return only a partial view of  
184 the correspondence between driving for real and driving in a simulator. As suggested by  
185 model-based and empirical evidence, attentional resources required by the driving task during  
186 real and simulated driving might differ, while behavioral measures may not (De Waard, 1996;  
187 Mehler et al., 2009, 2012; Yang et al., 2013). This would lead to wrongfully inferring  
188 (behavioral) absolute validity while validity on other dimensions would not be established.

189 The objective of this study was therefore to assess driving simulation validity by  
190 confronting measures of behavioral (speed) and psychological (mental workload) validity. In  
191 this purpose, we selected a route where participants could drive freely and were likely to  
192 adopt a similar speed in real and simulated conditions. This made it possible to examine  
193 whether mental workload level differed between real and simulated driving while speed  
194 measures had a good correspondence in both driving conditions. It also appears that workload  
195 level in virtual environments has received little attention and results are controversial. One  
196 reason could be the use of a variety of measures in isolation, which are known to dissociate  
197 (e.g., Brookhuis & de Waard, 2002; Faure et al., 2016; Yeh & Wickens, 1988). We therefore  
198 used a multiple measurement approach of workload, with subjective, physiological (eye-blink  
199 rate) and secondary task performance measures.

## 200 **2. Method**

### 201 **2.1. Participants**

202 Twenty-four participants (8 females) took part in the experiment. They had a mean age  
203 of  $43.5 \pm 8.9$  years, ranging from 27 to 56 years old. All participants had their driver's license  
204 for a minimum of two years. The self-reported experience ranged from 9 to 36 years (mean =  
205  $23 \pm 9.6$  years) and the self-reported total mileage ranged from 20,000 to 420,000 km (mean =  
206  $300,000 \pm 196,000$  km). All participants had normal or corrected-to-normal vision, and were  
207 naïve as to the aims of the experiment. Before they took part in the study, all gave their  
208 informed consent which was approved by the local Ethics Committee. The ethical  
209 considerations and principles of the World Medical Association Declaration of Helsinki  
210 regarding experimentation were respected.

### 211 **2.2. Experimental set-up**

#### 212 **2.2.1. Real vehicle**

213 The field experiment was run using a Renault Clio III with manual transmission. It was  
214 equipped with a GPS sensor (GlobalSat Technology BR-355, Taiwan), a Controller Area  
215 Network (CAN) interface (PEAK System PCAN-USB, Germany), and three cameras (4Kam  
216 Wideboy Sport, United-Kingdom). GPS data were logged with 1 Hz, and vehicle-related data  
217 were logged between 10 and 50 Hz depending on the data. All data sources were  
218 synchronized on a computer (Toshiba Satellite Pro A300, Japan) running a data acquisition  
219 and synchronization software (Intempora RTMaps, France).

#### 220 **2.2.2. Fixed-base simulator**

221 The experiment in the virtual environment was run using a fixed-base driving simulator  
222 composed of a force-feedback steering wheel, foot pedals, gearbox (Logitech G25), and a  
223 Renault Clio driving seat (Figure 1). The simulated environment was displayed on three  
224 screens 0.53 m wide and 0.30 m high (resolution: 1,280 x 1,024 pixels; refresh rate: 30 Hz).

225 The distance between the participant's head and the center of each screen was 1 m. The visual  
226 angle subtended by this configuration was  $97^\circ$  horizontally and  $30^\circ$  vertically. The images  
227 were calculated and projected at the participant's eye height, with the simulated viewing angle  
228 aimed at the vanishing point of the simulated scenario. Engine sounds and environmental  
229 noises were generated by five speakers. The speedometer was displayed in the lower part of  
230 the central screen. The central mirror was superimposed on the visual environment in the  
231 upper part of the central screen. Data were collected at a sampling rate of 30 Hz.



232

233

Figure 1. Illustration of the fixed-base driving simulator.

### 234 2.2.3. Eye-tracker

235 Two identical Pertech head-mounted eye-trackers were used to record eye measures in  
236 the instrumented vehicle and the driving simulator, respectively. This monocular eye-tracker  
237 uses a pupil tracking technology with an image processing algorithm to define the ocular  
238 direction (50 Hz sampling rate) and has  $0.25^\circ$  of accuracy (according to the manufacturer  
239 specification). A seven-point calibration was performed at the beginning of each real and  
240 simulated experimental trial. Eye measures data were synchronized with simulated and real  
241 car data so that they could be analyzed relative to the road geometry and any driving  
242 measures.

### 243 2.2.4. Driving environment

244 The experimental route (Figure 2) was a 7.5 km long section of the D786 road near  
245 Saint-Brieuc (France) between Plérin (GPS coordinates: 48.549687, -2.796007) and Binic  
246 (GPS coordinates: 48.6014398, -2.8356539) and consisted of three different road sections.

247 The first road section was a circa 4.2 km long dual 2-lane carriageway road including a  
248 roundabout. The posted speed limit was 110 km/h and then 70 km/h before entering the  
249 roundabout (limited to 50 km/h). ~~except at the approach to the roundabout where it was set at~~  
250 ~~70 km/h~~. After the roundabout, the posted speed limit was set again at 110 km/h and then  
251 90 km/h at the end of the section (hereafter labeled highway environment). The roundabout  
252 was excluded from the analysis as the corresponding speed limit (50 km/h) was not in line  
253 with high-speed environment and free driving conditions. This initial decision was reinforced  
254 by the fact that in real driving conditions all participants had to stop due to the presence of  
255 other vehicles on the roundabout. The second road section was a circa 1.8 km undivided two-  
256 lane rural road. The posted speed limit was 90 km/h in the first section and 70 km/h in the last  
257 winding section (hereafter labeled rural environment). The third road section was a circa  
258 1.5 km urban single carriageway road passing through the town of Binic. The posted speed  
259 limit was 50 km/h (hereafter labeled urban environment).

260 With the objective of minimizing the impact of traffic uncertainties and congestion on  
261 the driving task, we selected based on traffic surveys carried out on several sections of the  
262 experimental road two periods of time (9:30 am to 11 am and 2:30 pm to 4 pm) during which  
263 traffic density was the lowest. Furthermore, in case of particular events during the driving  
264 sessions in the real environment that could not be reproduced in the simulated driving session  
265 (e.g., red traffic light, pedestrians, vehicle overtaking), the corresponding sequences were  
266 excluded from data analyses. These sequences represented less than 5% of the data.

267

268

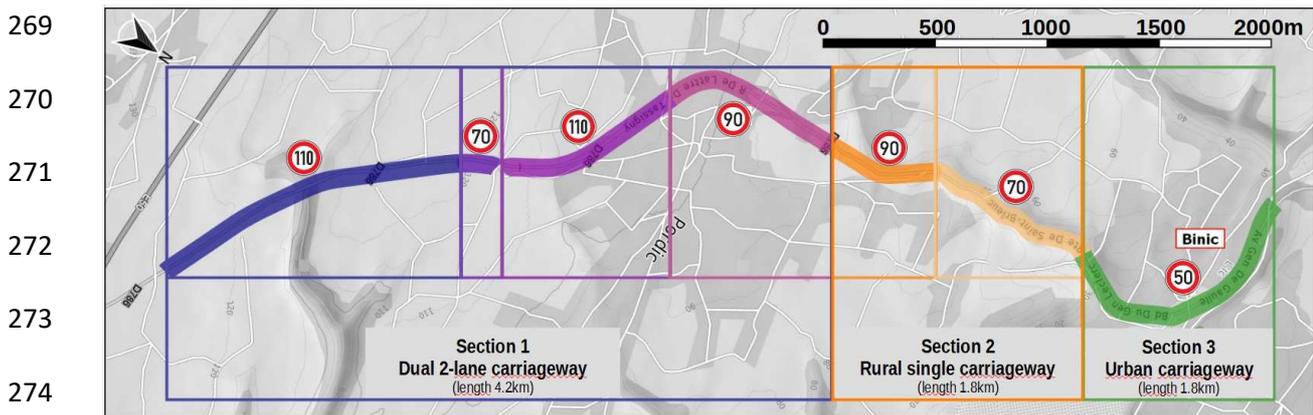


Figure 2. Overview of the experimental route and its different subsections.

All the important features of the real-world environment were replicated in a simulated 3D virtual world (see Figure 3). This included the three different driving environments (dual 2-lane carriageways and single carriageways in rural and urban contexts, respectively), roundabout, road signs, and buildings especially in the urban section. The road signs were positioned in the visual database at their corresponding locations in the real road. The traffic light in the actual traffic conditions was turned to green in the simulated scenario. In addition, traffic running in the opposite direction to that of the participant's car was introduced in the virtual environment. The traffic density corresponded to that surveyed during the two periods of time chosen for the on-road session. The simulated vehicles were of several types (passenger vehicles, commercial vehicles, trucks) and colors and obeyed the speed limit proper to the speed sections in the three driving environments.

Based on the classification of the effects of driving task demands on mental workload level (see Patten et al., 2004, 2006) and previous empirical results showing that drivers' mental workload is differently affected depending on the driving contexts (e.g., Faure et al., 2016; Mackenzie & Harris, 2017; Michaels et al., 2017; Patten et al., 2006), these three different driving environments were selected with the aim of triggering different levels of mental workload. The highway environment consisted of a straight carriageway located in the countryside with monotonous surrounding conditions, so that drivers would be undistracted



294  
295

296  
297

298

299 Figure 3. Illustration of the three different driving environments (top: highway environment in  
300 the first section at 110 km/h; middle: two-lane rural road in the section at 90 km/h; bottom:  
301 urban environment) in the simulator and in real situation. Views in the real situation were  
302 taken from Street View database (Google).

303

304 by the roadside environment and could control their vehicle in a consistent manner. This  
305 environment was therefore supposed to trigger a low level of mental workload. The rural  
306 environment represented a winding single carriageway lined with trees. In this scenario, an  
307 intermediate level of workload was expected as it required a greater degree of steering control  
308 than the highway environment. Lastly, the urban environment represented a visually rich  
309 neighborhood (street furniture, buildings, signs, traffic lights and intersections). Drivers had  
310 to adjust their trajectory and their speed and had to pay attention to their surrounding  
311 environment as they progressed through the environment. The urban section was therefore  
312 assumed to impose a high level of mental workload. The rationale for selecting these road  
313 sections was to examine whether the effects of the driving environment on drivers' mental  
314 workload varied in an absolute (i.e., sameness of numerical values between the two systems)  
315 or relative (i.e., same order and direction and similar or identical magnitude in the two  
316 systems) manner during real and simulated driving.

### 317 **2.3. Experimental design and task**

318 The experimental route was driven twice by all participants, one with the instrumented  
319 vehicle in real condition and one with the simulator in the 3D world. As the driving  
320 environment is known to affect drivers' mental workload level, we used a within-subject  
321 design with two driving conditions (real road and driving simulator) and three driving  
322 environments (highway environment, rural environment and urban environment). In the real  
323 driving condition, participants were asked to drive in the way that they were used to. In the  
324 driving simulator condition, they were asked to drive as if they were in a real car on a real  
325 road, while obeying the traffic regulations and speed limits.

326 While participants were performing the primary driving task, they also had to perform a  
327 tone reaction time task. This subsidiary secondary task consisted of responding verbally  
328 ("top") to an auditory stimulus (100 ms duration, 1.5 kHz amplitude) as quickly as possible.

329 To prevent any anticipation, the inter-stimulus interval varied between 7 and 10 s. It was  
330 adapted from the standard subsidiary secondary task (motor response to visual stimulus) as  
331 one of its drawbacks is the occurrence of manual and visual interference with the driving task  
332 (see Cantin et al., 2009; Jahn et al., 2005). Auditory stimuli were pre-recorded. In the real  
333 vehicle, sounds were played by an independent speaker. In the simulator, they were embedded  
334 into the engine sound. The participants' responses were recorded by the eye-tracker  
335 microphone. Compliant with the subsidiary task paradigm (Cain, 2007), participants were  
336 instructed to prioritize the primary driving task while responding to the auditory stimuli as  
337 rapidly as possible.

#### 338 **2.4. Procedure**

339 On their arrival, the nature and requirements of the experiment were explained to the  
340 participants. They signed an informed consent form and demographic data were collected.  
341 They drove the real and simulated condition in a counterbalanced order.

342 For the field test, participants sat in the instrumented vehicle and were invited to adjust  
343 the position of the seat, steering wheel and mirrors. They were then equipped with the eye-  
344 tracker before it was calibrated. Two experimenters also sat in the vehicle. The first one was  
345 seated at the front of the vehicle to record traffic events. The second was at the back to  
346 supervise data recording. When necessary due to the counterbalanced order of presentation of  
347 the two driving conditions, participants were trained with the audio-vocal reaction time task.  
348 Then, they drove the vehicle from the lab to a parking lot at the entrance to the course, located  
349 15 km away, which gave them time to get accustomed to the experimental vehicle. Once the  
350 quality of the ocular tracking had been checked, the experimenter seated in front reminded the  
351 participant of the instructions of the driving session. The participants then drove on the  
352 experimental course and performed the audio-vocal reaction time task concurrently. Once the  
353 total experimental route had been completed, they stopped in a parking lot and completed the

354 mental workload assessment questionnaire (NASA-TLX scale from Hart & Staveland, 1988)  
355 before removing the eye-tracker. Finally, one of the experimenters brought the vehicle back to  
356 its starting point.

357 For the simulator test, a similar procedure was used. Participants first made sure they  
358 were comfortable when sat in the simulator. One of the experimenters then fitted and  
359 calibrated the eye-tracker. When necessary (i.e., when they drove in the simulator condition  
360 first), participants familiarized themselves with the subsidiary secondary task. Then, they  
361 underwent a training period with the simulator which consisted of driving on a winding rural  
362 road with bends varying in length, curvature and direction (13 km long without any other  
363 traffic). Finally, a reminder of the instructions for the driving and the reaction time task was  
364 given. At the end of the driving session, participants completed the NASA-TLX questionnaire  
365 before removing the eye-tracker. The experiment lasted approximately 1 hr. 30.

## 366 **2.5. Dependent Variables**

### 367 **2.5.1. Speed**

368 The participant-determined speed was used to compute the mean speed (km/h) over the  
369 entire experimental route. As the speed limit varied within and between driving environments,  
370 the mean speed was also calculated for each speed section ( $n = 7$ ) for each environment. Since  
371 most comparative studies were run in a unique driving environment, this made it possible to  
372 examine the potential differences between real and simulated driving for different posted  
373 speed limits and sections.

### 374 **2.5.2. Blink rate**

375 Eye blinking has proven to be a useful and reliable signature of drivers' mental  
376 workload variations (e.g., Lohani et al., 2019), with a downward trend in the eye blink rate  
377 when the primary task difficulty increased (e.g., Faure et al., 2016; Hancock et al., 1990;  
378 Veltman & Gaillard, 1996; Wiberg et al., 2015). Capitalizing on previous research,

379 endogenous blinks corresponded to eye closures lasting between 80 (Benedetto et al., 2011,  
380 2014; Faure et al., 2016; McIntire et al., 2014) and 500 ms (Stern et al., 1984). Blink events  
381 were captured according to a modified version of the algorithm developed by Pedrotti et al.  
382 (2011) for low-speed eye-trackers, and the blink rate (number of blinks per minute) was  
383 computed in each driving environment and condition.

### 384 **2.5.3. Audio-vocal reaction time**

385 Reaction time to the secondary task was defined as the time lag between tone onset and  
386 voice onset. Before the mean response time (ms) was computed for each participant, outlier  
387 data were removed following a procedure similar to Cantin et al. (2009). Outliers  
388 corresponded to reaction times shorter than 150 ms or slower than each participant's mean  
389 reaction time plus two standard deviations. The mean response time was computed for each  
390 driving environment and condition.

### 391 **2.5.4. Subjective workload**

392 The NASA-TLX was used to assess the operator's perceived workload according to six  
393 dimensions: mental demand, physical demand, temporal demand, performance, effort, and  
394 frustration. The mean ratings on each dimension and the global score were computed for each  
395 driving condition (following the unweighted procedure).

## 396 **2.6. Statistical Analysis**

397 All statistical analyses were done with  $p$  set at .05. Speed differences between real and  
398 simulated driving were examined for the entire experimental route and each speed section  
399 ( $n = 7$ ). Speed results were first checked for distribution normality using Shapiro-Wilk tests  
400 for small samples. Pairwise comparisons were conducted using paired  $t$  tests when the data  
401 sets were normally distributed or Wilcoxon signed-rank tests when one of the distributions  
402 was not normally distributed. Effect sizes were reported as Cohen's  $d$ . Because of technical  
403 problems during real driving conditions, these analyses were performed with 21 participants.

404 The possible effects of the driving condition and the driving environment on blink rate  
405 and reaction time to the secondary task were assessed using 2 (driving conditions: real  
406 driving, simulated driving)  $\times$  3 (driving environment: highway, rural, urban) repeated  
407 measures analyses of variance (ANOVAs). Mauchly's test of sphericity was run to determine  
408 whether the sphericity assumption was violated. When necessary, the degrees of freedom  
409 were adjusted with the Greenhouse-Geisser correction. Tukey's honestly significant  
410 difference (HSD) procedure was run to follow up on significant effects. For each significant  
411 effect, partial eta-squared ( $\eta_p^2$ ) was computed to determine the proportion of total variability  
412 accounting for the effect. Note that the reaction time analysis was performed with 21  
413 participants because of technical problems during real driving condition.

414 Subjective workload ratings were analyzed following similar procedures as described  
415 above. Differences between real and simulated driving were first compared on global demand.  
416 To explore the origin of global demand differences, a repeated-measure ANOVA was run on  
417 mean ratings, with driving conditions and NASA-TLX subscales as within-participant factors.

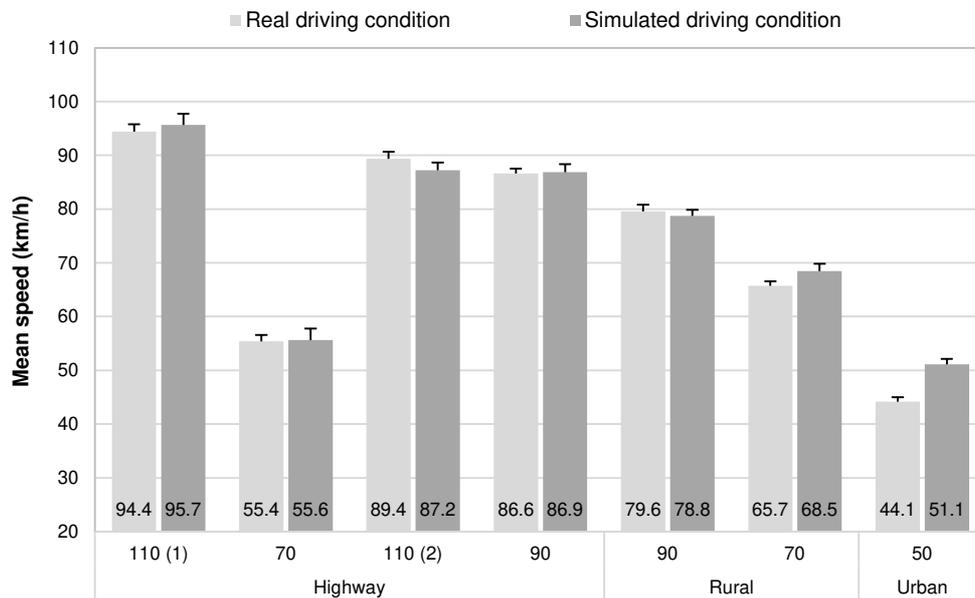
418

### 419 **3. Results**

#### 420 **3.1. Speed**

421 Statistical comparison of the mean driving speed computed for the entire route revealed  
422 that speed did not differ between real ( $73.6 \pm 3.4$  km/h) and simulated ( $74.8 \pm 5.4$  km/h)  
423 conditions ( $t(20) = -1.18, p = .25$ ). Analyses comparing speed between the two driving  
424 conditions in each speed section showed that speed did not differ in the highway environment  
425 (110 km/h:  $Z = 1.02, p = .30$ ; 70 km/h:  $Z = .08, p = .93$ ; 90 km/h:  $t(20) = 1.93, p = .07$ ;  
426 110 km/h:  $t(20) = -0.21, p = .83$ ) nor in the rural environment (90 km/h:  $t(20) = .81, p = .43$ ;  
427 70 km/h:  $Z = 1.69, p = .09$ ). However, in the urban environment, speed was significantly

428 higher in simulated than real condition ( $Z = 4.01, p < .001, d = -1.74$ ). Mean speed in each  
 429 driving condition, driving environment and speed section is provided in Figure 4.



430 Driving environment and speed section (km/h)

431 Figure 4. Mean speeds as a function of driving condition (real and simulated) for each driving  
 432 environment and speed section. Bars represent standard error of the mean.

433

434

### 435 3.2. Eye blink rate

436 The repeated measures ANOVA on eye blink rate revealed a significant main effect of

437 the driving condition ( $F(1,23) = 17.54, p < .001, \eta_p^2 = .43$ ), due to more frequent blinks

438 during real driving ( $11 \pm 9.7$  blinks/min) than simulated driving ( $4.1 \pm 4$  blinks/min), a

439 significant main effect of the driving environment ( $F(2,46) = 5.86, p < .01, \eta_p^2 = .20$ ), and a

440 significant interaction between the two factors ( $F(1.28,29.37) = 7.81, p < .01, \eta_p^2 = .25$ ).

441 Pairwise comparisons for the main effect of the driving environment showed that more blinks

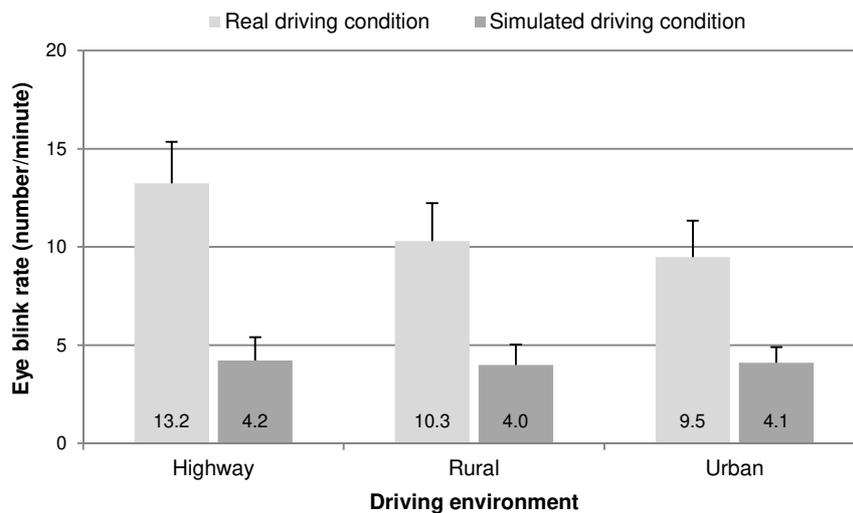
442 were made when driving on highway ( $M = 8.7 \pm 9.5$  blinks/min) than when driving in rural

443 ( $M = 7.1 \pm 8.2$  blinks/min) and urban settings ( $M = 6.8 \pm 7.5$  blinks/min), which did not

444 differ. The interaction between driving condition and driving environment (see Figure 5)

445 revealed the same effect of the driving environment as above in the real driving condition. In

446 the simulated driving condition, however, the blink rate did not vary significantly with the  
 447 driving environment. Note that the difference between real and simulated driving conditions  
 448 was significant in the three driving environments. It should also be mentioned that the eye  
 449 blink frequency decreased in the simulated as compared to the real driving condition in 21 out  
 450 of 24 participants.



451

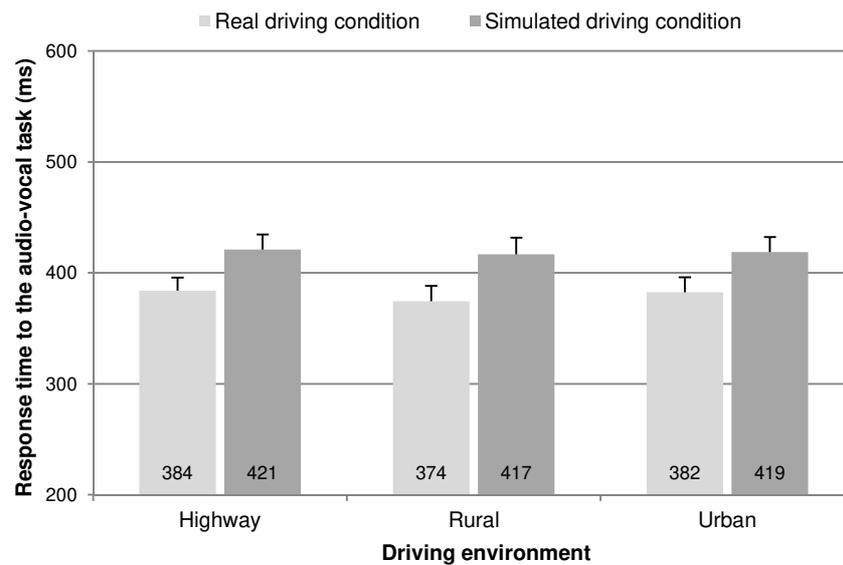
452 Figure 5. Interaction between driving condition (real and simulated) and driving environment  
 453 on eye blink rate. Bars represent standard error of the mean.

454

455

### 456 3.3. Response time to the audio-vocal task

457 The repeated measures ANOVA on the response time to the audio-vocal task, with  
 458 driving condition and driving environment as within-participant variables, yielded a  
 459 significant main effect of the driving condition ( $F(1,20) = 13.18, p < .01, \eta_p^2 = .40$ ). No other  
 460 main effect or interaction was observed. The response time was higher when driving the  
 461 simulator ( $419 \pm 65$  ms) than when driving for real ( $380 \pm 60$  ms). Note that the mean  
 462 response time increased in the simulated as compared to the real driving condition in 18 out of  
 463 21 participants. The mean response time as a function of driving condition and driving  
 464 environment is given for information in Figure 6.



465

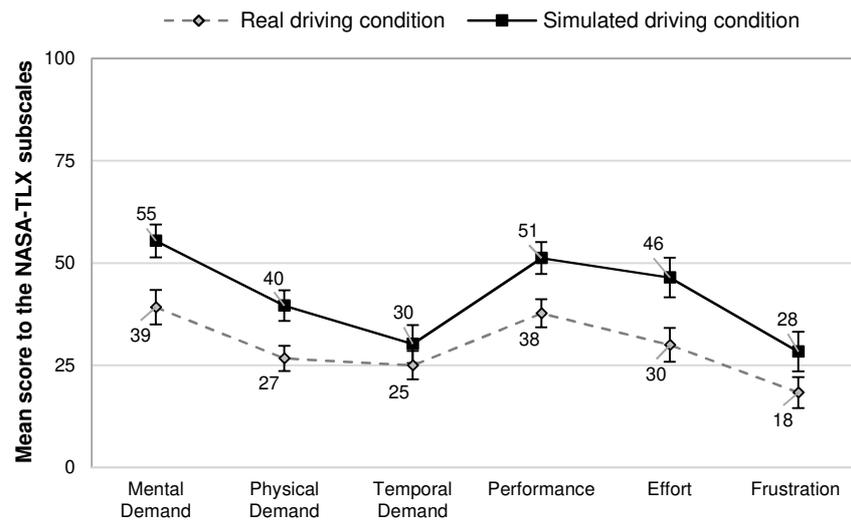
466 Figure 6. Mean response time (ms) to the audio-vocal task as a function of driving condition  
 467 (real and simulated) and driving environment. Bars represent standard error of the mean.

468

469

#### 470 3.4. Subjective workload (NASA-TLX)

471 A first *t*-test was run to compare the global score of subjective workload (unweighted  
 472 mean of all six subscales) between real and simulated driving. This analysis revealed a  
 473 statistically significant difference ( $t(23) = -3.75, p < .01, d = -.75$ ), with perceived workload  
 474 higher in simulated ( $M = 41.9$ ) than in real conditions ( $M = 29.5$ ). Global score in the  
 475 simulated condition outmatched the one in the real driving condition in 18 out of 24  
 476 participants (two participants had the same score in the two conditions). To examine the  
 477 workload profile and determine whether it changed with driving conditions, a 2 (driving  
 478 conditions: real, simulated)  $\times$  6 (subscales: mental demand, physical demand, temporal  
 479 demand, performance, effort, and frustration) repeated measures ANOVA was performed on  
 480 the ratings. Results revealed a significant main subscale effect ( $F(5,115) = 15.20, p < .001,$   
 481  $\eta_p^2 = .40$ ). Pairwise comparisons indicated that the effect was due to significantly higher  
 482 ratings of mental demand and performance (inverted scale: the higher the score, the worse the  
 483 performance) compared to physical demand, temporal demand and frustration, and to higher  
 484 ratings of effort compared to temporal demand and frustration (see Figure 7).



485

486 Figure 7. Workload profile (mean rating for each subscale) as a function of driving condition  
 487 (real and simulated). Bars represent standard error of the mean.

488

489

#### 490 4. Discussion

491

492

493

494

495

496

497

498

499

500

501

502

503

504

To date, few studies have tracked driving simulator validity using mental workload measures and they have found contradictory findings (e.g., Diels et al., 2011; Johnson et al., 2011; Milleville-Pennel & Charron, 2015). In this context, the present study aimed at determining whether mental workload varied between on-road and simulated driving while behavioral measures (speed) did not. To achieve this, we selected a free driving condition expected to induce a good correspondence in speed behavior. If the mental workload level differed according to the driving condition, the objective was to examine whether the effect of driving environment on mental workload was similar (relative validity) or not between simulated and real driving.

The findings confirmed previous studies (e.g., Ahlström et al., 2012), showing that the mean speed over the entire route had a very good correspondence between simulated (74.8 km/h) and on-road driving (73.6 km/h). However, mental workload level was higher in the simulator as revealed by scores to the NASA-TLX, response times to the subsidiary task and eye blink rate.

505           The similar speed but increased mental workload results in the simulator are consistent  
506 with the predictions we made from De Waard's model (1996). In real world driving,  
507 controlling the vehicle in free driving conditions is automated and requires few attentional  
508 resources. In driving simulation, on the other hand, vehicle handling is not automated. This  
509 could be explained by the low physical validity of the simulator (e.g., lack of inertial cues)  
510 which disrupts the naturalness of the perceptual-motor loop. This disruption may then call for  
511 an increase in the attentional resources allocated to vehicle handling to preserve driving  
512 performance. In this vein, Sahami and Sayed (2013) argued that when participants start to  
513 drive a simulator, they have to adapt their skills to the simulator outputs. Consequently, they  
514 need time to transfer their already existing driving-related perceptual-motor skills to operate  
515 the simulator appropriately (Ronen & Yair, 2013; Sahami & Sayed, 2010). However, this  
516 period imposes a mental effort, which lasts until driving becomes automatic (Sahami &  
517 Sayed, 2013). As an example, Benedetto et al. (2014) examined mental workload level during  
518 the first periods of exposition to a simulator and showed that mental workload level decreased  
519 with repetitions. In the present study, the non-significant effect of the driving environment on  
520 mental workload measures in the simulator suggests that mental workload level did not  
521 decrease with time on task; however, this may be confounded with the order of presentation  
522 of the environments, the urban one, which was encountered last, being known to increase task  
523 demands (Faure et al., 2016; Michaels et al., 2017). One way forward on this topic would be  
524 to examine together the adaptation period and mental workload to explore whether or not  
525 workload level decreases with the adaptation of the perceptual-motor loop to the simulator.  
526 This has significance as the allocation of additional mental resources to a level of driving  
527 activity known to be automated may interfere with the allocation of attention toward other  
528 driving tasks (e.g., Mackenzie & Harris, 2017; see also Wynne et al., 2019).

529 To broaden these results, the validity of driving simulators and VR systems for studying  
530 human behavior has been assessed through different user-centered dimensions (Malaterre &  
531 Fréchaux, 2001; Morice et al., 2008; Pinto et al., 2008). However, some studies have pointed  
532 out that these dimensions might be differently affected by the properties of a VR system.

533 Morice et al. (2008) argued that behavioral validity may be deteriorated by the system  
534 properties, while subjective validity (i.e., the extent to which the simulated experience is  
535 subjectively close to the real situation) remains good. Using a virtual ball-bouncing task, they  
536 found that bouncing performance deteriorated for small delays while participants reported a  
537 conscious perception of system latency for relatively high ones. In riding simulation, Lobjois  
538 et al. (2016) showed that visual roll used to elicit the leaning sensation when driving round a  
539 bend was preferred to no roll, although large visual roll hindered steering behavior. On the  
540 other hand, the combined evaluation of behavioral and psychological dimensions of validity  
541 also suggests a discrepancy between these two dimensions. As reported earlier, some studies  
542 showed that subjective workload ratings may be higher even when vehicle handling is well  
543 preserved in a simulator (Alm, 1995; Blaauw, 1982; Diels et al., 2011). Here, we showed that  
544 behavioral validity showed a good correspondence between on-road and simulated driving,  
545 but mental workload was higher in the simulator. These results thus support the idea that for  
546 in-depth evaluation of a VR system, user-centered effects are also to be examined at the level  
547 of psychological validity, i.e., at the level of the underlying processes.

548 Results also revealed that speed behavior was similar in 6 out of 7 speed sections. In the  
549 highway and rural environments, mean speed in each section did not vary between simulated  
550 and real driving. In the urban environment, however, participants drove faster in the simulator  
551 than in real driving conditions. In other words, speed was similar between both driving  
552 conditions in the less demanding environments, but it was higher when driving the simulator  
553 in the more demanding environment. This pattern of results is not consistent with previous

554 findings showing that participants drive faster in the simulator for the easiest driving  
555 situations (Bella, 2008; Bham et al., 2014; Bittner et al., 2002; Branzi et al., 2017). Nor is it  
556 consistent with other models which predict that reducing speed is a way to cope with task  
557 difficulty (e.g., Fuller, 2005). A potential perspective that could account for these results is a  
558 combined effect of increased workload and range of speed. Regarding the workload effect, it  
559 is well known that an increase in cognitive workload leads to visual tunneling and reduces  
560 glances to dashboard information and mirrors (Engström et al., 2005; Recarte & Nunes, 2000,  
561 2003; Reimer, 2009; Reimer et al., 2012; Tsai et al., 2007; Victor et al., 2005). It may then be  
562 that increased workload induced by driving a simulator leads to fewer fixations on the  
563 speedometer. Regarding the speed range effect, Wu et al. (2017) proposed that the perception  
564 of speed might be biased differently according to the range of speed. They showed that  
565 participants who passively observed video footage from a driver's perspective overestimated  
566 low speeds (range from 8 to 40 km/h) but underestimated high speeds (range from 64 to 96  
567 km/h). In a recent study by Hussain et al. (2019), participants had to actively drive a simulator  
568 and drive in the field and estimate when they reached the required speed (50, 70, 80, or 100  
569 km/h). Without having access to the speedometer, participants, in both driving conditions,  
570 underestimated their own speed and so drove faster than the 50 km/h requested speed;  
571 however, they drove slower than the 100 km/h requested speed. Interestingly, when they had  
572 access to the speedometer, participants drove slower than the 50 km/h requested speed in the  
573 field (48 km/h); in the simulator, however, they underestimated their own speed and drove  
574 faster than the requested speed (51 km/h). For higher speeds, participants overestimated their  
575 own speed in both conditions similarly. Taken together, these results suggest that the possible  
576 visual tunneling due to increased workload in the simulator may have led participants to rely  
577 on speed estimation to select their own speed. As low speeds are underestimated while

578 actively driving (Hussain et al., 2019), this could explain why participants drove the simulator  
579 faster than the real car in the urban environment.

580       Regarding mental workload and the issue of relative validity of driving simulators,  
581 different effects of the driving environment were found depending on the mental workload  
582 measures. First, the effect of the driving environment on blink frequency was significant in  
583 the field. As expected, the blink rate decreased when the complexity of the driving context  
584 increased. However, on the simulator, the driving environment was not found to have any  
585 effect on the blink frequency. This pattern of results does not support relative validity. With  
586 regard to subsidiary task performance, no difference was found deriving from the driving  
587 environment either in real or simulated driving. Unlike eye blink frequency, this result  
588 supports relative validity. Alternatively, the non-significant effect of the driving environment  
589 in real driving conditions may suggest that variations in task demands from one environment  
590 to another were not sufficient to elicit differences on this variable. This stems from the  
591 sensitivity of subsidiary tasks to mental workload variations. These tasks are sensitive to short  
592 peaks in mental workload (Jahn et al., 2005), unlike other measures that need to integrate raw  
593 data over longer intervals. Given free driving conditions with no particular scenario (such as  
594 pedestrians stepping out into the road), driving for real in the urban environment may have  
595 been carried out without compromising the reaction time task. Beyond this potential limit, the  
596 fact that the effects of driving environment on eye blink frequency and response time do not  
597 converge in both settings reflects previous studies that concluded that there was (e.g., Reimer  
598 & Mehler, 2011) or was not (e.g., Santos et al., 2005) relative validity depending on the  
599 mental workload measures. Finally, scores on the NASA-TLX questionnaire may contribute  
600 to the issue of relative validity. Galy et al. (2018) highlighted the value of distinguishing  
601 between the different subscales of the NASA-TLX rather than using only the global score. In  
602 addition, scores greater than the midpoint of a subscale (50) indicates substantial workload

603 levels and can be seen as “drivers” of task workload (Greenlee et al., 2018). The analysis  
604 revealed that the global score was mainly driven by scores on the mental demand,  
605 performance and effort subscales. While these scores were above or approached the subscale  
606 midpoint in the simulator, the workload profile was similar in the real and simulated  
607 conditions. Although subjective workload ratings were not collected for each environment,  
608 the similar workload profile in both settings is a reasonable case for overall relative validity of  
609 driving simulation.

## 610 **5. Limitations and conclusion**

611 It is important to note several limitations to the present study. While we concluded that  
612 a low-cost driving simulator has behavioral validity, lateral control measures were not  
613 collected. However, these measures are sensitive to the difference between real and simulated  
614 driving (see Supplementary Table 1; but see also the recent review by Wynne et al., 2019) and  
615 are also known to be correlated with mental workload level (e.g., Jamson & Merat, 2005;  
616 Faure et al., 2016). If behavioral validity is supported by the present speed behavior data, the  
617 same may not be true with lateral control measures. The question also arises whether a similar  
618 dissociation between behavioral and psychological validity would be observed with a high-  
619 fidelity driving simulator. Alm (1995) and Klüver et al. (2016) concluded absolute behavioral  
620 validity of high-fidelity driving simulators based on lateral control measures, as compared to  
621 fixed-base counterparts (but see Hallvig et al., 2013; Harms, 1996). This suggests better  
622 attunement of the perceptual-motor loop in high fidelity simulators and it cannot be excluded  
623 that mental workload level may be similar in high-fidelity driving simulators and on-road  
624 driving. Although Alm (1995) showed that subjective workload ratings were higher in a  
625 moving-base simulator than when driving in the field, this topic needs further research.

626 To conclude, the present study renews evidence that in free driving conditions, speed  
627 behavior in simulated driving closely matches that in real driving. Nevertheless, workload

628 level, as measured by blink frequency, response time to a subsidiary task and subjective  
629 ratings, was consistently higher in the simulator. These results thus suggest that while  
630 behavioral validity can be concluded for speed measures, no similar conclusion can be drawn  
631 when attentional resources allocated to vehicle handling come into consideration. The  
632 allocation of more attentional resources to a level of driving known to be automated may  
633 however interfere with other driving tasks. This fact should therefore be taken into account  
634 when driving simulators are used to study tactical or strategic aspects of the driving task  
635 which require attentional resources, or to study drivers interacting with in-vehicle information  
636 or assistance systems. **A potential follow-up of this study would be to explore whether**  
637 **workload level can be mitigated with the adaptation of the perceptual-motor loop to the**  
638 **simulator.**

639 **References**

- 640 Ahlström, C., Bolling, A., Sörensen, G., Eriksson, O., & Andersson, A. (2012). *Validating*  
641 *speed and road surface realism in VTI driving simulator III*. Swedish National Road and  
642 Transport Research Institute, Report N° 745A.
- 643 Aksan, N., Hacker, S. D., Sager, L., Dawson, J., Anderson, S., & Rizzo, M. (2016).  
644 Correspondence between simulator and on-road drive performance: Implications for  
645 assessment of driving safety. *Geriatrics*, 1(1): 8.  
646 <https://doi.org/10.3390/geriatrics1010008>
- 647 Alm, H., 1995. Driving simulator as research tools: A validation study based on the VTI  
648 Driving Simulator. Drive II V2065: GEM Validation Studies. Reference  
649 GEM/TR/TRC/MK950327, VTI, Linköping, Sweden.
- 650 Bella, F. (2005). Validation of a driving simulator for work zone design. *Transportation*  
651 *Research Record*, 1937, 136–144. <https://doi.org/10.3141/1937-19>
- 652 Bella, F. (2008). Driving simulator for speed research on two-lane rural roads. *Accident*  
653 *Analysis & Prevention*, 40, 1078–1087. <https://doi.org/10.1016/j.aap.2007.10.015>
- 654 Benedetto, S., Pedrotti, M., Minin, L., Baccino, T., Re, A., & Montanari, R. (2011). Driver  
655 workload and eye blink duration. *Transportation Research Part F: Traffic Psychology*  
656 *and Behaviour*, 14, 199–208. <https://doi.org/10.1016/j.trf.2010.12.001>
- 657 Benedetto, S., Lobjois, R., Faure, V., Dang, N-T., Pedrotti, M., & Caro, S. (2014). A  
658 comparison of immersive and interactive motorcycle simulator configurations.  
659 *Transportation Research Part F: Traffic Psychology and Behavior*, 23, 88-100.  
660 <http://dx.doi.org/10.1016/j.trf.2013.12.020>
- 661 Bham, G. H., Leu, M. C., Vallati, M., & Mathur, D. R. (2014). Driving simulator validation  
662 of driver behavior with limited safe vantage points for data collection in work zones.  
663 *Journal of Safety Research*, 49, 53-60. <https://doi.org/10.1016/j.jsr.2014.02.012>
- 664 Bittner Jr., A. C., Simsek, O., Levison, W. H., & Campbell, J. L. (2002). On-road versus  
665 simulator data in driver model development. *Transportation Research Record*, 1803, 38–  
666 44. <https://doi.org/10.3141/1803-06>
- 667 Blaauw, G. J. (1982). Driving experience and task demands in simulator and instrumented  
668 car: A validation study. *Human Factors*, 24, 473–486.  
669 <https://doi.org/10.1177/001872088202400408>
- 670 Blana, E. (1996). Driving Simulator Validation Studies: A Literature Review. Institute of  
671 Transport Studies, University of Leeds, Working Paper 480. Retrieved from:  
672 <http://eprints.whiterose.ac.uk/2110/>
- 673 Blana, E., & Golias, J. (2002). Differences between vehicle lateral displacement on the road  
674 and in a fixed-base simulator. *Human Factors*, 44, 303–313.  
675 <https://doi.org/10.1518/0018720024497899>

- 676 Boer, E. R., Girshik, A. R., Yamamura, T., & Kuge, N. (2000). Experiencing the same road  
677 twice: A driver centered comparison between simulation and reality. In *Proceedings of*  
678 *the Driving Simulation Conference* (pp. 33–55). Paris, France: Driving Simulation  
679 Association.
- 680 Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014). Measuring  
681 neurophysiological signals in aircraft pilots and car drivers for the assessment of mental  
682 workload, fatigue and drowsiness. *Neuroscience & Biobehavioral Reviews*, *44*, 58–75.  
683 <https://doi.org/10.1016/j.neubiorev.2012.10.003>
- 684 Branzi, V., Domenichini, L., & La Torre, F. (2017). Drivers' speed behaviour in real and  
685 simulated urban roads – A validation study. *Transportation Research Part F: Traffic*  
686 *Psychology and Behaviour*, *49*, 1–17. <https://doi.org/10.1016/j.trf.2017.06.001>
- 687 Brookhuis, K. A., & de Waard, D. (2002). On the assessment of mental workload and other  
688 subjective qualifications. *Ergonomics*, *45*, 1026-1030.  
689 <https://doi.org/10.1080/00140130210166799>
- 690 Cain, B. (2007). A review of the mental workload literature. Defense Research and  
691 Development Toronto (Canada) (pp. 1–35).
- 692 Cantin, V., Lavallière, M., Simoneau, M., & Teasdale, N. (2009). Mental workload when  
693 driving in a simulator: Effects of age and driving complexity. *Accident Analysis &*  
694 *Prevention*, *41*(4), 763-771. <https://doi.org/10.1016/j.aap.2009.03.019>
- 695 Carsten, O., & Jamson, A. H. (2011). Driving simulators as research tools in traffic  
696 psychology. In B. E. Porter (Ed.), *Handbook of Traffic Psychology* (pp. 87–96).  
697 Academic Press, London.
- 698 Carsten, O., Kircher, K., & Jamson, S. (2013). Vehicle-based studies of driving in the real  
699 world: The hard truth? *Accident Analysis & Prevention*, *58*, 162-174.  
700 <https://doi.org/10.1016/j.aap.2013.06.006>
- 701 Chatziastros, A., Wallis, G. M., & Bülthoff, H. H. (1999). The effect of field of view and  
702 surface texture on driver steering performance. In A. G. Gale (Ed.), *Vision in Vehicles*  
703 *VII* (pp. 253-259). Elsevier, Amsterdam.
- 704 Chen, Y. A., Chung, Y. C., Proffitt, R., Wade, E., & Winstein, C. (2015). Attentional demand  
705 of a virtual reality-based reaching task in nondisabled older adults. *Journal of Motor*  
706 *Learning and Development*, *3*, 91-109. <https://doi.org/10.1123/jmld.2014-0067>
- 707 Deniaud, C., Honnet, V., Jeanne, B., & Mestre, D. (2015). The concept of “presence” as a  
708 measure of ecological validity in driving simulators. *Journal of Interaction Science*, *3*(1),  
709 1-13. <https://doi.org/10.1186/s40166-015-0005-z>
- 710 de Waard, D. (1996). *The Measurement of Drivers' Mental Workload*. Traffic Research Center,  
711 Groningen University, Netherlands.

- 712 Diels, C., Robbins, R., & Reed, N. (2011). Behavioural validation of the TRL driving  
713 simulator DigiCar: phase 1 – Speed choice. In L. Dorn (Ed.), *Driver Behaviour and*  
714 *Training*, vol. 5 (pp. 429-446). Farnham: Ashgate.
- 715 Engström, J., Johansson, E., & Östlund, J. (2005). Effects of visual and cognitive load in real  
716 and simulated motorway driving. *Transportation Research Part F: Traffic Psychology*  
717 *and Behaviour*, 8, 97–120. <https://doi.org/10.1016/j.trf.2005.04.012>
- 718 Faure, V., Lobjois, R., & Benguigui, N. (2016). The effects of driving environment  
719 complexity and dual tasking on drivers' mental workload and eye blink behavior.  
720 *Transportation Research Part F: Traffic Psychology and Behaviour*, 40, 78–90.  
721 <http://dx.doi.org/10.1016/j.trf.2016.04.007>
- 722 Fuller, R. (2005). Towards a general theory of driver behaviour. *Accident Analysis & Prevention*,  
723 37(3), 461-472. <https://doi.org/10.1016/j.aap.2004.11.003>
- 724 Galy, E., Paxon, J., & Berthelon, C. (2018). Measuring mental workload with the NASA-  
725 TLX needs to examine each dimension rather than relying on the global score: An  
726 example with driving. *Ergonomics*, 61(4), 517-527.  
727 <https://doi.org/10.1080/00140139.2017.1369583>
- 728 Gemou, M. (2013). Transferability of driver speed and lateral deviation measurable  
729 performance from semi-dynamic driving simulator to real traffic conditions. *European*  
730 *Transport Research Review*, 5, 217–233.
- 731 Godley, S. T., Triggs, T. J., & Fildes, B. N. (2002). Driving simulator validation for speed  
732 research. *Accident Analysis & Prevention*, 34, 589–600. [https://doi.org/10.1016/S0001-](https://doi.org/10.1016/S0001-4575(01)00056-2)  
733 [4575\(01\)00056-2](https://doi.org/10.1016/S0001-4575(01)00056-2)
- 734 Greenlee, E. T., DeLucia, P. R., & Newton, D. C. (2018). Driver vigilance in automated  
735 vehicles: Hazard detection failures are a matter of time. *Human Factors*, 60(4), 465–476.  
736 <https://doi.org/10.1016/j.human.2018.01.011>
- 737 Hallvig, D., Anund, A., Fors, C., Kecklund, G., Karlsson, J. G., Wahde, M., & Åkerstedt, T.  
738 (2013). Sleepy driving on the real road and in the simulator — A comparison. *Accident*  
739 *Analysis & Prevention*, 50, 44-50. <https://doi.org/10.1016/j.aap.2012.09.033>
- 740 Hancock, P. A., Wulf, G., Thom, D., & Fassnacht, P. (1990). Driver workload during  
741 differing driving maneuvers. *Accident Analysis & Prevention*, 22, 281–290.
- 742 Harms, L. (1996). Driving performance on a real road and in a driving simulator: Results of a  
743 validation study. In Gale, A. G., Brown, I. D., Haslegrave, C. M., Taylor, S. P. (Eds.),  
744 *Vision in Vehicles – V*. Elsevier, Amsterdam, pp. 19–26.
- 745 Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index):  
746 Results of empirical and theoretical research. *Advances in Psychology*, 52, 139–183.  
747 [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
- 748 Hussain, Q., Alhajyaseen, W. K. M., Pirdavani, A., Reinolsmann, N., Brijs, K., & Brijs, T.  
749 (2019). Speed perception and actual speed in a driving simulator and real-world: A

- 750 validation study. *Transportation Research Part F: Traffic Psychology and Behaviour*,  
751 62, 637-650. <https://doi.org/10.1016/j.trf.2019.02.019>
- 752 Jahn, G., Oehme, A., Krems, J. F., & Gelau, C. (2005). Peripheral detection as a workload  
753 measure in driving: Effects of traffic complexity and route guidance system use in a  
754 driving study. *Transportation Research Part F: Traffic Psychology and Behaviour*, 8(3),  
755 255-275. <http://dx.doi.org/10.1016/j.trf.2005.04.009>
- 756 Jamson, S. L., & Jamson, A. H. (2010). The validity of a low-cost simulator for the  
757 assessment of the effects of in-vehicle information systems. *Safety Science*, 48, 1477–  
758 1483. <https://doi.org/10.1016/j.ssci.2010.07.008>
- 759 Jamson, A. H., & Merat, N. (2005). Surrogate in-vehicle information systems and driver  
760 behaviour: Effects of visual and cognitive load in simulated rural driving. *Transportation  
761 Research Part F: Traffic Psychology and Behaviour*, 8, 79–96.  
762 <https://doi.org/10.1016/j.trf.2005.04.002>
- 763 Johnson, M. J., Chahal, T., Stinchcombe, A., Mullen, N., Weaver, B., & Bédard, M. (2011).  
764 Physiological responses to simulated and on-road driving. *International Journal of  
765 Psychophysiology*, 81, 203-208. <https://doi.org/10.1016/j.ijpsycho.2011.06.012>
- 766 Kappé, B., van Erp, J., & Korteling, J. H. (1999). Effects of head-slaved and peripheral  
767 displays on lane-keeping performance and spatial orientation. *Human Factors*, 41, 453–  
768 466. <https://doi.org/10.1518/001872099779610950>
- 769 Klee, H., Bauer, C., Radwan, E., & Al-Deek, H. (1999). Preliminary validation of driving  
770 simulator based on forward speed. *Transportation Research Record*, 1689, 33-39.  
771 <https://doi.org/10.3141/1689-05>
- 772 Klüver, M., Herrigel, C., Heinrich, C., Schöner, H.-P., & Hecht, H. (2016). The behavioral  
773 validity of dual-task driving performance in fixed and moving base driving simulators.  
774 *Transportation Research Part F: Traffic Psychology and Behaviour*, 37, 78–96.  
775 <http://dx.doi.org/10.1016/j.trf.2015.12.005>
- 776 Knapper, A., Christoph, M., Hagenzieker, M., & Brookhuis, K. (2015). Comparing a driving  
777 simulator to the real road regarding distracted driving speed. *European Journal of  
778 Transport and Infrastructure Research*, 15, 205-225.
- 779 Lobjois, R., Siegler, I. A., & Mars, F. (2016). Effects of visual roll on steering control and  
780 gaze behavior in a motorcycle simulator. *Transportation Research Part F: Traffic  
781 Psychology and Behaviour*, 38, 55-66. <http://dx.doi.org/10.1016/j.trf.2016.01.002>
- 782 Lohani, M., Payne, B. R., & Strayer, D. L. (2019). A review of psychophysiological measures  
783 to assess cognitive states in real-world driving. *Frontiers in Human Neuroscience*, 13:57.  
784 <https://doi.org/10.3389/fnhum.2019.00057>
- 785 Mackenzie, A. K., & Harris, J. M. (2017). A link between attentional function, effective eye  
786 movements, and driving ability. *Journal of Experimental Psychology: Human  
787 Perception and Performance*, 43(2), 381-394. <http://dx.doi.org/10.1037/xhp0000297>

- 788 Malaterre, G., & Fréchaux, J. (2001). Étude de la validité des simulateurs de conduite par  
789 comparaison de tâches réalisées en situation réelle et en simulation. In V. Cavallo & C.  
790 Berthelon (Eds.), *Facteurs perceptifs dans les activités de transport* (pp.149-156).  
791 Arcueil, France: INRETS.
- 792 McIntire, L. K., McKinley, R. A., Goodyear, C., & McIntire, J. P. (2014). Detection of  
793 vigilance performance using eye blinks. *Applied Ergonomics*, *45*, 354–362.  
794 <https://doi.org/10.1016/j.apergo.2013.04.020>
- 795 Mehler, B., Reimer, B., Coughlin, J. F., & Dusek J. A. (2009). Impact of incremental  
796 increases in cognitive workload on physiological arousal and performance in young adult  
797 drivers. *Transportation Research Record*, *2138*, 6-12. <https://doi.org/10.3141/2138-02>
- 798 Mehler, B., Reimer, B., & Coughlin, J.F. (2012). Sensitivity of physiological measures for  
799 detecting systematic variations in cognitive demand from a working memory task: An  
800 on-road study across three age groups. *Human Factors*, *54*, 396-412.  
801 <https://doi.org/10.1177/0018720812442086>
- 802 Michaels, J., Chaumillon, R., Nguyen-Tri, D., Watanabe, D., Hirsch, P., Bellavance, F., et al.  
803 (2017). Driving simulator scenarios and measures to faithfully evaluate risky driving  
804 behavior: A comparative study of different driver age groups. *PLoS ONE* *12*(10):  
805 e0185909. <https://doi.org/10.1371/journal.pone.0185909>
- 806 Milleville-Pennel, I., & Charron, C. (2015). Driving for real or on a fixed-base simulator: Is it  
807 so different? An explorative study. *Presence: Teleoperators and Virtual Environments*,  
808 *24*(1), 74-91. [https://doi.org/10.1162/PRES\\_a\\_00216](https://doi.org/10.1162/PRES_a_00216)
- 809 Morice, A. H. P., Siegler, I. A., & Bardy, B. G. (2008). Action-perception patterns in virtual  
810 ball-bouncing: Combating system latency and tracking functional validity. *Journal of*  
811 *Neuroscience Methods*, *169*, 255–266. <https://doi.org/10.1016/j.jneumeth.2007.11.020>
- 812 Mueller, J. (2015). *Driving in a simulator versus on-road: The effect of increased mental*  
813 *effort while driving on real roads and a driving simulator*. Montana State University.
- 814 Mullen, N., Charlton, J. L., Devlin, A., & Bédard, M. (2011). Simulator validity: Behaviors  
815 observed on the simulator and on the road. In D. L. Fisher, M. Rizzo, J. K. Caird, & J. D.  
816 Lee (Eds.), *Handbook of Driving Simulation for Engineering, Medicine, and Psychology*  
817 (pp. 13.1-13.18). CRC Press Inc.
- 818 O'Donnell, R., & Eggemeier, T. (1986). Workload assessment methodology. In K. R. Boff, L.  
819 Kaufman, & J. P. Thomas (Eds.). *Handbook of perception and human performance* (Vol.  
820 2, pp. 1–49). New York: Wiley Interscience.
- 821 Patten, C. J. D., Kircher, A., Östlund, J., & Nilsson, L. (2004). Using mobile telephones:  
822 Cognitive workload and attention resource allocation. *Accident Analysis & Prevention*,  
823 *36*, 341–350. [https://doi.org/10.1016/S0001-4575\(03\)00014-9](https://doi.org/10.1016/S0001-4575(03)00014-9)
- 824 Patten, C. J. D., Kircher, A., Östlund, J., Nilsson, L., & Svenson, O. (2006). Driver experience  
825 and cognitive workload in different traffic environments. *Accident Analysis &*  
826 *Prevention*, *38*, 887–894. <https://doi.org/10.1016/j.aap.2006.02.014>

- 827 Pedrotti, M., Lei, S., Dzaack, J., & Rötting, M. (2011). A data-driven algorithm for offline  
828 pupil signal preprocessing and eyeblink detection in low-speed eye-tracking protocols.  
829 *Behavior Research Methods*, *43*, 372–383. <https://doi.org/10.3758/s13428-010-0055-7>
- 830 Pinto, M., Cavallo, V., & Ohlmann, T. (2008). The development of driving simulators:  
831 Toward a multisensory solution. *Le Travail Humain*, *71*, 62–95.  
832 <https://doi.org/10.3917/th.711.0062>
- 833 Recarte, M. A., & Nunes, L. M. (2000). Effects of verbal and spatial-imagery tasks on eye  
834 fixations while driving. *Journal of Experimental Psychology: Applied*, *6*(1), 31–43.  
835 <http://dx.doi.org/10.1037/1076-898X.6.1.31>
- 836 Reed, M. P., & Green, P. A. (1999). Comparison of driving performance on-road and in a low-  
837 cost simulator using a concurrent telephone dialing task. *Ergonomics*, *42*, 1015-1037.  
838 <https://doi.org/10.1080/001401399185117>
- 839 Reimer, B. (2009). Impact of cognitive task complexity on drivers' visual tunneling.  
840 *Transportation Research Record*, *2138*, 13–19. <https://doi.org/10.3141/2138-03>
- 841 Reimer, B., & Mehler, B. (2011). The impact of cognitive workload on physiological arousal  
842 in young adult drivers: A field study and simulation validation. *Ergonomics*, *54*, 932-  
843 942. <http://dx.doi.org/10.1080/00140139.2011.604431>
- 844 Reimer, B., Mehler, B., Wang, Y., & Coughlin, J. F. (2012). A field study on the impact of  
845 variations in short-term memory demands on drivers' visual attention and driving  
846 performance across three age groups. *Human Factors*, *54*(3), 454–468.  
847 <https://doi.org/10.1177/0018720812437274>
- 848 Ronen, A., & Yair, N. (2013). The adaptation period to a driving simulator. *Transportation  
849 Research Part F: Traffic Psychology and Behaviour*, *18*, 94–106.  
850 <https://doi.org/10.1016/j.trf.2012.12.007>
- 851 Sahami, S., & Sayed, T. (2010). Insight into steering adaptation patterns in a driving  
852 simulator. *Transportation Research Record*, *2185*(1), 33–39.  
853 <https://doi.org/10.3141/2185-05>
- 854 Sahami, S., & Sayed, T. (2013). How drivers adapt to drive in driving simulator, and what is  
855 the impact of practice scenario on the research? *Transportation Research Part F: Traffic  
856 Psychology and Behaviour*, *16*, 41–52. <https://doi.org/10.1016/j.trf.2012.08.003>
- 857 Santos, J., Merat, N., Mouta, S., Brookhuis, K., & De Waard, D. (2005). The interaction  
858 between driving and in-vehicle information systems: Comparison of results from  
859 laboratory, simulator and real-world studies. *Transportation Research Part F: Traffic  
860 Psychology and Behaviour*, *8*, 135–146. <https://doi.org/10.1016/j.trf.2005.04.001>
- 861 Shechtman, O. (2010). Validation of driving simulators. *Advances in Transportation Studies*,  
862 Special Issue 2010, 53-62.
- 863 Stern, J. A., Walrath, L. C., & Goldstein, R. (1984). The endogenous eyeblink.  
864 *Psychophysiology*, *21*, 22–33. <https://doi.org/10.1111/j.1469-8986.1984.tb02312.x>

- 865 Törnros, J. (1998). Driving behaviour in a real and a simulated road tunnel — a validation  
866 study. *Accident Analysis & Prevention*, 30(4), 497-503. <https://doi.org/10.1016/S0001->  
867 4575(97)00099-7
- 868 Tsai, Y.-F., Viirre, E., Strychacz, C., Chase, B., & Jung, T.-P. (2007). Task performance and  
869 eye activity: Predicting behavior relating to cognitive workload. *Aviation, Space, and*  
870 *Environmental Medicine*, 78(5), B176–B185.
- 871 Veltman, J. A., & Gaillard, A. W. K. (1996). Physiological indices of workload in a simulated  
872 flight task. *Biological Psychology*, 42, 323–342. <https://doi.org/10.1016/0301->  
873 0511(95)05165-1
- 874 Victor, T. W., Harbluk, J. L., & Engström, J. A. (2005). Sensitivity of eye-movement  
875 measures to in-vehicle task difficulty. *Transportation Research Part F: Traffic*  
876 *Psychology and Behaviour*, 8(2), 167–190. <https://doi.org/10.1016/j.trf.2005.04.014>
- 877 Wade, M. G., & Hammond, C. (1998). Simulator validation: Evaluating driver performance in  
878 simulation and the real world. Report N°28, Minneapolis, MN: Minnesota Department of  
879 Transportation.
- 880 Wang, Y., Mehler, B., Reimer, B., Lammers, V., D'Ambrosio, L. A., & Coughlin, J. F. (2010).  
881 The validity of driving simulation for assessing differences between in-vehicle  
882 informational interfaces: A comparison with field testing. *Ergonomics*, 53, 404-420.  
883 <http://dx.doi.org/10.1080/00140130903464358>
- 884 Wiberg, H., Nilsson, E., Lindén, P., Svanberg, B., & Poom, L. (2015). Physiological  
885 responses related to moderate mental load during car driving in field conditions.  
886 *Biological Psychology*, 108, 115-125. <http://dx.doi.org/10.1016/j.biopsycho.2015.03.017>
- 887 Wu, C., Yu, D., Doherty, A., Zhang, T., Kust, L., & Luo, G. (2017). An investigation of  
888 perceived vehicle speed from a driver's perspective. *PLoS ONE*, 12(10): e0185347.  
889 <https://doi.org/10.1371/journal.pone.0185347>
- 890 Wynne, R. A., Beanland, V., & Salmon, P. M. (2019). Systematic review of driving simulator  
891 validation studies. *Safety Science*, 117, 138-151.  
892 <https://doi.org/10.1016/j.ssci.2019.04.004>
- 893 Yang, Y., Reimer, B., Mehler, B., & Dobres, J. (2013). A field study assessing driving  
894 performance, visual attention, heart rate and subjective ratings in response to two types  
895 of cognitive workload. In *Proceedings of the Seventh International Driving Symposium*  
896 *on Human Factors in Driver Assessment, Training and Vehicle Design*, June 17-20,  
897 Bolton Landing, New York. Iowa City, IA: Public Policy Center, University of Iowa,  
898 2013: 397-403. <https://doi.org/10.17077/drivingassessment.1518>
- 899 Yeh, Y., & Wickens, C.D. (1988). Dissociation of performance and subjective measures of  
900 workload. *Human Factors*, 30, 111-120. <https://doi.org/10.1177/001872088803000110>
- 901