# Validation of a driving simulator for research into human factors issues of automated vehicles

Nebojsa Tomasevic[1], Tim Horberry[1], Kristie Young[1] and Brian Fildes[1]

[1]*Monash University Accident Research Centre Melbourne, Australia*

Corresponding Author: Nebojsa Tomasevic, Accident Research Centre, 21 Alliance Lane, Monash University 3800, nebojsa.tomasevic@monash.edu, +61 3 9905 1818.

*This peer-reviewed paper was first presented as an Extended Abstract and Oral Presentation at the 2018 Australasian Road Safety Conference (ARSC2018) held in Sydney, NSW, Australia and first published in the ARSC2018 Proceedings in the form of an Extended Abstract. It was expanded into a 'Full Paper' and underwent further peer-review by three independent experts in the field. It is being reproduced here with the kind permission of the authors and is now only available in this edition of the JACRS.*

## Key Findings

- The driving simulator is a valid tool for human factors research in automated vehicles;
- Events and conditions with the best transfer of behavioural validity have been identified;
- Findings will be used for the design of future studies investigating automated driving;
- Further simulator validation issues were identified, e.g. simulator representation of on-road situations requiring high mental workload.

## Abstract

This study evaluated the behavioural validity of the Monash University Accident Research Centre automation driving simulator for research into the human factors issues associated with automated driving. The study involved both on-road and simulated driving. Twenty participants gave ratings of their willingness to resume control of an automated vehicle and perception of safety for a variety of situations along the drives. Each situation was individually categorised and ratings were processed. Statistical analysis of the ratings confirmed the behavioural validity of the simulator, in terms of the similarity of the on-road and simulator data.

## Keywords

Vehicle automation, driving simulator, human factors, validation, willingness to resume control, perception of safety

## Glossary

SAE – Society of automotive engineers
TH – Time headway
POS – Perception of safety
WTE – Willingness to engage automated driving system
WTRC – Willingness to resume manual control of the vehicle
TD – Traffic density
SC – Situation complexity

## Introduction

Driving automation is on the brink of becoming the mainstream from a technological point of view. The SAE classifies six levels of automation. These levels are summarised in Figure 1, with the deployment predictions being derived from multiple sources such as Chan (2017) and Litman (2015). The majority of academic research found is focussed on levels 2 and 3. Level 3 is acknowledged as being associated with the greatest number of human factors issues because it requires the driver to remain in the loop enough to regain manual control in the event of an emergency or if driving conditions move outside of the automation operational design domain (Logan et al, 2017).

There are many unanswered questions from a human factors perspective that are preventing legalisation of automated driving, such as transfer of control from automated to manual driving and driver acceptance of new technology. These questions are difficult to answer without proper testing. The obvious approach to this problem is the utilisation of driving simulators. Simulators provide a

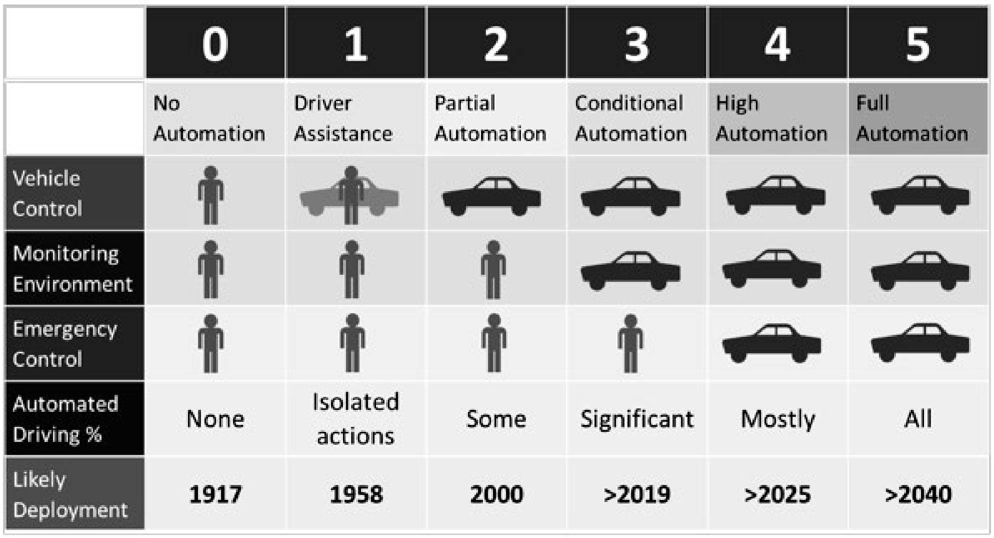


**Figure 1. SAE levels of automation and deployment predictions**

safe, economical and controlled environment in which to conduct automation research. However, this is an artificial environment and these differences may influence the subject's behaviour. Therefore, to be used in automation research, driving simulators need to reproduce similar driver responses to those occurring on the real road. Every driving simulator has its limitations which are directly related to the cues (visual, auditory tactile and vestibular) it is able to provide. Kaptain et al. (1996), state that if the set of cues important to the subject of the investigation is available in the simulator, the simulator may be as valid as a field experiment.

As research simulators are commonly developed independently of each other and have distinct parameters (Godley et al. 2002), it is necessary to validate them on an individual basis. Driving simulators are commonly validated for various specific aspects such as speed perception, vehicle dynamics, hazard perception and many more. Godley et al. (2002) evaluated a driving simulator for speed research establishing relative behavioural validity and relative validity for mean speed. McGehee et al. (2000) examined driver reaction and performance in an intersection crash scenario in the simulator and on a test track. The study produced statistically equivalent reaction times. Underwood et al. (2011) evaluated hazard perception in the simulator and on the road observing similar patterns in behaviour in both settings.

As automated driving is a new field, a study was needed to establish the behavioural validity of the available driving simulator. Behavioural validation involves:

- Comparison of two systems during identical tasks and circumstances in terms of system performance and/or driver behaviour
- Measurement of physical and/or mental workload (physiological measurements)
- Subjective criteria from drivers
- Evaluation of how well the simulator results align to real-world findings

There are very few studies concerning validity of the driving simulator for research into automated vehicles. Eriksson et al. (2017) explored workload differences between a driving simulator and on-road drives in an automated vehicle. In this validation study the authors argued that a driving simulator can be a valid tool for studying users' interactions with automated driving systems. Pariota et al. (2017) observed the effects of connected automated vehicles on car-following behaviour in driving simulators and an instrumented vehicle. Although there were some differences in behaviour between environments, a consistency in car spacing within each environment has been shown.

The current work is part of a larger investigation of human factors issues associated with automated driving. The overall research program aims to explore drivers' willingness to engage or disengage automated driving system, the perception of safety in automated driving and transfer of control between vehicle control modes. The aim of this study was to validate the use of a driving simulator for research in human factors of automated driving. More specifically, a relative behavioural validation study was conducted which will establish a level of credibility and transferability of the simulator results into the real world. To the knowledge of authors, no other validation study had been conducted to answer this specific question in the context of automated driving.

## Method

The study was conducted at the Monash University Accident Research Centre. The data collection was conducted under semi-controlled experimental conditions. The on-road drive was conducted on real roads and in the real traffic but followed a strict route. The simulator drive was programmed to replicate this on-road test route in terms of length, road conditions and other controllable parameters. No safety critical events were part of the experimental drives.

Since an automated vehicle was not available for the study, on-road automated driving had to be controlled by the human driver. Therefore, to keep experimental conditions the same across the settings, participants were aware that a human driver was used to represent automation in both drives. The participants were placed in the passenger seat and did not have access to a steering wheel and control pedals in both conditions. The researcher was in the driver's seat and controlled the vehicle. Participants were instructed to assume a situation in which they were behind the controls of a level 3 automated vehicle that was operating in an automated mode for the entire duration of the drive and that they could resume manual control of the vehicle at any time, but their task was just to answer the experimenter's questions.

The same procedure was followed in the simulator. This way, both experimental conditions were kept as similar as possible. This included obstructing speedometer from the participant in the simulator since the speedometer in the car was not visible from the passenger's seat.

# Participants

There were 20 participants, 11 males and 9 females, ranging in age from 21 to 64 years, with an average age of 36.8 years (SD = 11.2). The median number of years of driving experience was 14.5 (IQR: 9-24.75). Participants were recruited from both Monash University (post-graduate and undergraduate students or staff) and outside using personal contacts. Ethics approval was obtained from Monash University Human Research Ethics Committee. Participants were required to have a full driver's licence and drive at least 6,000 km per year. They were paid $30 for their participation. The total duration of the experiment was between 90 and 105 minutes.

# Equipment

## Instrumented car

The experimental car was an instrumented Holden Commodore VE. It had rear wheel drive and automatic transmission. In addition to the existing instrumentation, a wide-angle camera was used to record the driving scene and audio cues.

## Driving simulator

The MUARC Automation Driving Simulator (Figure 2) consisted of two seats mounted on separate motion bases. Both seats moved in unison. The simulator vehicle represented a car with an automatic transmission. Visuals were presented on three 46" high brightness bezel-less displays. Each display had a resolution of 1080p and the image refresh rate was 60Hz.

The driver and the passenger both had a 140° of horizontal field of view and a 45° vertical field of view. The sound was presented via left, right and centre satellite speakers and a subwoofer. Each motion base produced three degrees of freedom of movement as well as vibration. The same wide-angle camera from the instrumented car was used to record simulator drives and audio cues.

## Experimental questions

**Figure 2. Automation driving simulator setup**

A tablet (iPad) was used to collect answers during both simulator and on-road drives. There were between 20 and 25 questions for each drive and the final overall question completed after the end of drive. Each question consisted of part A and part B. Part A (Figure 3) asked participants to rate willingness to resume control of the vehicle in that situation. The four categories were: very willing, willing, unwilling and very unwilling. Part B (Figure 3) asked participants to rate perception of safety in that situation using a linear scale from 1 to 100 (1 for very unsafe and 100 very safe).

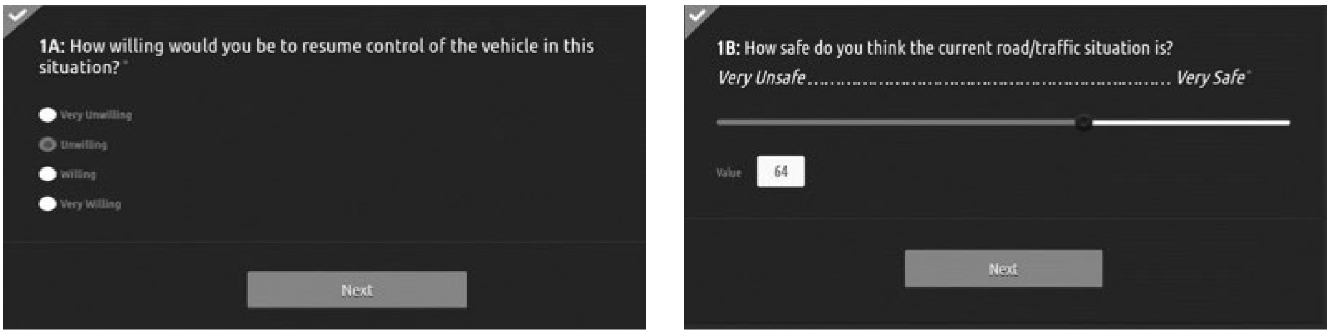


**Figure 3. Example of Part A (Willingness to resume control) and B (Perception of safety) question at the decision point**

### Experimental Drives

The real road and simulator routes were selected to resemble each other as much as possible, taking into account available equipment, time constraints and resources. Overall factors that had to be considered were:

- The total duration of each drive needed to be kept under 30 minutes;
- Total travelled distance during drives needed to be limited to under 20 km;
- The proportion of freeway driving vs urban/residential driving had to be similar;
- Time of the day was between 11:00 and 15:00. This prevented sun glare situations and provided an optimum visibility on the road;
- Peak traffic conditions had to be avoided; and
- Adverse weather conditions had to be avoided (dry roads only).

The following matching criteria between on-road and simulator scenes were used:

- Road lane width;
- Speed limits;
- Number of roundabouts;
- Number of turns;
- Number of freeway entries and exits;
- Number of road bends;
- Traffic density and composition;
- Number of signalised intersections;

The simulator drives were scripted and therefore the same events were presented to each participant. However, during the on-road drives not all events were encountered by every participant. Only events that occurred in both the simulator and on-road drive were analysed.

## Experimental Procedure

Participants completed an informed consent form and read the experimental instructions. They were then given a brief introduction to automated vehicles and presented with a definition of willingness:

- Ready of eager to do something;
- Disposed or inclined;
- Prepared, or
- Acting or ready to act gladly.

This was followed by a demographics questionnaire that also included questions about driving habits, subjective driving skills and attitudes toward technologies.

Participants completed the drives in a counterbalanced order. Half of the participants completed the simulator drive first and the other half completed the on-road drive first. Only one researcher was involved in the experiment.

During the drives, participants were given a tablet which was used to record ratings for willingness to resume control (WTRC) and perception of safety (POS). During the drives, participants were instructed to observe the road and wait for the researcher's verbal instruction: "Ready ... Now!". The instructions were given with enough lead time for participant to recognise the situation ahead. After hearing this cue, participants were instructed to stop observing the road and quickly complete Part A and Part B of the question. After completing the question, participants would continue observing the road until the next question.

At the end of the drives, participants were asked to rate their overall willingness to engage (WTE) automated driving system as well as their perception of safety of the entire drive.

## Data Collection and Processing

During the drives, the following data were collected:

- Video recordings of the road scene;
- Experimental drive questionnaire;
- GPS and vehicle data in on-road drive only;
- Simulator data during simulator drive only;
- Pre-drive and post-drive well-being questionnaires (simulator only).

Using video recording, each decision point was coded for several parameters. They were: time, event name, environment, speed limit, road division, number of lanes, road shape, traffic density, situation complexity and participant comments. These parameters were later used in selecting data for statistical analysis. Traffic density and situation complexity of each event were rated as low, medium and high according to the criteria below.

Traffic density (criteria partially based on Strategic Highway Research Program 2, SHRP2) Levels of Traffic density (VTTI, 2015):

Low:

- Free flow, no lead traffic (0-1 cars ahead within 5s time headway (TH), minimum TH > 3s);
- Freedom to select speed, change lanes and make turns (No vehicles in left or right lanes relative to the participant within 20m radius).

**Table 1. Percentages of events with Levels of Traffic Density (TD) and Situation Complexity (SC)**

| | TD Road | TD Simulator | SC Road | SC Simulator |
|---|---|---|---|---|
| **Low** | 64.24% | 70.45% | 56.53% | 57.39% |
| **Medium** | 30.41% | 20.34% | 40.47% | 28.91% |
| **High** | 5.35% | 8.99% | 3.00% | 13.49% |

Medium:

- Free flow with some restriction (1-3 cars ahead within 5sTH, 2-3s TH);
- Freedom to select speed, change lanes and make turns (vehicle or vehicles in left or right lanes relative to the participant, within 10 – 20m radius).

High:

- Forced traffic flow conditions (3+ cars ahead within 5 seconds TH, minimum TH < 2s);
- Limited freedom to select speed, change lanes and make turns (vehicle or vehicles in left or right lanes relative to the participant, within 10m radius).

Situation complexity levels (partially based on Cabral et al., 2016):

Low:

- No significant cognitive processing is required (clear road, smooth and predictable traffic).

Medium:

- Some cognitive processing required (traffic ahead, approaching intersections or turns).

High:

- Medium to intensive cognitive processing required (dealing with vulnerable or unpredictable road users, complex intersections, aggressive drivers, reduced visibility);
- Critical decision making (merging, overtaking, potential emergency braking).

Based on these criteria, levels of traffic density (TD) and situation complexity (SC) were assigned to every individual event. Distributions of these levels across all events are presented in Table 1.

## Data Analysis

The purpose of the statistical analysis was to determine whether there were differences between ratings (WTRC/WTE and POS) given for similar decision points in both experimental environments (simulator and on-road). Generalised Estimating Equations model was used for statistical analysis. This model is used to estimate the parameters of the generalised linear model with the possible unknown correlation between outcomes. It can be used for both ordinal (WTRC/WTE) and interval data (POS). The data analysis was done by comparing dependent variables (ratings for WTRC and POS recorded during experimental drives in two environments).

Processing of the data resulted in a single rating for each category (individual events and conditions) per participant. In the GEE models, participants were the subject variable and experimental environment (simulator or on-road) were independent variables. WTRC/WTE and POS were the dependent variables. In cases where multiple records existed for a category, the median value was used for ordinal variables (because the data were non-normally distributed) and the mean for linear variables. The correlation matrix that represented the within-subject dependencies was estimated as part of the model.

## Results

Results of the data analysis are presented in Table 2. The table contains a list of all tests conducted on events and driving conditions. Results are primarily expressed as p-values for both WTRC (WTE for the Final question) and POS (Figure 4).

The results for the final questionnaire item, which represents overall WTE and POS ratings for the whole drive revealed that there were no significant differences across the on-road and simulator environments for both WTE ($p=0.315$) and POS ($p=0.324$).

There were no significant differences across environments for WTRC and POS ratings for free driving on the freeway, short time headway, left bend, roundabout, give way/stop sign, congestion, stopped bus, and pedestrians.

Mixed results were obtained for free driving on urban roads where POS was significantly different across the environments, while there was no significant statistical difference in WTRC. Events that produced significant statistical differences in both WTRC and POS were uphill road and merging on the freeway.

Statistical test on levels of traffic density (TD) and situation complexity (SC) indicated that there were no significant statistical differences between on-road and simulator environments. The only exceptions were WTRC for medium TD on the freeway where significant differences were found across environments ($p=0.018$) and POS for medium SC on the freeway ($p=0.045$).

**Table 2. Test results**

| | Mean POS road | Mean POS simulator | SD road | SD sim | p(POS) | p(WTE/ WTRC) |
|---|---|---|---|---|---|---|
| **Final Question** | 70.75 | 73.30 | 3.59 | 3.89 | 0.315 | 0.324 |
| **Free Driving (Freeway) ay) Freeway)** | 71.21 | 76.26 | 3.35 | 3.15 | 0.053 | 0.180 |
| **Free Driving (Urban)** | 69.08 | 75.75 | 3.29 | 2.98 | **0.000** | 0.143 |
| **Short Time Headway** | 48.15 | 52.35 | 5.54 | 5.96 | 0.517 | 0.06 |
| **Left Bend (Freeway)** | 75.79 | 76.51 | 3.41 | 3.42 | 0.811 | 0.210 |
| **Roundabout** | 67.96 | 62.84 | 5.98 | 5.45 | 0.340 | 0.739 |
| **Give Way/ Stop Sign** | 65.36 | 67.58 | 4.04 | 4.60 | 0.492 | 0.657 |
| **Merging (Freeway)** | 56.63 | 74.15 | 4.20 | 3.12 | **0.000** | **0.002** |
| **Changing Lanes** | 62.58 | 51.65 | 5.17 | 4.54 | **0.033** | 0.482 |
| **Congestion*** | 75.93 | 61.79 | 7.52 | 5.96 | 0.127 | 0.089 |
| **Stopped Bus*** | 67.73 | 53.95 | 6.44 | 6.56 | 0.109 | 0.191 |
| **Pedestrians*** | 56.16 | 61.45 | 7.49 | 5.20 | 0.309 | 0.300 |
| **Uphill road*** | 72.43 | 86.55 | 4.17 | 2.38 | **0.000** | **0.015** |
| **Low TD (Urban)** | 68.58 | 70.47 | 3.13 | 3.17 | 0.371 | 0.951 |
| **Medium TD (Urban)** | 64.42 | 63.54 | 4.05 | 4.78 | 0.808 | 1.000** |
| **Low TD (Freeway)** | 73.48 | 72.75 | 2.67 | 3.45 | 0.796 | 0.065 |
| **Medium TD (Freeway)** | 56.53 | 55.86 | 4.54 | 4.83 | 0.815 | **0.018** |
| **Low SC (Urban)** | 71.05 | 73.72 | 2.90 | 3.10 | 0.259 | 0.191 |
| **Medium SC (Urban)** | 62.98 | 60.45 | 3.76 | 4.70 | 0.304 | 0.701 |
| **Low SC (Freeway)** | 73.05 | 76.40 | 2.59 | 3.39 | 0.125 | 0.187 |
| **Medium SC (Freeway)** | 56.70 | 63.38 | 4.47 | 3.80 | **0.045** | 0.216 |
| **High SC (Freeway)*** | 61.33 | 47.30 | 8.58 | 5.15 | 0.160 | 0.968 |

*Events that did not have a full dataset (< 50%)

**Repeated GEE model analysis with only two categories of WTRC (willing and unwilling)

Significant values ($p < 0.05$) are shown in bold.

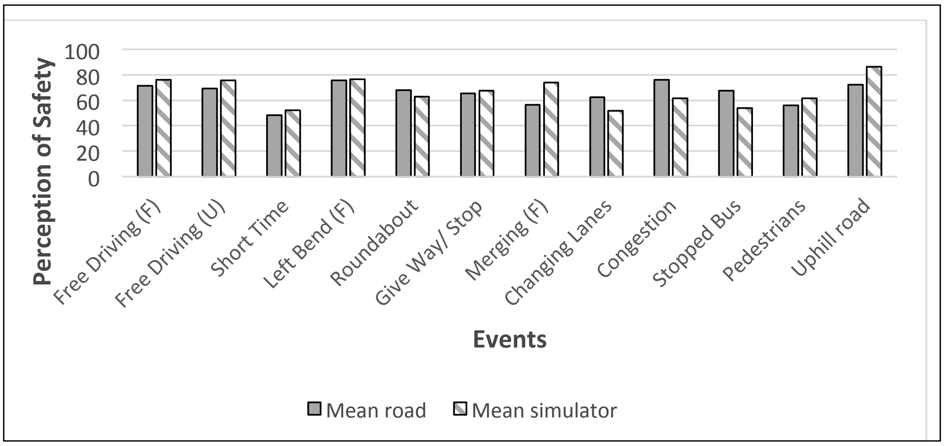


**Figure 4. Perception of safety during events**

## Discussion

The results showed that for the large majority of events, there was no statistical difference ($p > 0.05$) in ratings of WTRC, WTE and POS when comparing the two driving environments. This suggests that these events are well represented in the simulator when compared to the on-road environment in the context of the research question. In their research on driving simulator validity, Kaptain et al. (1996), stated that if the results between the simulator and the field experiment are similar, the simulator is shown to be valid for investigating the studied driving task. Another important element of a succesful behavioural validation study is a carefully designed experimental procedure (Blana, 1996).

From the perspective of further research, it is more interesting to understand what were the differences in experimental conditions that may have contributed to the significant statistical differences. Only two events produced significant statistical differences in ratings for both POS and WTE between experimental conditions. They were merging onto the freeway and to a lesser extent unrestricted driving on an urban road.

Merging onto a freeway could be classified as a high-risk event. This event involved multiple simultaneous manoeuvres (changing lanes, adjusting speed, finding gaps, and continuously scanning the scene) while travelling at a relatively high speed, often in medium or high TD. In comparison with the on-road event, the simulator freeway merging event was simpler (lower TD) and more predictable, therefore demanding less mental workload. Moreover, we speculate that an increase in workload demand exponentially augments perceived risk between the two experimental conditions. Although the merging event in the simulator could be made more demanding by increasing traffic density and speed, further research is needed to answer how exactly perceived risk and mental workload correlate under the simulator and on-road conditions. The exact relationship will, of course, be affected by the specifications of each individual simulator.

Uphill driving on the urban road was intended as a relatively simple and undemanding event so the perceived risk should not be such an important factor. However, statistical test results indicated significant differences in the ratings between environments. Due to limitations in the selection of roads, not all experimental conditions could be accurately matched. In the simulator drive, this event occurred on the four-lane road, while in the on-road drive it occurred on a two-lane road with occasional parked cars on both sides of the road. To participants, the on-road event may have appeared less safe than the simulator event and thus, influenced their WTRC ratings. These observations are supported by Fildes et al. (1989) who found that road width and number of lanes had the strongest influence on judgements of safety and travel speed, while the roadside environment also had an effect but to a lesser degree. Finally, it is believed that the differences in WTRC for Medium TD on the freeway and POS for medium SC on the freeway are due to the challenges in creating a realistic freeway driving environment in the simulator.

It is important to accurately represent an event in the simulator; however, differences between the real and the simulated environments related to simulator measurements and mental workload emerge whatever the cost of a driving simulator is. Harms et al. (1996), observed that increasing the face validity of the VTI driving simulator did not necessarily enhance the overall behavioural validity of the simulator. More research is needed when investigating on-road situations that create a high mental workload and their representation in the simulator. This will be especially important in the case of take-over requests. In addition, it is important to understand the precise conditions under which drivers are willing to engage or disengage an automated driving system. A future study to investigate this is currently being undertaken by the authors.

Given that an automated vehicle was not available for this study and Level 3 Automated vehicles are not legally allowed to travel on Australian roads, we adopted a protocol whereby participants sat in the front passenger seat of the real and simulated vehicles which were driven by an experimenter. Participants were asked to imagine that he or she was in the driver's seat of an automated vehicle and answer the questions from this perspective. This method may, of course, lead to differences in participants' perception of safety and trust in vehicle automation. However, we estimate only a small impact of these limitations because the main task was to enter ratings in the questionnaire (willingness to resume control of the vehicle and perception of safety during the drive) and not to drive or respond to take over requests. We were also interested in comparing ratings across the on-road and simulated environments, which were kept as similar as possible in terms of the automation protocol.

## Conclusions

The results confirmed the relative behavioural validity of the MUARC automated driving simulator. We argue that if certain limitations of the driving simulator are taken into account absolute behavioural validity can be confirmed.

These findings will be used for the design of future simulator experiments investigating willingness to resume control or engage an automated driving system, the associated perception of safety and driver behaviour during transfer of control.

## Acknowledgements

This work was funded by Monash University research scholarship and RACV scholarship award.

## References

Blana, E. (1996). *Driving Simulator Validation Studies: A Literature Review. Bilingualism* (Vol. 110). Leeds, UK.

Cabrall, C. D. D., & De Winter, J. C. F. (2017). What Makes Driving Difficult ? Perceived Effort Follows Semantic Complexity Factors of Driving Scenes. *Proceedings Road Safety & Simulation International Conference 2017*

*(RSS2017)*. The Hague.

Chan, C.-Y. (2017). Advancements, prospects, and impacts of automated driving systems. *International Journal of Transportation Science and Technology*, 6(3), 208–216.

Eriksson A, Banks V A, S. N. A. (2017). Transition to Manual: comparing simulator with on-road control transitions Eriksson. *Accident Analysis and Prevention*, *102*, 227–234.

Fildes, B.N., Leering, and A.C. & Corrigan, J. McM. (1989). Speed *Perception 2: Driver's judgements of safety and travel speed on rural curved roads and at night*, Federal Office of Road Safety, Report CR60, Department of Transport and Communications, Canberra

Godley, S. T., Triggs, T. J., & Fildes, B. N. (2002). Driving simulator validation for speed research. *Accident Analysis and Prevention*, *34*(5), 589–600.

Harms, L., Alm, H., & Tornos, J. (1996). The influence of situation cues on simulated driving behaviour: a summary of three validation studies. *Paper presented at the Symposium on the Design and Validation of Driving Simulators*. ICCTP'96 Valencia.

Kaptein, N., Theeuwes, J., & Van Der Horst, R. (1996). Driving Simulator Validity: Some Considerations. *Transportation Research Record: Journal of the Transportation Research Board*, *1550*(November), 30–36.

Litman, T. (2015). Autonomous Vehicle Implementation Predictions: Implications for Transport Planning. Victoria Transport Policy Institute, Victoria.

Logan, D.B., Young, K.L., Allen, T., Horberry, T., 2017. Safety benefits of cooperative ITS and automated driving in Australia and New Zealand. Report AP-R551-17. Austroads, Sydney, October 2017.McGehee, D. V., Mazzae, E. N., & Baldwin, G. H. S. (2000). Driver Reaction Time in Crash Avoidance Research: Validation of a Driving Simulator Study on a Test Track. In *Proceedings of the IEA 2000/ HFES 2000 Congress* (pp. 320–323).

Pariota, L., Bifulco, G. N., Markkula, G., & Romano, R. (2017). Validation of driving behaviour as a step towards the investigation of Connected and Automated Vehicles by means of driving simulators. *5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems, MT-ITS 2017 - Proceedings*, 274–279.

Underwood, G., Crundall, D., & Chapman, P. (2011). Driving simulator validation with hazard perception. *Transportation Research Part F: Traffic Psychology and Behaviour*, *14*(6), 435–446.

VTTI. (2015). SHRP2 Researcher Dictionary for Video Reduction Data. Blacksburg, Virginia: Virginia Tech Transportation Institute.