Cooperative Truck Platooning Systems Trial

Driver Experience Assessment
Final On-Road Trial
September 13, 2022

Contents

1 Executive Summary 5

2 INTRODUCTION 8

3 DRIVER DEMOGRAPHICS 16

4 ANALYSIS PROCESS 16

3.1 Hypothesis Driven Data Analysis 16

3.2 Multi-Modal Data Analysis 17

5 QUESTIONNAIRES 18

4.1 Overview 18

4.2 Analysis of Non-Platooning Activities 19

4.2.1 Modified Risk Perception Questionnaire (MRPQ) 19

4.2.2 Driver Stress Inventory (DSI) 21

4.3 Analysis of Platooning Activities 25

4.3.1 Karolinska Sleepiness Scale (KSS) 25

4.3.2 Driving Activity Load Index (DALI) 26

4.3.3 Trust in Automation (TiA) 29

4.3.4 Acceptance Scale for Advanced Transport Telematics (AATT) 31

4.3.5 Equipment and Comfort (E&C) 33

4.4 Conclusion 35

6 TABLET-BASED COGNITIVE TASKS (TBCT) 36

6.1 Overview 36

6.1.1 Reaction Time (RT) 36

6.1.2 Judgment Task 37

6.1.3 Memory Task 37

6.1.4 Control Task 38

6.2 Collected Data 39

6.3 Analysis 39

6.3.1 Reaction Time (RT) 39

6.3.2 Judgment Task 40

6.3.3 Memory Task 42

6.3.4 Control Task 43

6.4 Conclusion 44

6 EYE TRACKER DATA....45
6.1 Overview....45
6.2 Data Acquisition....45
6.3 Collected Data and Preprocessing....46
6.4 Analysis....47
6.4.1 Equipment validation and comparison....47
6.4.2 The effect of the ambient light on pupil diameter....48
6.5 Conclusion....49
7 BRAIN DATA....50
7.1 Overview....50
7.2 Data Acquisition....51
7.3 Collected Data and Preprocessing....51
7.4 Analysis....53
7.4.1 Equipment validation and comparison....53
7.4.2 The effect of time of day on brain oscillations....54
7.5 Conclusion....56
8 MULTI MODAL ANALYSIS....57
8.1 Cut-ins Events....57
8.2 Hard Braking Events....58
9 OUTLIER ANALYSIS....61
9.1 Low Platooning Engagement....61
9.1.1 The effect of time of day on brain oscillations....61
9.1.2 Modified Risk Profile Questionnaire....62
9.1.3 Driver Stress Inventory (DSI)....62
9.1.4 Sleep quality and alertness....63
9.1.5 Trust in Automation (TiA)....64
9.1.6 Acceptance Scale for Advanced Transport Telematics (AATT)....65
9.1.7 Driving Activity Load Index (DALI)....66
9.1.8 Conclusion....67
9.2 Alberta Trucker Convoy Protest....68
10 SUMMARY....71
11 CONCLUSION, LESSONS LEARNED, & FUTURE WORK:....75
11.1 Lessons Learned and Areas for Improvement....75
11.2 Future Work....77

Appendix A: Human-Machine Interface Challenges 79
System Reliability 79
Driver Distractions 79
Driver Trust 79
Appendix B: On-Road Trials Trip Configurations 81
Appendix C: Data Collection Protocol 83
Appendix D: Questionnaire Responses 89
Appendix E: Tablet-Based Cognitive Tasks (TBCT) 92
Reaction time task 92
Judgment task 92
Memory task 93
Control task 94
Appendix F: Eye Tracking System (Smart Eye) 95
General Specifications and features for Smart Eye Pro Cameras 95
Extracted Measures 95
References 98

Cooperative Truck Platooning System Trial

Title	Cooperative Truck Platooning System (CTPS) Trial
Client	Government of Canada – Transport Canada
Contract Number	T8009-190376/001/SL dated 2021-02-09
Project Start Date	March 1, 2021
Project End Date	September 29, 2022

Document Revision History

Issue No.	Issue Date	Status	Authors	Comments
1.0	13 September 2022	Final Release	Reyhaneh Bakhtiari	Version circulated for Final review.

Document Purpose

This document provides details of the drivers' experience assessment during the cooperative truck platooning system (CTPS) on-road trial that includes the test matrix, questionnaire, cognitive tasks, data collection devices, data analysis methods, and results.

Confidentiality

The information contained in this document is the property of the Alberta Motor Transport Association. Except as specifically authorized in writing by the AMTA, the holder of this document shall keep the information contained herein CONFIDENTIAL and shall protect same in whole or in part from disclosure and dissemination to third parties and use same for evaluation purposes only.

Note: The University of Alberta ("Contractor") is furnishing this report "as is". Contractor does not provide any warranty of the contents of the report whatsoever, whether express, implied, or statutory, including, but not limited to, any warranty of merchantability or fitness for a particular purpose or any warranty that the contents of the report will be error-free.

1 Executive Summary

This main purpose of the current report is to assess driver experience of cooperative truck platooning system (CTPS). To fully understand the impact of automation and platooning on drivers, psychophysiological data were collected during their trips and behavioral and cognitive tools measured their performance across various tests before and after each trip.

As the project landed, the naturalistic nature of a study in an industrial setting with many safety regulations imposed some constraints and challenges on the project. In particular, observations on the platooning performance during the human factors on-road pilot study, between October 21st and November 3rd, 2021, reflected in the Interim Report, deliverable #, impacted the scope of the project. In what follows, some of the challenges during the study that resulted in deviation from the initial methodology have been briefly reviewed:

- The observation and the provided feedback during the on-road pilot study, resulted in an update on the scope of the project; changing the six months' data collection period that was supposed to start in February 2022, to a three weeks of data collection in January 2022.

- Feedback from the on-road trial resulted in Changing the protocol to include an in-truck instructor/trainer to accompany the drivers. Having a second person in the truck cab worked against a hypo-vigilant driving state and one of the main hypotheses of the study regarding the passive fatigue, as passengers affect the social context of a driving scenario, and the outcome due to changes in attention.

- Based on the sample size calculation, it is required to have 20 drivers to have enough power to find differences across conditions. By the end of December 2021, 15 drivers volunteered to participate in the study, in either day or night working shifts, and we were expecting more drivers would join our study as we start data collection. Besides, in the original design, drivers were paired together based on their regular working hours, and each driver pair was required to drive together in all six round trips between Calgary and Edmonton. Due to short notice schedule change of the project, and reduced data collection duration, nine drivers were randomly matched together with each driver participating between one to five driving sessions.

- Due to changes on the timing of the project, shake dawn trips were not completed before the on-road trials to capture the issues in real life scenarios. This resulted in not getting the glitches of the data collection pipeline and having a lower data quality than expected.

- Driving in a system that has not been fully developed for commercialized production, led to drivers experiencing the stress and unexpected issues of a newly developing technology vs. the experiencing the automation or platooning system.

- A set of accurate, reliable non-intrusive, and wearable equipment and measurement tools are required to measure drivers' psychophysiological data . In general, such devices are limited, less accurate and more expensive compared to the in-office research equipment. Ambiotex, the company that provided wearable heart rate variability t-shirts, filed bankruptcy in December 2021, and the alternative device from Movisens was not accessible by the advanced on-road trial date in January 2022.

- Due to safety considerations, Alberta Transport only approved platooning intervals of 4s and 5s during night time. Not being able to try the 3s interval for night shift drivers was another source of bias in the design. This did not impact the final project, as no night time data was collected for this study.

The changes, and more specifically reducing the data collection duration, having instructor/trainer in the cab revised the focus of the project to become a proof of concept and exploratory in nature, where the emphasis became to validate a working methodology in a field operation test. This has been verified by showing the equipment validation of the eye tracker data and brain data collection devices and successfully demonstrating the impact of extended hours of driving and time of the day on various collected data.

- This study was successful in showing that real-time psychophysiological data could be collected from professional drivers while conducting real-world, on-road trucking operations on a major highway.

- The questionnaires and accompanying scales used in the study showed high validity and reliability. Insight from these subjective measures also identified the need for greater consideration of macro ergonomics prior to soliciting and engaging drivers (e.g., schedule design, length of research days, impact on worker and home life, driver training requirements, researcher training requirements, communication structures, etc.) and micro ergonomics (e.g., cab design, seat features, workloads, etc.).

- The TBCT consisted of a battery of 4 cognitive tests that have been extensively used to assess fitness to drive in the context of aging, drug use, and various workplace stressors. The main preliminary finding from TBCT was that there was sensitivity in the spatial-attention RT task to the long workday, but the other measures were predominantly insensitive. This was an important finding at this stage because it could have implications for future studies.

- Real-time eye tracking data was continuously collected as a measure of driver fatigue and vigilance. We were able to verify that the quality of the collected data in a real-world on-road trucking environment is as good as data collected in the laboratory on a stationary table. There seemed to be some loss in the quality of gaze direction data from the beginning to the end of three hours driving session, but it remained measurable. The eye tracking data showed a clear sensitivity to the amount of ambient sunlight in the truck. That is, the diameter of the drivers' pupils was sensitive to the low light of early morning (larger diameter) and the brighter light of later in the day (smaller diameter). This is encouraging because pupil diameter is well known to correlate with aspects of fatigue and vigilance. Thus, our findings suggest that dash-mounted eye tracking has the strong potential to be a valuable experimental tool in future studies of this nature.

- Comfortable ambulatory EEG technology is relatively new and has not been extensively tested in real-world skilled performance scenarios like commercial trucking over many hours. We showed that good quality data can be collected in a real-world on-road trucking study and those oscillation frequencies known to be associated with fatigue and vigilance can be collected and quantified. The electrode array was sufficient to collect the data over the length of the trial, but the lateral electrodes showed less connection with slightly more loss of signal over time compared to the central electrodes. We observed some interesting alpha, beta, and theta results when comparing the early and later parts of the trips in line with previous studies showing that fatigue can be associated with changes in these frequency bands, perhaps due to changes in arousal, or the participants attempting to stay vigilant when fatigued. We are encouraged by these results and think it would be valuable to study fatigue in the context of EEG while driving more thoroughly.

It should be emphasized that, as a result of changes to the experimental structure and sample size, the original hypotheses could not be tested and no conclusions can therefore be formulated on the effects of automation and/or platooning on driver experience on the basis of this dataset.

The main lessons learned from the project listed as follows:

- In a multi-disciplinary study, where human participants specifically high skilled workers and safety regulation are involved, it is advantageous to the project, to involve he human factors team during the pre-planning and project planning stages to identify and communicate the critical unanticipated system and process flaws in the system, instead of facing them at later stage of the project, to minimize the risk of unpredicted changes in the project.

- We learned that the shake-down trials were a very critical stage, and we would have benefited from more time to analyze the data to catch glitches, and also have more time to practice data collection pipeline with researchers.

- We learned the importance of involving drivers' input in design and know their consideration when working with high skill workers over an extended length of time for several sessions and communicate the relevant aspect of the project.

The future directions are:

- Despite all the changes in the project, we were able to collect more than 100 hours of brain and eye-tracker data from drivers, as well as events during the road.

- We are continuing on analyzing the data, using non-traditional hypothesis driven models, such as machine learning and regression based models. We plan to prepare and submit a methodological paper about this study to help other transportation and human factors researchers.

- We are planning to investigate the validity of using Muse brain sensing headband with a 256 electrode EEG system on a driving simulator.

2 INTRODUCTION

The Alberta Motor Transport Association (AMTA), in collaboration with Alberta Transportation, Bison Transport, Pronto, Solaris Fatigue Management, Tantus and University of Alberta, conducted platooning trials on public roads to better understand the operational impacts, benefits and limitations of interaction with advanced driver assistance system (ADS) including platooning technology. This document provides details of drivers' experience assessment during on-road trials including test matrix, questionnaires, cognitive tasks, data collection devices, data analysis methods, and results.

This study was intended to be the first of its kind, implemented by a collaborative team highly invested in seeing the project through to completion. Based on the designed methodology, and the selected data collection equipment, including the eye-tracker and brain sensing headband, the conducted power analysis resulted in selecting 20 drivers as the sample size for this study. The study was originally planned with 5 main outcome measures to test 8 separate hypotheses about driver passive fatigue, vigilance, stress, and workload related to critical aspects of truck Platooning.

Due to project constraints outside of our control including system reliability issues and subsequent driver distraction and trust ramifications (Appendix A), the final study included nine drivers participating in the on-road trials between January 12, 2022, and January 30, 2022. On-road trials were conducted on Highway 2 between Calgary and Edmonton. Platooning was to be completed on Highway 2 between Airdrie and Leduc, a distance of 234 km (Figure 1). To consider a trip viable for inclusion in the assessment, copilot or platooning was required to be engaged for at least 30 minutes, as has been shown in a previous study to be long enough to demonstrate driver performance decrements, particularly a substantial loss of engagement (Saxby et al., 2013). It is also a minimum time to see the impact of time-on-task effect on cognitive tasks (Randall et al., 2014).. Appendix B shows the duration of active copilot and platooning for each truck in each trip, as well as the trip configuration.

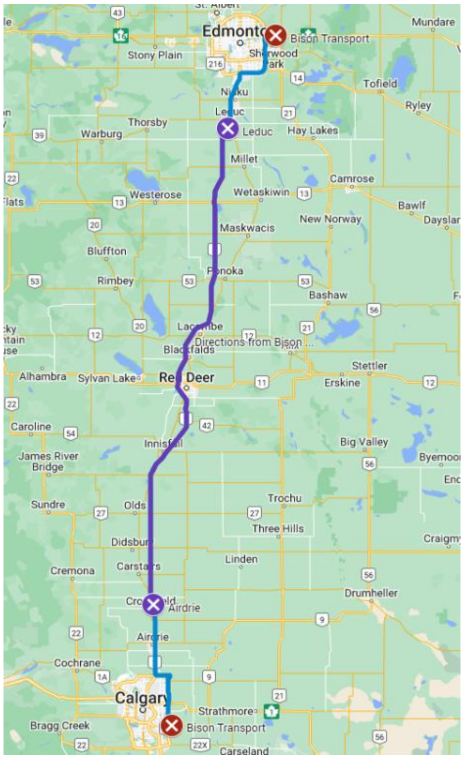


Figure 1: On-road trial location. Drivers started platooning trip between Airdrie and Leduc, as identified by the purple marker on the map.

All trips originated from the Bison-Transport yard just outside of Calgary, located at 234090 Wrangler Rd, Rocky View, Alberta, T1X 0K2. Before starting the trip, the drivers performed a set of tablet-based cognitive tasks (TBCT), and a set of questionnaires which took approximately 20 minutes. TBCT have been extensively used to assess fitness to drive in the context of aging, drug use, and various workplace stressors. We chose to include these measures in the present study to complement the continuous monitoring data and the questionnaire data. During this time, researchers initiated the data acquisition and continuous data collection systems that included the brain activity sensing headband and dash mounted eye-tracker systems. The protocol for performing these tasks is provided in Appendix C. These tasks formulated the Pre Trip session. Figure 2 shows the on-road trial diagram.

Continuous data was collected while drivers collected their loads and then proceeded to Edmonton via Highway 2. Trailer loads were attached to the truck cab at either the Bison yard or an alternate yard location as defined by dispatch. Drivers were instructed to activate the platooning/copilot system between Airdrie and Leduc. This portion of the trip is referred to as Trip 1 throughout this document as noted in Figure 2.

After arriving in Edmonton, drivers dropped off their loads, and picked up their return load. Trial dispatchers tried to use the Bison Transport Yard at 80 Liberty Rd, Sherwood Park, AB T8H 2J6 as often as possible. Customer yard pickups added additional time to the route and commonly occurred throughout the trials.

Researchers met up with the drivers at either the AMTA facility located at 3599 56 Ave E, Edmonton International Airport, AB, T9E 0V4 or at the Bison Transport yard in Sherwood Park. At this point drivers were requested to answer a set of questions and perform tablet based cognitive tests (TBCT). As the drivers completed their questionnaires and TBCT activities, researchers stopped the previous continuous data collection sessions, backed up trail data, and started a new continuous data collection session. To reduce the cumulative trip time, it was decided to remove the TBCT tests from the Post Trip 1 trial schedule after January 19th. This portion of the trip is referred to as Post Trip 1 in Figure 2.

The return trip from Edmonton to Calgary is referred to as Trip 2 throughout this document. Prior to leaving for Trip 2, drivers were reminded to activate the copilot/platooning system between Leduc and Airdrie.

Upon arrival in Calgary, drivers dropped off their load and met researchers for one final assessment at Bison Transport, 234090 Wrangler Road, Rocky View, T1X 0K2. At Bison Transport, drivers performed a final set of TBCT activities and completed a set of questionnaires while researchers shut down the data collection systems and backed up the continuous data collection completing the Post Trip 2 session.

It should be noted that the drivers started their trips during sunrise, experienced the mid-day light in Edmonton, and ended their drive in Calgary during sunset in January. Data collected during the day was influenced by the drivers' biological clock and eye behavior was subject to changes in the ambient light during differing intervals of the day.

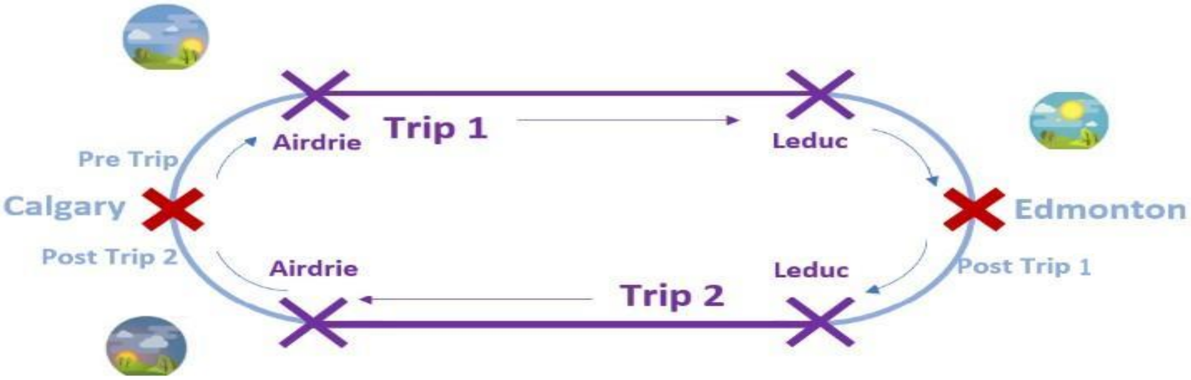


Figure 2: Road Diagram

Platooning involves two trucks driving at close distance (3s or 5s intervals), while the SAE level 2 automation (referred as copilot throughout this document) are activated (i.e. both speed and steering wheel are controlled by the automation system). It should be noted that the platooning interval of 3s and 5s, was later accurately measured to be 4s and 6s respectively. These intervals signify the outer limit for platoon engagement based on theorized fuel economy factors, which is one of the key drivers for implementing automated systems within the commercial vehicle industry. The lead and follower truck driver experiences are not the same, as, besides their physical location differences, the follower truck receives operational signals from the lead truck resulting in a less demanding job as proposed by the platooning technology developers. To investigate the platooning impact, we need to disentangle various impacting factors and assess each factor separately. Table 1, shows the various conditions that drivers were supposed to participate in according to the original plan.

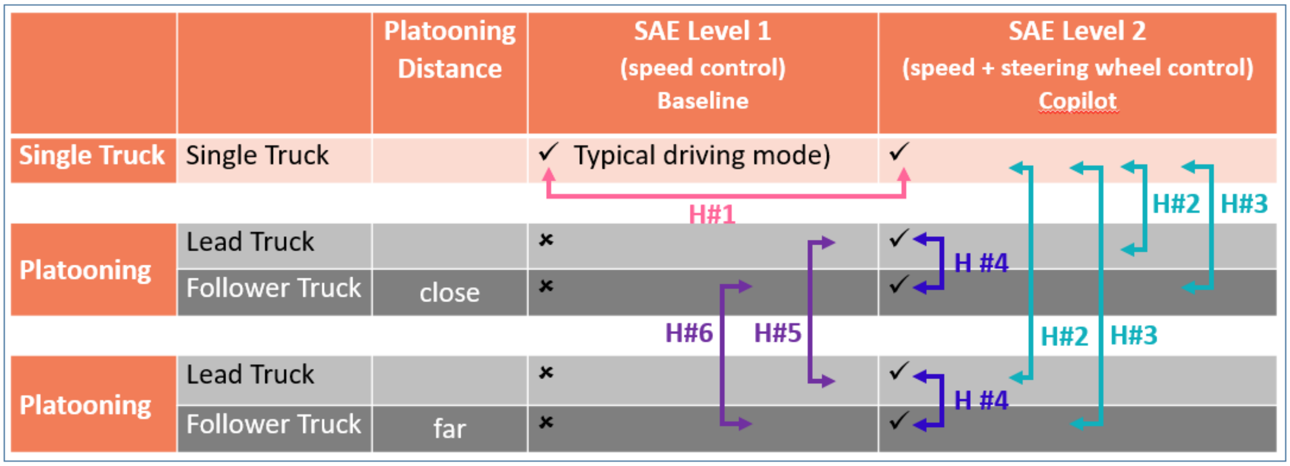
The first layer is the impact of SAE Level 2 on drivers' experience. In the commercial truck industry, compared to other driving situations, where full manual driving is used, the default mode of driving is SAE Level 1, in which the speed is controlled by the automation system. This is referred as the single truck baseline condition throughout the design. To study the impact of automation level, the drivers' performance while driving in Single Truck while driving in baseline (SAE Level 1) vs copilot (SAE Level 2) was compared (H#1).

The next layer is driving in platooning (two trucks at a close distance). This only happens while SAE Level 2 is activated. Therefore, to specifically examine the impact of this condition, the performance while driving platooning is compared with the single truck copilot condition (H#2, H#3).

The next factor is the truck driver's role in either the lead or follower truck (H#4). This was designed by making sure each driver would drive in each of these roles.

Finally, to assess the platooning distance on drivers' experience all drivers tried both 3s and 5s platooning intervals while driving in either lead or follower roles (H#5, H#6).

Table 1: Hypothesis testing matrix



		Platooning Distance	SAE Level 1 (speed control) Baseline	SAE Level 2 (speed + steering wheel control) Copilot
Single Truck	Single Truck		✓ Typical driving mode)	✓
Platooning	Lead Truck		×	✓
	Follower Truck	close	×	✓
Platooning	Lead Truck		×	✓
	Follower Truck	far	×	✓

H #1: investigate the automation impact

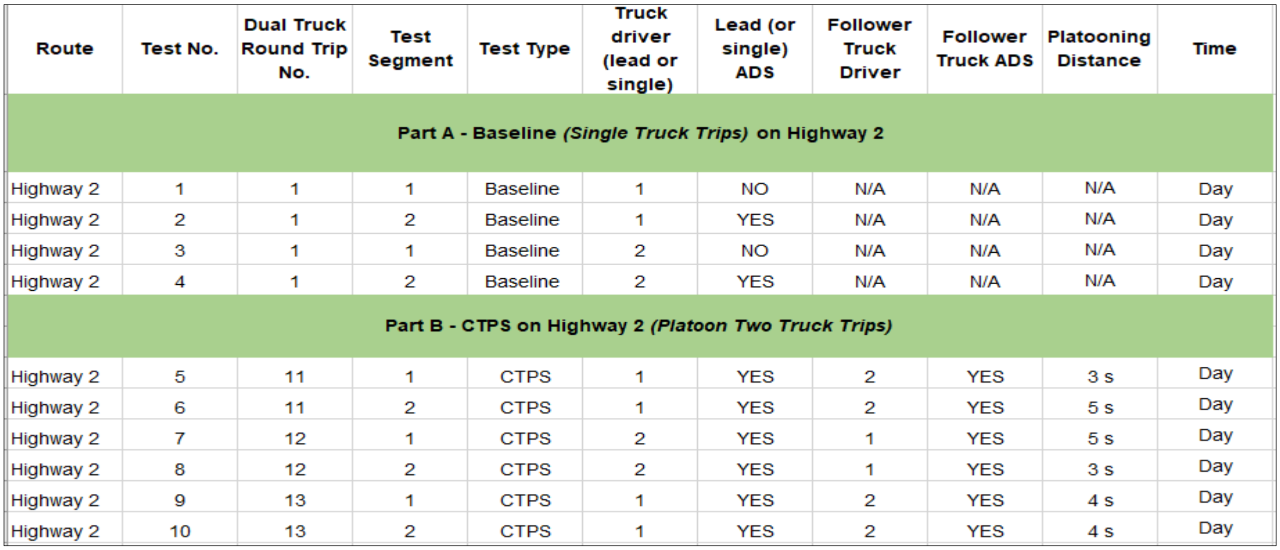
H #2 & H #3: investigate the impact of driving in close distance situation

H #4: Investigate the difference between two truck roles (Lead vs Follower)

H #5, & H #6: investigate the impact of platooning distance on each driver role

To meet the necessary power analysis objectives, it was determined that 20 drivers would be required to participate in the on-road trials. The initial test matrix (Table 2) was designed in such a way that all 20 drivers would experience each condition (single truck baseline, single truck copilot, driving in lead or follower role during platooning, at a 3 or 5 second platooning distance) and the order in which they performed each condition was counterbalanced to have minimal impact from confounding factors such as different exposure experience, time of the day, etc. Due to an accelerated timeline for the trial, a significantly reduced number of drivers (9) participated in the trials. Scheduling limitations were also experienced as Bison Transport assigned drivers for each operational run. Table 3 shows the actual driving conditions experienced by each driver.

Table 2: Initial Test matrix



Route	Test No.	Dual Truck Round Trip No.	Test Segment	Test Type	Truck driver (lead or single)	Lead (or single) ADS	Follower Truck Driver	Follower Truck ADS	Platooning Distance	Time
Part A - Baseline *(Single Truck Trips)* on Highway 2										
Highway 2	1	1	1	Baseline	1	NO	N/A	N/A	N/A	Day
Highway 2	2	1	2	Baseline	1	YES	N/A	N/A	N/A	Day
Highway 2	3	1	1	Baseline	2	NO	N/A	N/A	N/A	Day
Highway 2	4	1	2	Baseline	2	YES	N/A	N/A	N/A	Day
Part B - CTPS on Highway 2 *(Platoon Two Truck Trips)*										
Highway 2	5	11	1	CTPS	1	YES	2	YES	3 s	Day
Highway 2	6	11	2	CTPS	1	YES	2	YES	5 s	Day
Highway 2	7	12	1	CTPS	2	YES	1	YES	5 s	Day
Highway 2	8	12	2	CTPS	2	YES	1	YES	3 s	Day
Highway 2	9	13	1	CTPS	1	YES	2	YES	4 s	Day
Highway 2	10	13	2	CTPS	1	YES	2	YES	4 s	Day

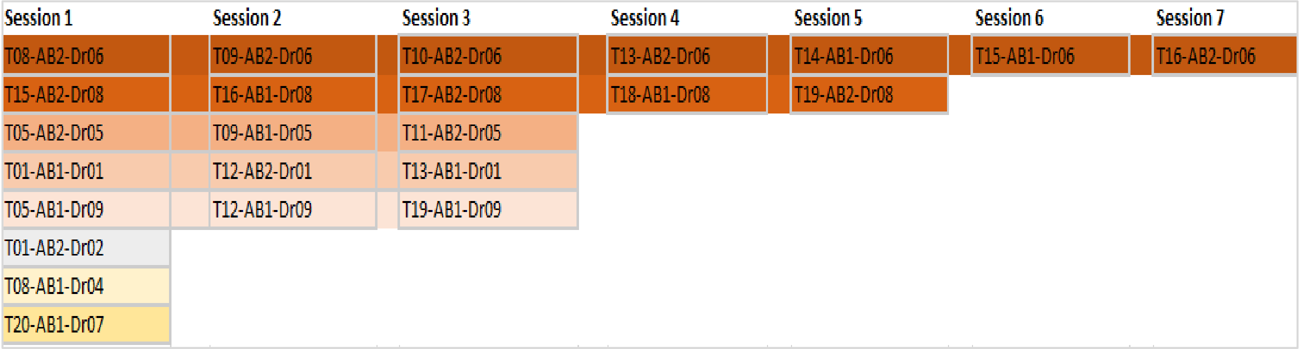
Table 3: Trips completed by drivers 1-9 (Y: successful trip completed N: successful trip was not completed)

	Single Truck		Platooning, Lead role (AB1)			Platooning, Follower role (AB2)		
Driver	**Baseline**	**Copilot**	**3 sec**	**4 sec**	**5 sec**	**3 sec**	**4 sec**	**5 sec**
Dr01	Y	Y	Y	Y	Y	N	N	N
Dr02	Y	N	Y	Y	N	Y	N	Y
Dr03	N	N	Y	N	Y	Y	N	N
Dr04	Y	N	Y	Y	Y	Y	N	Y
Dr05	Y	Y	Y	N	N	Y	N	N
Dr06	Y	N	Y	N	Y	Y	Y	Y
Dr07	Y	N	Y	N	Y	Y	Y	Y
Dr08	N	N	Y	N	Y	Y	Y	Y
Dr09	Y	N	Y	Y	Y	Y	N	N

As part of the initial plan, contingency trips were allotted to replace operational runs with low platooning engagement or low quality data. A project scope amendment mid trial reduced the project timeline and removed the ability to repeat runs that yielded insufficient platooning or data outcomes. As a result, the collected data for assessing drivers' experience was heavily unbalanced and the project was redefined as a pilot study. The revised focus of this study is a proof of concept that is exploratory in nature and will emphasize validating the proposed methodology as the primary objective. Therefore, no conclusions can be reached from this data with regards to any of the original hypotheses (i.e. the impact of automation and/or platooning on driver experience).

In Table 4, the data collected from the Pre Trip tablet-based cognitive tasks for each of the drivers is shown. For three drivers, data was collected during only 1 session. Other drivers were able to complete 3, 5, or 7 sessions. Throughout this study, whenever there were multiple instances of a specific driver's data, the last session was selected for analysis, as this approach maximized the chance of capturing the platooning effect on the driver, rather than their exposure to a new technology.

Table 4: Available Pre Trip tablet-based cognitive tasks for each driver



Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
T08-AB2-Dr06	T09-AB2-Dr06	T10-AB2-Dr06	T13-AB2-Dr06	T14-AB1-Dr06	T15-AB1-Dr06	T16-AB2-Dr06
T15-AB2-Dr08	T16-AB1-Dr08	T17-AB2-Dr08	T18-AB1-Dr08	T19-AB2-Dr08		
T05-AB2-Dr05	T09-AB1-Dr05	T11-AB2-Dr05				
T01-AB1-Dr01	T12-AB2-Dr01	T13-AB1-Dr01				
T05-AB1-Dr09	T12-AB1-Dr09	T19-AB1-Dr09				
T01-AB2-Dr02						
T08-AB1-Dr04						
T20-AB1-Dr07						

The analysis of continuous data centered on data collected between Airdrie and Leduc, during which copilot or platooning was planned to be activated. Therefore, each trip was split into 3 sections: the first 45 minutes (Section 1), the last 45 minutes (Section 3), and the rest (Section 2) (Figure 3).

Within each section, and for each trip separately, 10 minutes of clean data were identified for analysis and are referred to as Segments. Reference to Seg 1, Seg 2, and Seg 3 will be used throughout this document. The definition of clean data will be reviewed for each modality (eye-tracker and brain data) later in the report. When comparing different conditions such as Trip 1 vs. Trip 2, or the impact of driving role (lead vs follower) data in Seg 3 was considered as the best representative of the trip.

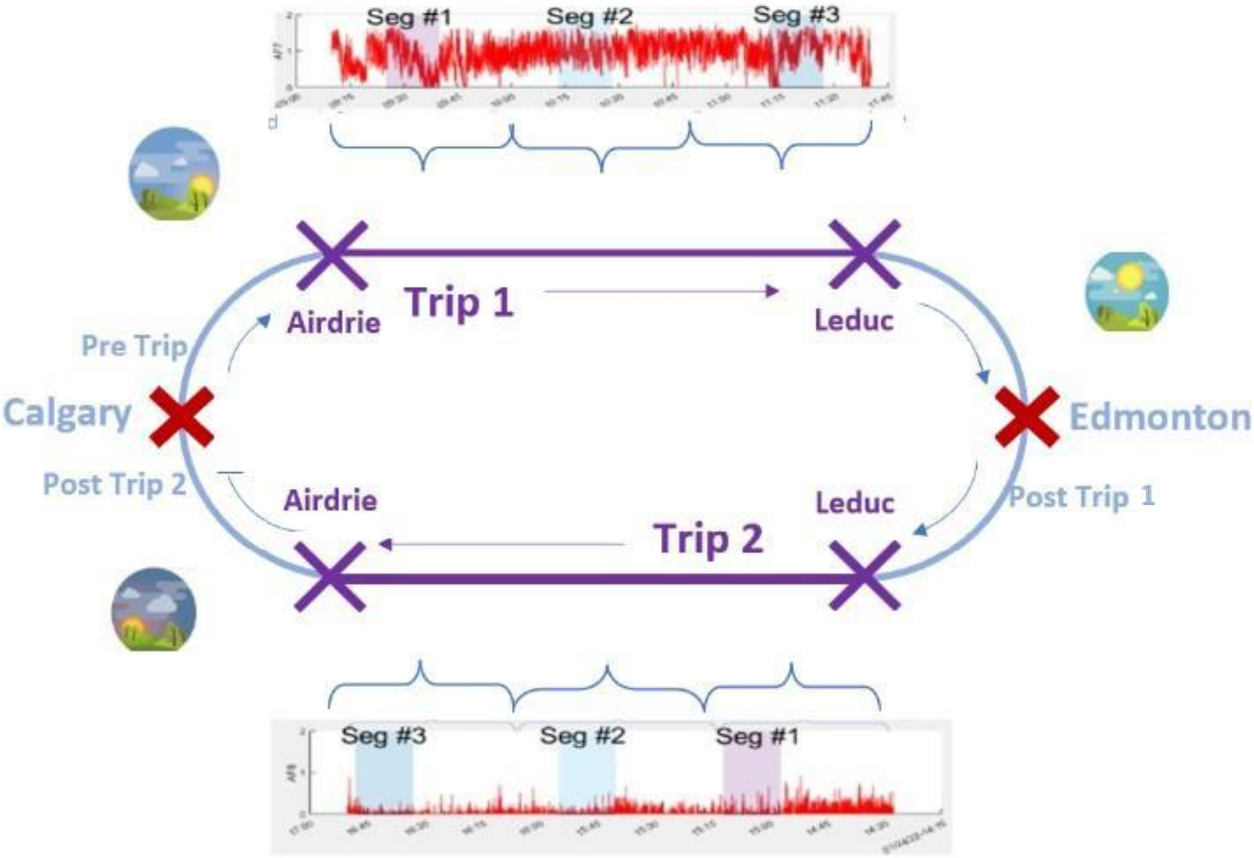


Figure 3: Road Segment

Chapter 2 reviews the driver demographics while Chapter 3 provides an overview of the original hypothesis and multimodal data analysis. Chapters 4 and 5 focus on the data collected during pre and post trips sessions (i.e., questionnaire and TBCT, respectively). Chapters 6 and 7 focus on results from the continuous data collected during driving (EEG-headband and eye-tracker). In chapter 8, the impact of some on-road events on drivers' psychophysiological data is presented. Chapter 9 addresses the trips that did not meet the necessary criteria for inclusion in the analysis and the underlying conditions that led to this situation. Chapter 10 and 11 provide a trial summary, a conclusion, and recommendations for future studies.

It should be noted that the analyses presented are based upon the hypothesis-driven comparison as portrayed in Table 1. All conditions were tested using both within-group analysis (each driver participated in both conditions to be compared), and between group analysis (the participants of the groups to be compared are not the same). In the results section, those results that were significant (p-val<.05) or had a trend to be significant (p-val<.1) or were meaningful are presented.

Throughout this report, orange asterisks on the figures are used to represent a trend for a change (p-val<.1) and red asterisks as a significant change (p-val<.05) across conditions.

Additional factors impacted the planned data collection:

- As proposed in the methodology, it was initially planned to collect the drivers' heart rate data using an Ambiotex t-shirt with a recording device. Ambiodex filed for bankruptcy in November 2021, and the EcgMove 4 from Movisens was selected as an alternative device. Although an order was placed for these heart rate monitors, the devices were not received in time to use during the on-road trials.

- Based on feedback provided by the drivers during the early driver experience assessment that was completed on Highway II, we changed the protocol to include an in-truck instructor/trainer to accompany the drivers. Having a second person in the truck cab worked against a hypo-vigilant driving state. It is well known that passengers affect the social context of a driving scenario, and the outcome due to changes in attention (Chan, et al., 2016). Also, if the second person is evaluating the participant's performance and offering feedback, the impact of their presence is likely greater. We expect that the second person in the truck reduced monotony for the drivers and perhaps increased vigilance and the strategies undertaken by the drivers. We were unable to test this since all the trips had the second instructor on board (no control to compare to). As such, the above points are speculative, but supported in the literature. This will be addressed again in the report conclusion.

- Six student researchers were dedicated to collect data, three in Calgary and three in Edmonton. In Calgary, one student contracted COVID, one was required to work remotely, and the third student had limited schedule availability resulting in reduced coverage in Calgary. An additional individual was sent to Calgary to assist but limited training resulted in data that was not adequately collected and discarded from the analysis. By January 17, the human factors team was at full capacity and able to support full time data collection in Edmonton and Calgary.

- Fitbits were provided to the drivers to collect their sleep history before participation in the study. Drivers were requested to wear their devices, but compliance was poor. Researchers planned to contact each driver two weeks before their participation to remind them of the necessity of wearing the Fitbit. Due to accelerated timelines and operational scheduling this was unfortunately not completed. Consequently, only data from one driver was available. This step is mentioned further below.

- The initial hypothesis of this study was focused on assessing the impact of a platooning system on passive fatigue. This assumed that the platooning system was commercially advanced. As discussed in the interim report, in meetings, and during the on-road trials, limitations and challenges associated with the system maturity affected the performance and workload of the drivers. This effectively negated the possibility of analyzing the impact of passive fatigue while driving in a naturalistic setting. Although not directly measured, it is reasonable to assume the presence of a trainer/observer during the on-road trials further negated the naturalistic driving environment and further reduced the potential to observe the onset of passive fatigue.

- Accomplished on-road trials occurred during the daytime. Nighttime driving was not investigated, mainly due to drivers' scheduling and the long working hours, which did not allow having more than one round trip in 24 hours. Therefore, nighttime driving that increases passive fatigue in drivers could not be investigated.

- Due to safety regulations, the distance between the two trucks was 3s (actually 4s) at the minimum. This quite large distance can be another factor to decrease the chances of drivers getting into passive fatigue.

3 DRIVER DEMOGRAPHICS

Nine male truck drivers participated in the study. The trails included drivers aging from 25 to 66 years of age (mean±std: 44.0±13.1, min: 25 yrs, max: 66 yrs). Commercial driving experience ranged from 4 to 45 years (mean±std: 18. 3± 12.4, min: 4yrs, max: 45 yrs). Based on the inclusion criteria for this study, only drivers with a minimum of two years commercial driving experience were considered for the study, and all drivers are above this threshold.

4 ANALYSIS PROCESS

In the proposal, two main approaches to analyze data were outlined: 1) Hypothesis driven data analysis and 2) multi-modal data analysis. Although passive fatigue is no longer the focus of this study, driver stress, workload levels, and the impact of the new technology could still be investigated. In this chapter, an overview of these two approaches and how they were used in this study to analyze the different modalities are discussed.

4.1 Hypothesis Driven Data Analysis

Based on the original methodology, the following eight hypotheses (the first six hypotheses are shown in Table 1) were defined to investigate the effect of platooning on drivers in various conditions.

Hypothesis 1: Driving in the ADS engaged system will decrease vigilance and increase passive fatigue compared to driving in ADS disengaged.
This hypothesis was not investigated due to the low number of participants in baseline and copilot trials.

Hypothesis 2: While ADS is engaged in both the lead and follower truck in CTPS, the driver in the lead truck will experience increased vigilance and decreased passive fatigue compared to driving in the single truck with ADS engaged.
This hypothesis was not investigated due to the low number of participants in copilot trials.

Hypothesis 3: While ADS is engaged in both the lead and follower truck in CTPS, the driver in the follower truck will experience decreased vigilance and increased passive fatigue compared to driving in the single truck with ADS engaged.
This hypothesis was not investigated due to the low number of participants in copilot trials.

Hypothesis 4: While ADS is engaged in both lead and follower trucks, the driver in the follower truck will experience decreased vigilance and increased passive fatigue compared to the lead truck driver.
When enough data was available, this hypothesis was investigated.

Hypothesis 5: Longer distances between platooning vehicles will decrease vigilance and increase passive fatigue compared to driving in shorter platooning distances for the lead driver.
This hypothesis was investigated based on initial data. The collected data was not balanced and after adjusting the Trip number (whether it was performed in Trip 1 or Trip 2), there was not enough data to comparing this hypothesis.

Hypothesis 6: Shorter distances between the platooning vehicles will increase passive fatigue and decrease vigilance compared to driving with longer platooning distances for the follower driver.
This hypothesis was investigated based on initial data. However, as the collected data was not balanced, after adjusting Trip number (whether it was performed in Trip 1 or Trip 2), there was not enough data for comparing this hypothesis.

Hypothesis 7: Drivers will demonstrate increased levels of trust and user acceptance as they gain experience driving in CTPS with the ADS engaged.
A comparison of the drivers' very first and last exposure to platooning was used to investigate this hypothesis. Due to unbalanced data, some drivers only participated in one CTPS trip and some participating in as many as seven trips.

Hypothesis 8: Drivers will demonstrate increased levels of fatigue in their second trip of day (Trip 2) compared to their first trip (Trip 1).
There was sufficient data in some measures to investigate this hypothesis.

We have adopted an exploratory approach to this study as participant numbers yielded an underpowered and unbalanced trial. This forced us to abandon the statistical tests and significance levels typically associated with human behavioral research. In several instances, we sought to explore observations that were of practical significance even if they did not reach statistical significance. Depending on the number of drivers in each group, one-sample or two-sample tests were implemented to check for differences between conditions. For each test, a normality test (Shapiro test) was performed, and depending on the result of the normality test, a t-test or a non-parametric t-test (Wil-Cox test) was used. As passive fatigue was no longer the focus, a two-tailed test was done to look for differences in either direction (increase or decrease). Due to the exploratory nature of this study, the results with p-val<.1 were also reported as trend-level effects. Importantly, no adjustment for the multiple-comparison problem was performed.

4.2 Multi-Modal Data Analysis

In the project trip segment, over three hours of continuous data was collected from various physiological sources associated with stress and fatigue, including brain activity based on the EEG. Different aspects of driving performance such as speed, steering wheel behavior, accelerations and braking, lane drifting, and interaction with other road users such as cut ins, were being collected simultaneously. This multi-modal data set provides a unique opportunity to have a full picture of how driving in platooning with an ADS engaged system impacts the brain activity, physiological signals, and driving behavior, and how these aspects interact with each other.

Having unbalanced data imposes challenges for systematic multi-modal data analysis. Therefore, we restricted the analysis for this approach to a few cases rather than a complete multi-modal analysis as given in Chapter 8.

5 QUESTIONNAIRES

5.1 Overview

Various questionnaires were used to collect personal opinions on the drivers' state of alertness, driving behaviors, and overall driver experience with the platooning system. They have been categorized as non-platooning related questionnaires and platooning related questionnaires.

- Modified Risk Perception Questionnaire (MRPQ): The MPRQ has been used in previous naturalistic driving studies (Antin, et al., 2011; Dingus et al., 2015) to assess both the risk perception of truck drivers and their willingness to engage in risky behaviors.

- Driver Stress Inventory (DSI): The DSI utilizes five scales (Aggression, Dislike of Driving, Fatigue Proneness, Hazard Monitoring, and Thrill Seeking) to assess specific personality traits that may control vulnerability to stress and fatigue.

- Karolinska Sleepiness Scale (KSS): Drivers provided subjective judgments on their level of sleepiness/alertness using the Karolinska Sleepiness Scale (KSS). The KSS is a nine-point Likert scale designed to measure intrusions into what should be periods of wakefulness.

- Driving Activity Load Index (DALI): DALI is a revised version of the NASA-TLX, adapted to the driving task. It has been validated to identify workload origins and evaluate mental workload while using in-vehicle systems (Pauzié, 2008; Pauzié & Manzano, 2007; Kim & Ji, 2013; Zakerian et al., 2018) It utilizes 6 subscales; 1) Effort of attention; 2) Visual demand; 3) Auditory demand; 4) Temporal pressure; 5) Interference; 6) Situation stress.

- Trust in Automation (TiA): TiA assesses and accounts for the effects of trust in Level II SAE technology automated driver assist (ADS) operator behavior (Körber et al., 2018). It has five sub-scales including Familiarity, Intention of Developers, Reliability/competence, Trust in Automation, Understanding/predictability. The professional drivers in this pilot study completed the TiA as part of their pre and post trip questionnaires.

- Acceptance Scale for Advanced Transport Telematics (AATT): To assess and account for effects of user acceptance on trust of Level 2 SAE technology automated driver assist operator behavior, participants completed an Acceptance Scale for Advanced Transport Telematics (AATT) (Van der Laan et al., 1996). AATT has two categories, a Usefulness and a Satisfactory scale.

- Equipment and Comfort (E&C): Drivers' subjective opinion on the vehicle equipment and its comfort compared to their normal operating equipment (regular truck) was collected using a questionnaire designed specifically for this project.

The MRPQ and the DSI were completed by the drivers during the Introduction session, prior to any truck familiarization.

The Karolinska Sleepiness Scale (KSS) was administered during the Pre Trip session. During Post Trip 1 sessions, both the KSS and DALI scores were collected. Finally, driver responses to DALI, E&C, AATT, and TiA were captured in Post Trip 2 sessions. Table 5 demonstrates the data collection process for the Introduction, Pre Trip, Post Trip 1 and Post Trip 2 sessions. Questionnaire data were collected successfully in 62% of trips, 79% starting Jan 17, when full researcher coverage was available.

Table 5: Questionnaire Administration

Questionnaire	Introduction Session	Pre Trip Session	Post Trip 1 Session	Post Trip 2 Session
Modified Risk Perception Questionnaire (MRPQ)	☑			
Driver Stress Inventory (DSI)	☑			
Karolinska Sleepiness Scale (KSS)		☑	☑	
Driving Activity Load Index (DALI)			☑	☑
Trust in Automation (TiA)				☑
Acceptance Scale for Advanced Transport Telematics (AATT)				☑
Equipment and Comfort (E&C)				☑

A summary of key driver responses for each questionnaire can be found in Appendix D.

5.2 Analysis of Non-Platooning Activities

During the Introduction session, each driver completed the MRPQ and DSI. These questionnaires assess underlying factors that can impact the driving experience.

5.2.1 Modified Risk Perception Questionnaire (MRPQ)

To assess both the risk perception of truck drivers and their frequency of risky behaviors (willingness to engage), the Modified Risk Perception Questionnaire (MRPQ), customized to accommodate for Canadian factors, was issued to the drivers during the introduction session. The MRPQ was devised by combining the Cox Assessment of Risk Driving Scale (CARDS) and the DeJoy Risk Perception Questionnaire (RPQ) (Antin et al., 2011) and has been utilized in two naturalistic driving projects (Antin et al., 2011; Dingus et al., 2015).

The risk perception battery contains 33 items and researchers have used it "to gauge the participant's perception of dangerous or unsafe driving behaviors or scenarios and…to gauge the frequency and a participant's willingness to engage in dangerous, unsafe, or risky behaviors," (Dingus et al., 2015). It questions a drivers' propensity (0=never to 3=often) in the previous 12 months to participate in various driving behaviors (e.g., run a red light, drive when sleepy or falling asleep, or change lanes suddenly to get ahead in traffic or speed for the thrill of it).

Researchers in this pilot study wanted to explore driver risk perception as part of the demographic profile before their driving experience with the platooning system, and to explore their general tendency towards a lower or higher risk perception, as it affects their acceptance of, and trust in ADS. In 1990, Tränkle et al. noted young male drivers rated traffic situations as less risky than did older male drivers. Before them, several studies found younger male drivers demonstrated a higher propensity to be involved in a high-speed accident, incur accidents due to loss of control of the vehicle, run red lights, and speed in times of low illumination, on rural roads, and around curves (Ellinghaus & Schlag, 1984; Jonah, 1986; Knoflacher, 1979; Michiels & Schneider, 1984; Schlag, 1987; *Statistisches Bundesamt*, 1987).

More recently, Liu et al., (2021) found that older drivers (between 45-55, specifically) scored highest in risk perception and in driving performance or skill. Thus, the age range (25 to 66 years of age) and average age (44.0±13.1) of the drivers in this pilot study may largely explain their low risk attitude (avg. 16.89).

However, in 2002, Weber et al, explored risk attitudes and risk taking in various domains (financial, health and safety, gambling, ethics, and social decisions) and demonstrated that risk taking is domain specific. In other words, risk taking depends on the activity itself, and thus Weber et al. (2002) thus introduced a dimension of relativity to the concept of risk perception and risk taking. Content domain differences in risk taking were based on differences on the perceived cost versus benefit relationship/ratio of the specific activity rather than global perceived risk attitudes. Sitkin and Weingart (1995) dubbed this an "interactional model of risk-taking, in which situational characteristics as well as person-centered characteristics jointly influence risk-taking" (Weber et al., 2002). Situational factors influence risk attitude and acceptance as well.

Bell et al., (2018) noted that in a comparison of conventional driving vs connected/automated driving that risk perceptions were significantly lower in the former and demonstrated a reduction in risk perception. Other researchers, though, have suggested that as the novelty of the automated driving experience lessened, the perception of cost versus benefit may have also influenced the acceptance of ADS technology (Brell, et al., 2019). Thus, how drivers perceive conventional driving, which is where they have the most experience, can shape their responses about risk attitude towards ADS until they have more experience with driving ADS. Meaning, we may see different results with more experience in ADS. At this time and in this pilot study, drivers' general risk attitudes in relation to driving with ADS may not be completely reflected in their MRPQ responses.

Figure 4 identifies the four highest-ranking risky behaviors based on driver responses to the Modified Risk Perception Questionnaire (MRPQ). Altogether, drivers failed to generate an average score above a 1.31 (1 = rarely, 2 = sometimes) reflecting a low to moderate risk attitude. Distracted behaviors associated with doing other things while driving is the most frequently cited risky behavior and may be associated with the monotony of driving. Still, it is evident that overall risky behaviors are relatively low in terms of frequency of occurrence, with the remaining 29 behaviors demonstrating even lower frequencies. This may suggest that these drivers would experience lower stress responses, fewer distractions, and take a more cautious approach during adaptation to a platooning system until such time that the novelty wears off, when the drivers may look for more distractions to stay engaged with the driving task.

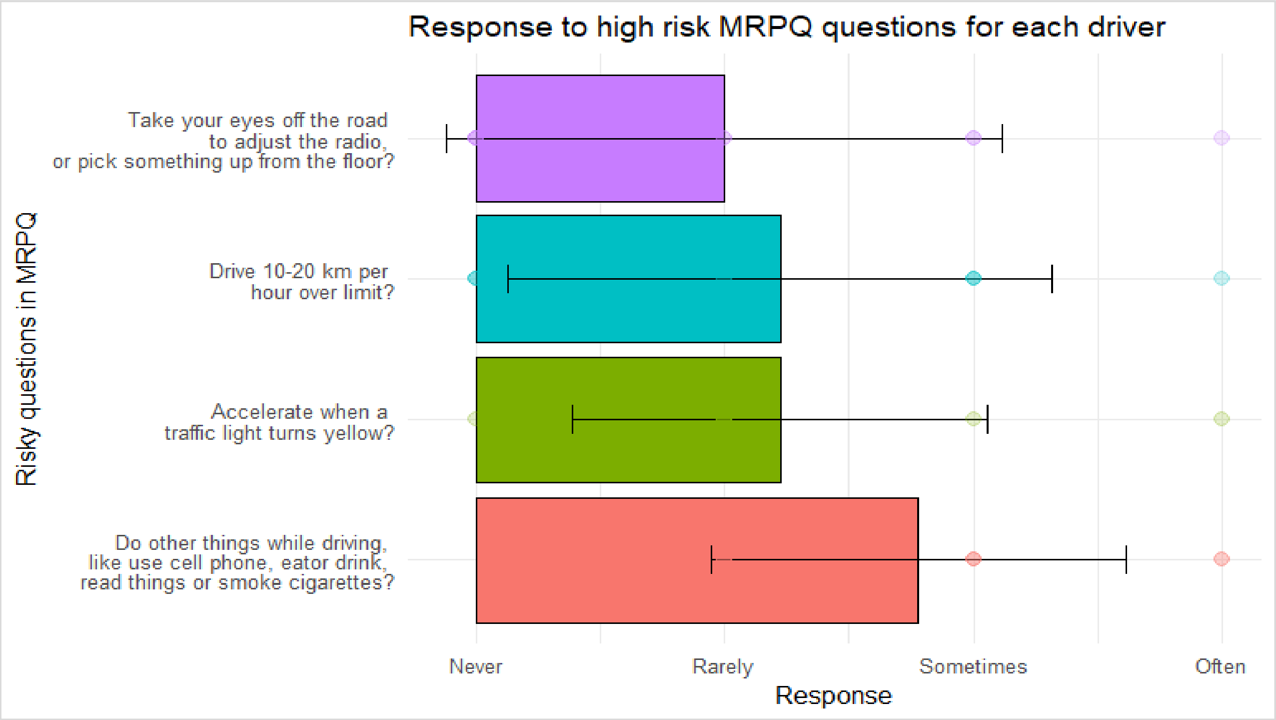


Figure 4: Average of the drivers' response to the top four identified risk behaviors in MRPQ. The group standard deviation from the mean is shown as an error bar

5.2.2 Driver Stress Inventory (DSI)

The DSI was chosen to compare driver traits against physiological data that was to be captured during the platooning activities and the various study conditions as outlined in the original hypotheses (Qu et al., 2016). Due to the changes in the study design, the resulting analysis of the DSI was limited to providing an overview of personal driver traits that were identified with no comparison to specific driving conditions.

The Driver Stress Inventory (DSI) utilizes five scales to assess specific personality traits that may control vulnerability to stress and fatigue: Aggression, Dislike of Driving, Fatigue Proneness, Hazard Monitoring, and Thrill Seeking. It utilizes a 10-point rating scale for each question.

Table 6 demonstrates how the categories are constructed.

Table 6: DSI Categorization Breakdown

Trait	Number of Questions	Maximum Score
Aggression	12 Questions	120
Dislike of Driving	12 Questions	120
Hazard Monitoring	8 Questions	80
Thrill Seeking	8 Questions	80
Fatigue Proneness	8 Questions	80

Unfortunately, only the first four categories were captured due to an entry error related to the Fatigue Proneness questions and therefore, this fifth category was not part of this analysis. However, this proved of minor consequence as fatigue was no longer the focus of this study. This adjusted the maximum number of questions to 40.

Additionally, a second coding error allowed Dr06 to skip over 25 of the 40 questions and was subsequently removed from this analysis.

Studies have demonstrated that drivers under stress will demonstrate performance impacts, both negative (increased risky behavior) and positive (decreased risky behavior).

Driver stress is defined as the process of facing a situation where the perceived demand, mostly defined based on previous experiences, internal body sensations, and external stimuli, is higher than the available resources (Tavakoli, et al., 2022). In a comparison study of automated driving vs non-automated, automated drivers demonstrated slower response times to sudden traffic events, recorded higher levels of distress, and higher predictions of being prone to fatigue (Neubauer et al., 2010).

Larger vehicles, like trucks, and shorter distances with a higher standard deviation in the distance to the lead vehicle are associated with higher heart rate and more negative emotions, indicating higher stress. Additionally, highway driving at higher speeds generally produces less stress-like indicators (Tavakoli et al., 2022).

In a study (Qu et al 2016) using the DSI, it was noted that those drivers who reported traffic violations (e.g., speeding, ignoring traffic signs or markings) within the previous three years correspondingly recorded higher aggression and thrill-seeking behaviors and lower hazard monitoring behaviors compared to those with no reported violations. In comparison to minor accidents, only hazard monitoring and fatigue showed weak correlations. No drivers within this study reported having a minor or major accident, nor a convicted moving violation within the previous three years.

Average DSI scores for the drivers in each individual category are presented in Figure 5. Higher scores reflect greater levels of the categorized trait.

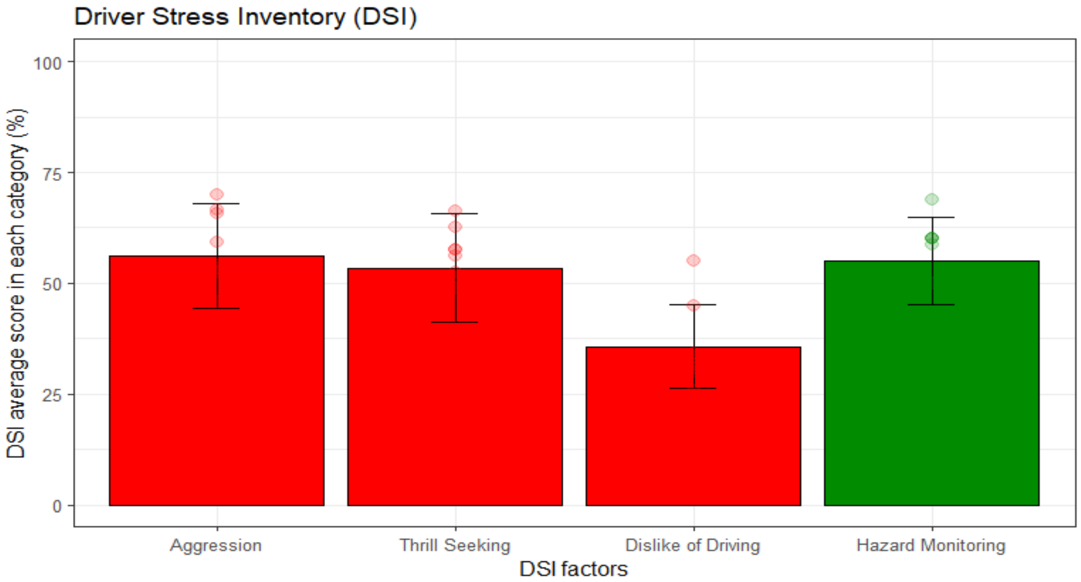


Figure 5: Average of driver response in each category. Standard deviation is shown using error bars.

The highest ranking scale on the DSI by the drivers was the Aggression scale. Aggressive driving can be a result of personal factors unrelated to driving, or as a reactive consequence to driving conditions such as incorporating a confrontational approach to risky behaviors of other drivers (e.g., tailgating, cut ins, etc.). High Aggression scores have been correlated to more frequent, and deliberate, road safety violations such as speeding and risk overtaking maneuvers (Matthews, et al., 1997).

It is theorized that the ability to monitor for hazards is directly impacted by the alertness of an individual to their surroundings (situational awareness). A higher score within the Hazard Monitoring scale suggests an increased detection of hazards or changing conditions. However, one study suggested that when factoring in age considerations, older drivers did not fare as well at hazard detection compared to younger drivers, but it was theorized that this was balanced by the older drivers' better alertness levels (Matthews et al., 1998). An alternative explanation may be that older drivers, based on years of experience, are more capable at discerning between the level of risk associated with various hazards and therefore, were able to focus their attention on other aspects of the driving task. Regardless, higher hazard monitoring reflects a more cautious approach to driving and may correspond to a higher stress response. While it is interesting that the drivers in this study had similar ranking in hazard monitoring and aggressive driving, it is not possible to derive inferences from this data due to the low number of drivers.

Thrill seeking behaviors differ from aggressive behaviors in that the former enjoy risk taking behaviors, while aggressive drivers report higher depression following such actions (Zuckerman, 1979). Thrill seeking behaviors were less evident amongst the drivers participating in this study in comparison to the other DSI scales. This is a preferred behavior for a professional or commercial driver and is likely a component of their extensive training and influenced by the extent of their driving experience, more so than the general public.

High Dislike scores have demonstrated the strongest correlation to higher stress responses (Matthews, 1993; Matthews, et al., 1991) and a subsequent lack of self-confidence and perceived control over the driving task, which has in turn been shown to increase situational stress and worry while driving in traffic (Matthews, 1993; Matthews et al., 1997). Specifically, high Dislike scores reflect more variability with regard to lateral positioning during both follower and open road driving, and an increase in operational control errors (Matthews et al., 1998). Conversely, a decrease in the number of overtaking occurrences may correspond to a reduced risk by demonstrating more cautious behavior (Matthews et al., 1997). The driver scores here are relatively low, which could likely be attributed to the fact that these are individuals who chose driving as a career path, therefore less likely to not like the driving aspect, as well as being highly trained in the act of driving. Again, their overall exposure level to driving is also much greater than the average person in a non-driving profession, which is likely to correspond to a greater sense of self-confidence and control.

Interestingly, a large naturalistic driving study on driver behavior and risk of crash occurrence did not find any correlation between an increase in crash risk and any of the scales in the DSI (Klauer, et al., 2006). However, research in 2010 identified links between accidents and three of the scales; Aggression, Dislike of Driving, and Hazard Monitoring (Öz et al., 2010).

Figure 6 provides a composite DSI risk profile by combining all four scales of an individual driver. To calculate the risk profile score, the hazard monitoring score was subtracted from the sum of the three remaining scales. Total score values can range from a minimum of 0 to a maximum of 320. Actual scores ranged from a low of 67 to a high of 139. The risk profile suggests that DR05 takes a more cautious approach to driving compared to the other drivers, while DR03 appears to have the highest risk profile but remained below the 50th percentile, as did all the drivers, which could reflect the nature of being a professional driver as opposed to the general population of adult drivers.

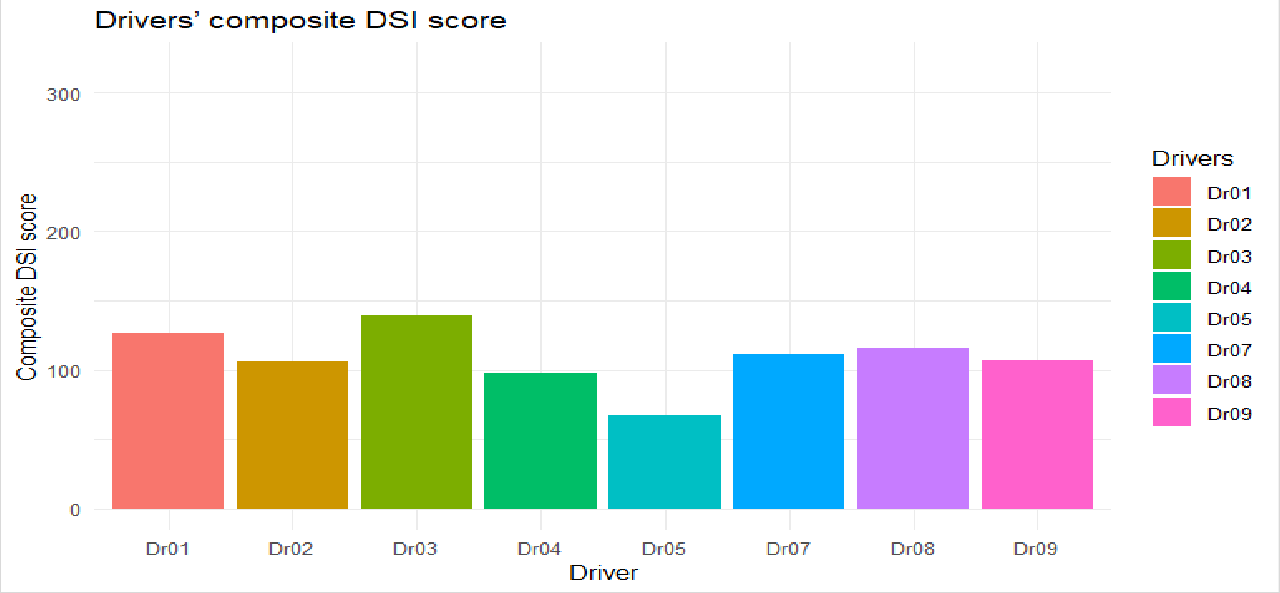


Figure 6: Drivers' composite DSI score

5.3 Analysis of Platooning Activities

Platooning activities were analyzed using five different questionnaires/scales; the KSS, DALI, TiA, AATT, and E&C.

KSS data was collected at Pre Trip and Post Trip 1 to determine alertness states at the beginning of the trip. Post Trip data was limited as there were several factors impacting the collection of the data including trucks requiring maintenance, drivers being behind schedule and the truck convoy protest.

DALI was collected at Post Trip 1 and Post Trip 2. To assess the platooning workload on drivers, DALI responses were compared across different conditions (lead vs. follower truck, platooning vs. copilot, etc.) if sufficient samples were available.

TiA and AATT were collected at Post Trip 2. It should be noted that drivers' participation was very imbalanced due to operational scheduling difficulties. Some drivers completed only two platooning sessions while others had four. First and last session responses were compared with each other to see if there were any notable changes over time due to increased exposure to the system.

5.3.1 Karolinska Sleepiness Scale (KSS)

A score of 7 or greater indicates high levels of drowsiness and elevated risk for the onset of a microsleep. Scores between 1-3 correlate to strong levels of alertness. Higher scores during the Pre Trip assessment could indicate a poor or deprived sleep period the night before. If higher scores were recorded during Post Trip 1, it could reflect circadian factors as they are heading into a natural circadian low period in the afternoon, or it could reflect the gradual onset of fatigue. Without physiological data to compare against, it is not possible to make any further determination.

In both the Pre Trip and Post Trip 1 sessions (Figure 7, Figure 8), seven of the nine drivers did not score higher than 3 on any of their trips, indicating high alertness levels, and suggesting that drowsiness was not affecting the drivers to any significance at those times. However, Dr04 scored a 7 in one Pre-Trip evaluation. This was an outlier score, as his three other Pre Trip scores registered high alertness (3 or less). When comparing against other personal data collected during the initial survey, this driver reported very good subjective sleep data (good sleep latency value, sufficient quality and quantity of sleep, consistent schedules), had a "Robin" chronotype (neither early morning or night orientation), and had a low Epworth Sleepiness (ESS) score. This driver did have an elevated screening score for Obstructive Sleep Apnea (OSA), but based on a clinical evaluation, had been medically cleared. Therefore, it is likely that on this particular trip he had a poor sleep the night before.

Dr09 had consistently higher scores in comparison to all other drivers and was responsible for all four of the remaining elevated scores (4-6) during the Pre Trip sessions but still remained below the threshold of concern (7 or higher). This driver also had the only elevated Post Drive scores. When comparing against other personal data collected during the initial survey, this driver reported very good subjective sleep data (good sleep latency value, sufficient quality and quantity of sleep, consistent schedules), had a morning chronotype, and low ESS and OSA scores. However, this driver's normal departure time begins between 1300 and 1630, which could explain the higher KSS values as he was required to have a much earlier arrival and departure time for this study (8-8:30 AM).

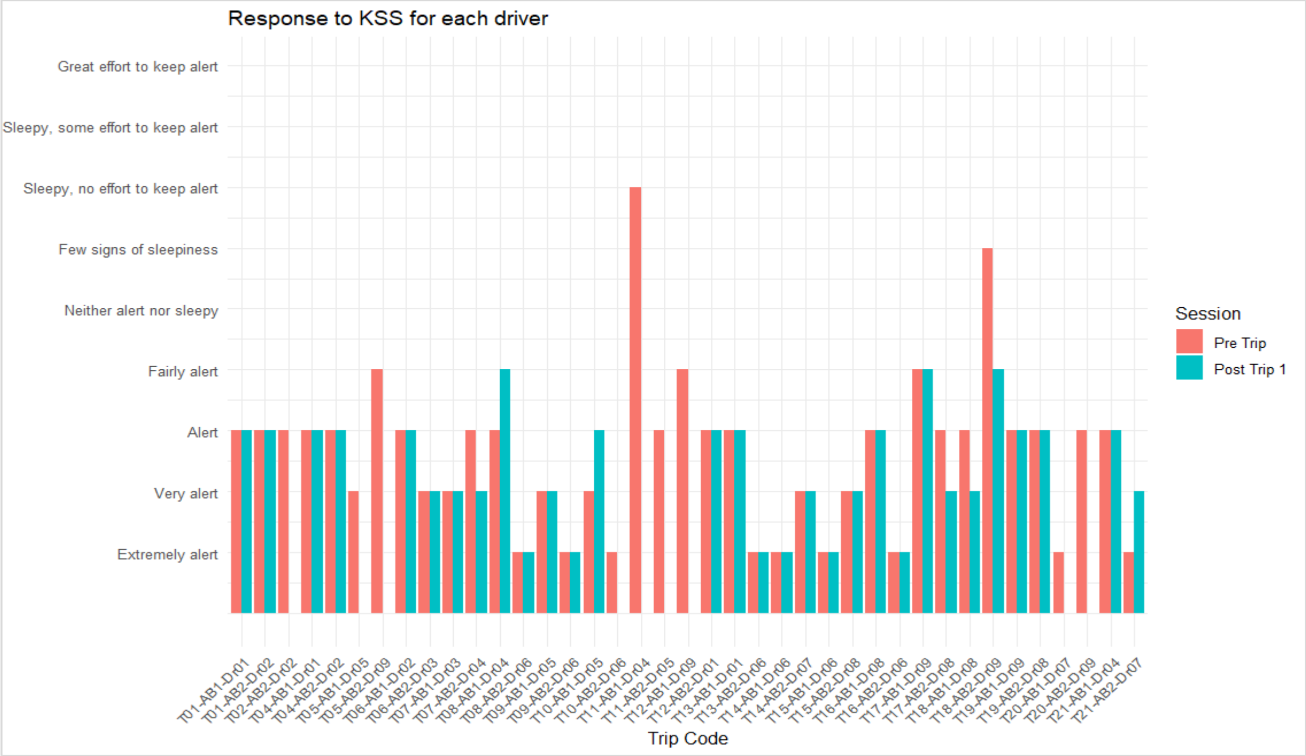


Figure 7: Drivers' responses to KSS questionnaire in Pre Test and Post Trip Sessions

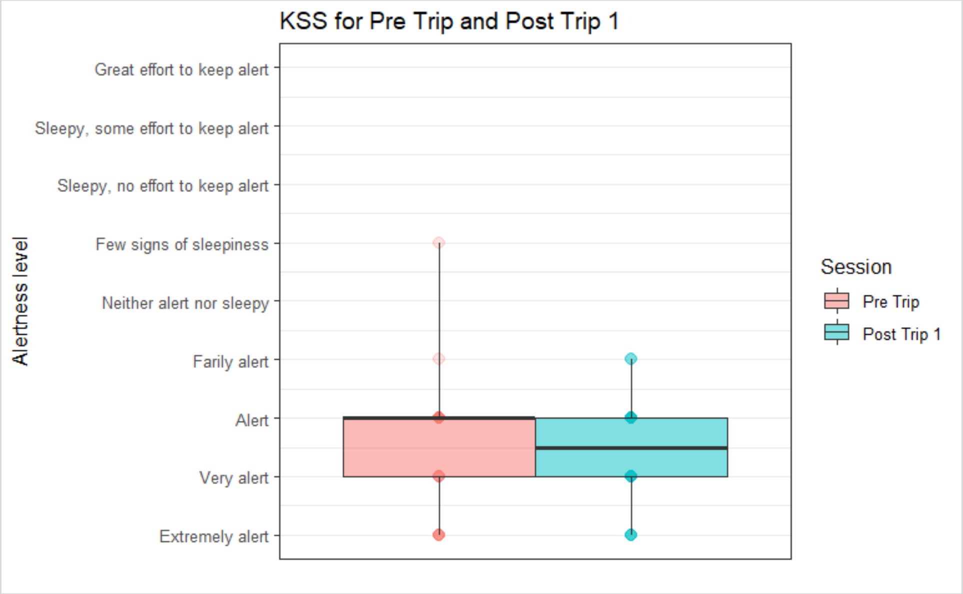


Figure 8: Alertness level during Pre Trip and Post Trip 1 sessions

5.3.2 Driving Activity Load Index (DALI)

In this section we explore drivers' DALI response after platooning Post Trip 1 and Post Trip 2 in support or refutation of Hypothesis 7. Based on the normality test (Shapiro test) data is normally distributed

(Figure 9). The t-test shows that distraction is significantly less in Trip 1 compared to Trip 2 (df =5, p-val = .042, diff = -0.83, t-test = -2.71) (Figure 10). No other subscales were found significant between the two trips.

The higher distraction scores (i.e., the level of disturbance or attention devoted to a non-driving task)—noted in Trip 2 vs Trip 1 (Figure 10) may be due to the familiarity and comfort with the Level II automated driving system. After one trip the professional drivers in this study, who typically deal with very dynamic situations on a regular basis, no longer react as intensely to the novelty of Level II automation and platooning and/or to integrate more tasks into their duties/checklist (Pauzié, 2008). If, shortly after engagement with the Level II automation and platooning, the professional truck drivers in this study perceived the driving task as less demanding, they may have sought novelty or experienced more mind wandering (Galéra et al., 2012). Similarly, reduced attention even after a short time working with an automated system was demonstrated in Remotely Piloted Aircraft (RPA) or Unmanned Aerial Vehicles (UAS) operators, who, when monitoring an automated system, consistently demonstrated vigilance tasks decrement at about an hour or as early as 20-35 min after they engaged with the system (Cummings, et al., 2013; Thompson et al., 2006).

Conversely, after initial familiarization and depending on their trust in the system, the truck drivers in this pilot study may have engaged with and become pre-occupied by the platooning system to the point of distraction. Furthermore, individual drivers will vary greatly in their distractibility, mind wandering, ability to focus and in their sensation seeking (Pauzié, 2008; Thiffault & Bergeron, 2003). Comparing DALI with eye tracking and EEG data as well as risk taking and sensation seeking personality traits may elucidate the reasons behind the significant change in driver distraction from Trip 1 to Trip 2. Galéra et al., (2012) noted that eye behavior in the form of increased eye blinking, less complex eye movements, and less pupil diameter responsiveness to task related events are prevalent during mind wandering. Finally, a fatigued brain is more susceptible to distraction, and we could not conclusively ascertain driver fatigue state in this pilot study. Even though both overload and underload due to driving task-related/relevant (instrument monitoring, and/or operating communication equipment) or irrelevant/non-task related activities (traffic, cell phone use, mind wandering and/or presence of an observer) may contribute to driver distraction. We do not have sufficient data to concluded why DALI distraction scores were significantly less for Trip 1 than Trip 2 (Catherine et al., 2012).

Therefore, without objective data to corroborate or verify, we cannot identify the nature of the distraction the drivers experienced nor attribute it to other than expected decline in directed attention (Cummings et al., 2013).

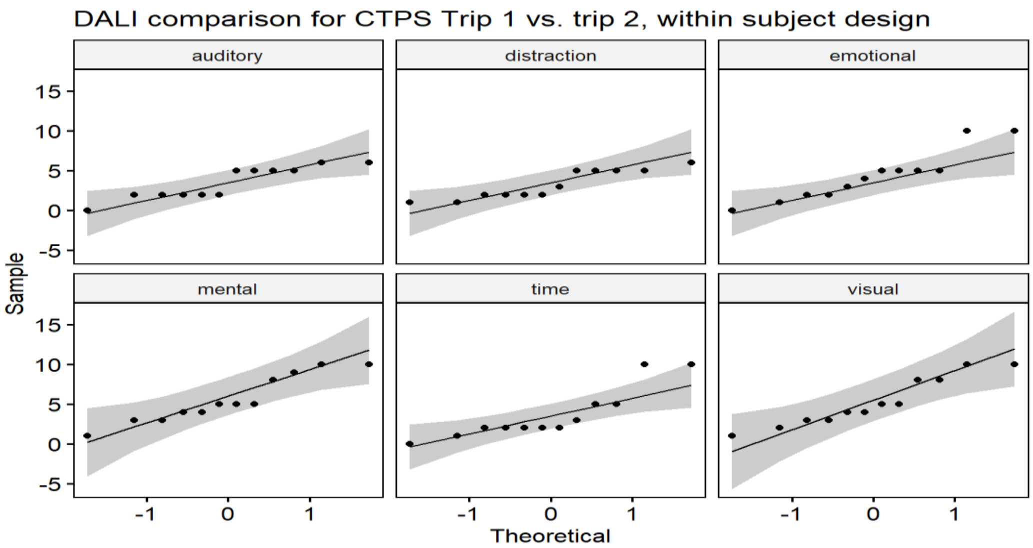


Figure 9: QQ-Plots for DALI. Distraction (p-val > .05) follows a normal distribution based on the normality test (Shapiro test).

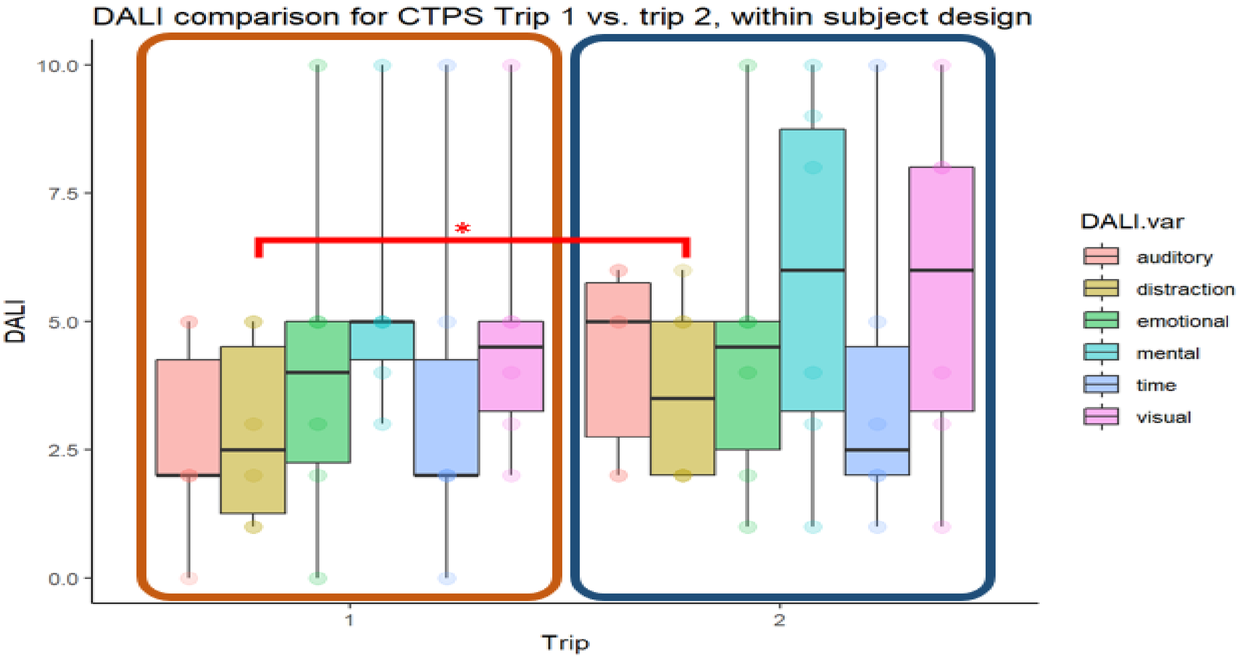


Figure 10: DALI Comparison for CTPS Trip 1 vs Trip 2. Box-plots for DALI. Distraction is significantly less in Trip 1 compared to Trip 2 (p-val = .042).

5.3.3 Trust in Automation (TiA)

Based on the normality test (Shapiro test) Familiarity data is normally distributed (Figure 11). More exposure to the platooning system is expected to increase the familiarity to the system. Therefore, a one-tailed t-test was performed to compare the first vs last session. The t-test shows that there is a trend that Familiarity is less in the First session compared to the Last session (df = 4, p-val = .052, diff = -1.8, t-test=-2.09) (Figure 12). The non-parametric t-test (Wilcox test) shows that there is a trend that Understanding/Predictability is greater in the First session compared to the Last session (df = 4, p-val = .089, diff = 1.4, V = 10) (Figure 12).

The drivers' overall responses show greater variability and higher overall scores for the Reliability & Competence subscale ("The system is capable of interpreting situations correctly," "The system works reliably," "The system is capable of taking over complicated tasks," etc.) (Figure 13). Drivers understand the benefits of technological assistance such as ADS in general but their level of trust in the reliability and competence of the system depends on their own expectations of and experience with automated driving. Furthermore, drivers' responses for Understanding/Predictability (Figure 12), may at first appear paradoxical or counterintuitive, but could simply reflect the effect of drivers' expectations of the Level 2 Platooning ADS vs their experience with the same; whereas, Familiarity ("I already know similar systems," "I have already used similar systems") forms expectations, influences trust and is itself influenced by experience with similar systems (Körber et al., 2018). Thus, familiarity with this system would naturally increase with time spent using or interacting with the system, assuming no unplanned events occurred outside of normal driving parameters.

It is theorized that if participants complete the TiA prior to first learning about the ADS it can shape their reported levels of trust (Körber et al., 2018). In a 2015 study, the items of the TiA related to Understanding/Predictability measures ("The system state was always clear to me," "--The system reacts unpredictably," "I was able to understand why things happened" and "It's difficult to identify what the system will do next") were answered before participants drove with the system engaged based off information received in the introduction session and the in-cab familiarization without the on-road driving experience (Körber et al., 2015). In particular, participants displayed higher levels of trust in the system, demonstrated decreased monitoring and reflected higher engagement with a non-driving task, in other words, a distraction (Körber et al., 2018). The Post Trip decrease in the Understanding/Predictability score may accurately reflect the effect of the practical driving experience.

In this pilot study, the TiA was administered Post Trip 1, after the drivers had become familiar with the trucks and the ADS. However, a downward trend was noted when comparing first and last trip sessions, suggesting that their initial understanding/predictability declined over time. Further exploration of this trend against telematics and eye behavior data, as well as a comparison with drivers' acceptance of the automation scores, would provide more insight, but only based on meeting the necessary power analysis criteria.

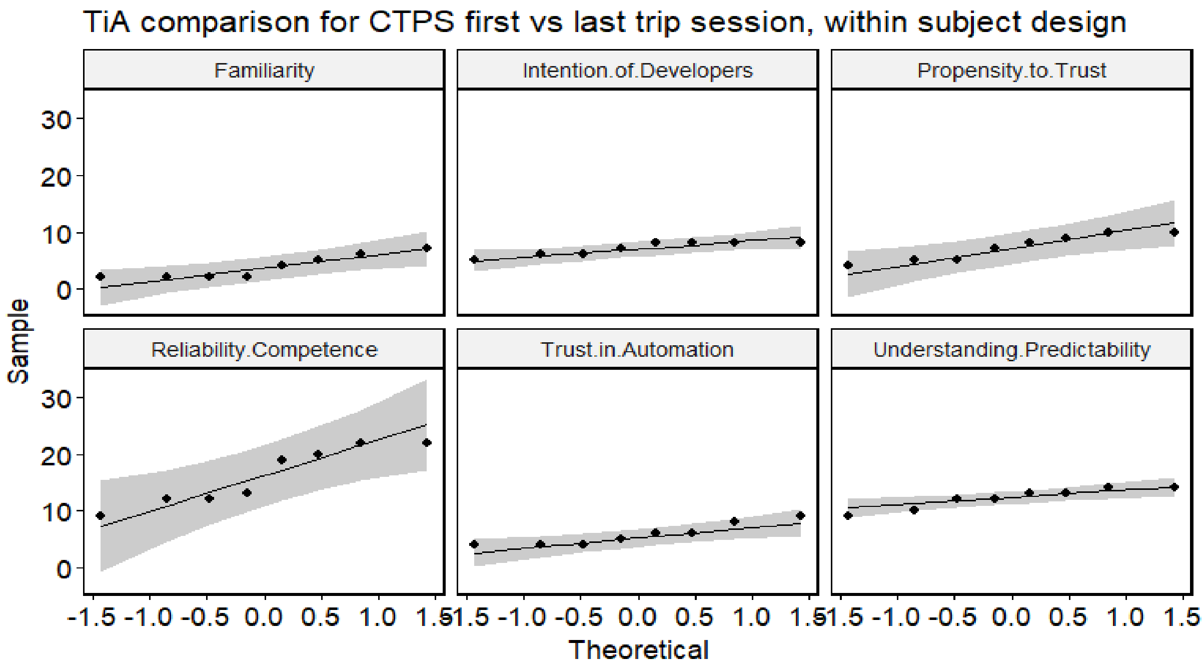


Figure 11: QQ-Plots for TiA. Understanding/Predictability (p-val < .05) deviate from a normal distribution based on the normality test (Shapiro test).

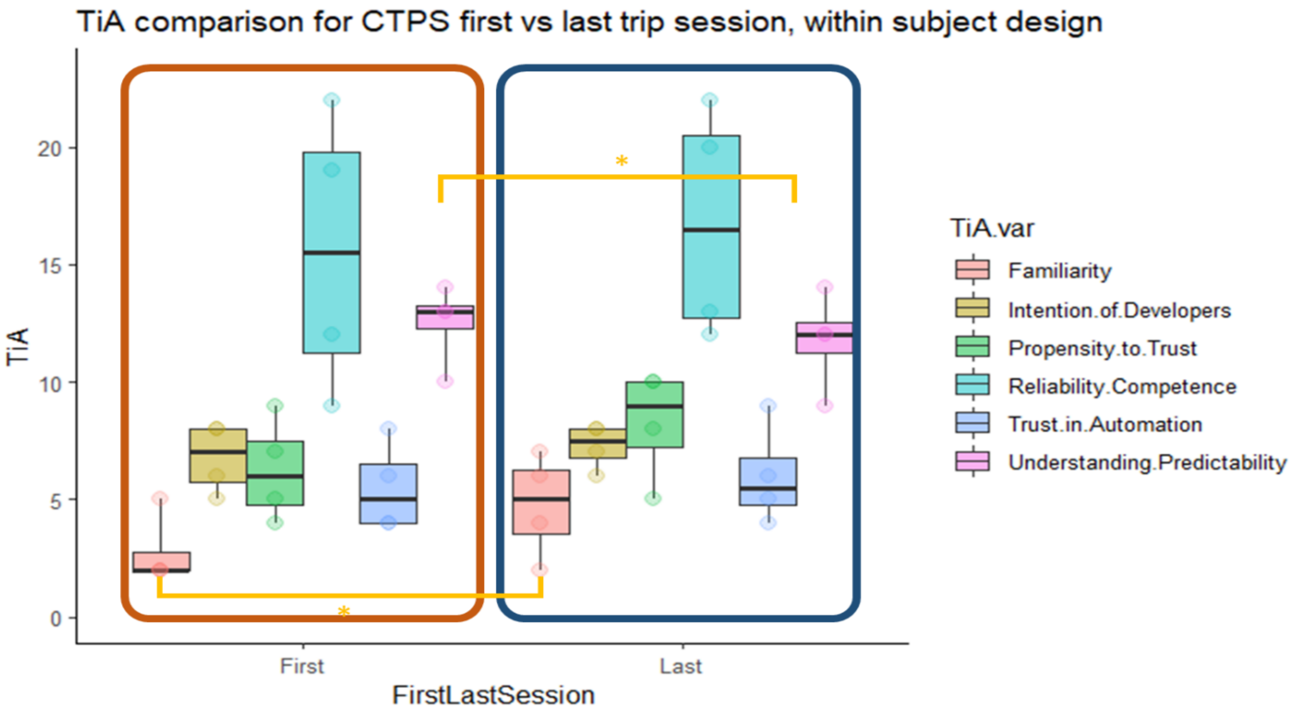


Figure 12: TiA Comparison for CTPS First vs Last Trip Session. Familiarity is less in the First session compared to the Last session (p-val = .052), on-tailed. Understanding/Predictability is greater in the First session compared to the Last session (p-val = .089).

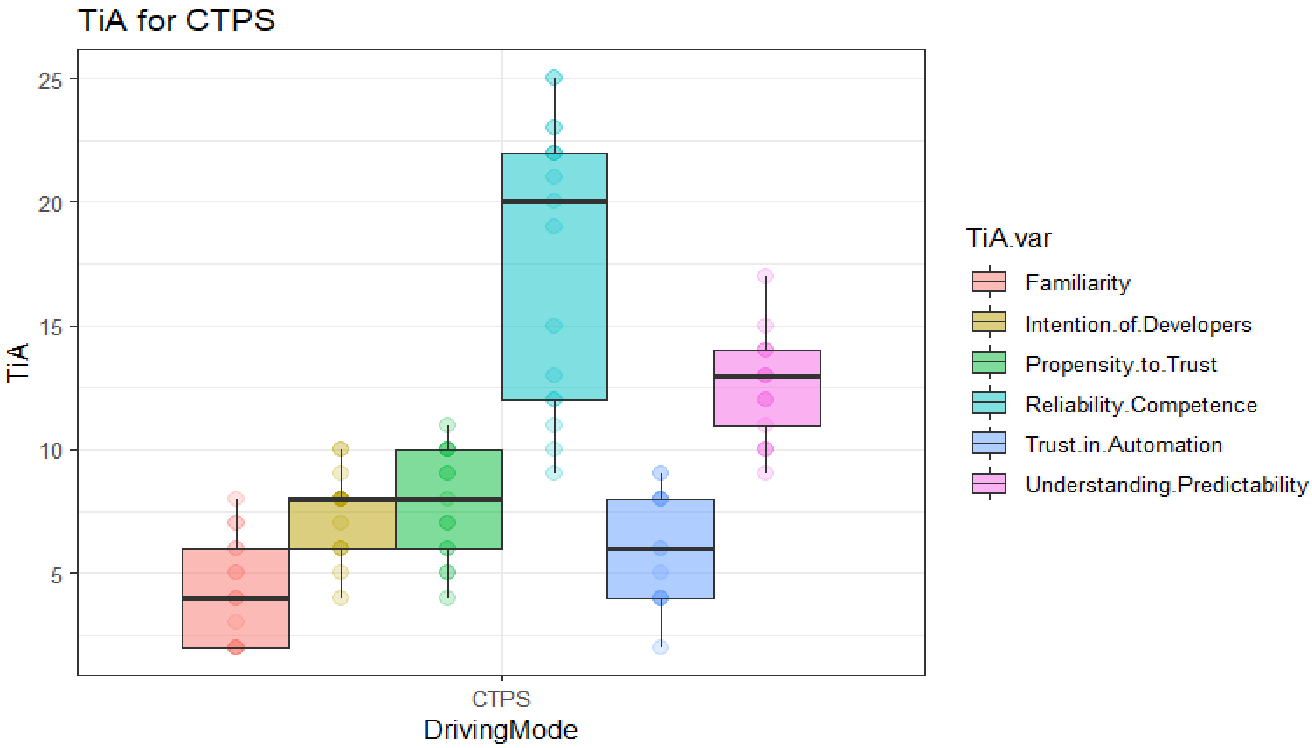


Figure 13: Trust in Automation for all drivers and all sessions

5.3.4 Acceptance Scale for Advanced Transport Telematics (AATT)

Acceptance Scale for Advanced Transport Telematics (AATT) measures driver acceptance of the platooning and automation system using a 9 item Likert scale. Van der Laan's Simple Scale (Van der Laan et al., 1996) contains two subscales, a Usefulness (useful, good, effective, assisting and raising alertness) and a Satisfactory (pleasant, nice, likable and desirable) scale.

There was no change reflected in these scales from the first to the last session (Figure 14). However, for both trucks, driver responses favored Usefulness (Figure 15). This is in line with previous research. Zoellick et al. (2021) found that in 77% of cases (97 studies across research topics and N = 4,095 participants), driver responses to the Usefulness scale were higher than the Satisfactory scale, similar to the results in this pilot study. Furthermore, in a study of 1054 participants, of whom 77.3% had previously been in an autonomous vehicle, but of whom only 46% had high experience with it, 71% considered autonomous vehicles as beneficial or useful (Ayoub et al., 2021). Zoellick, et al. (2021) propose this may be a bias or "method effect" of the Simple Scale itself. Overall, they found the Simple Scale and its two subscales to be reliable (Zoellick et al., 2021).

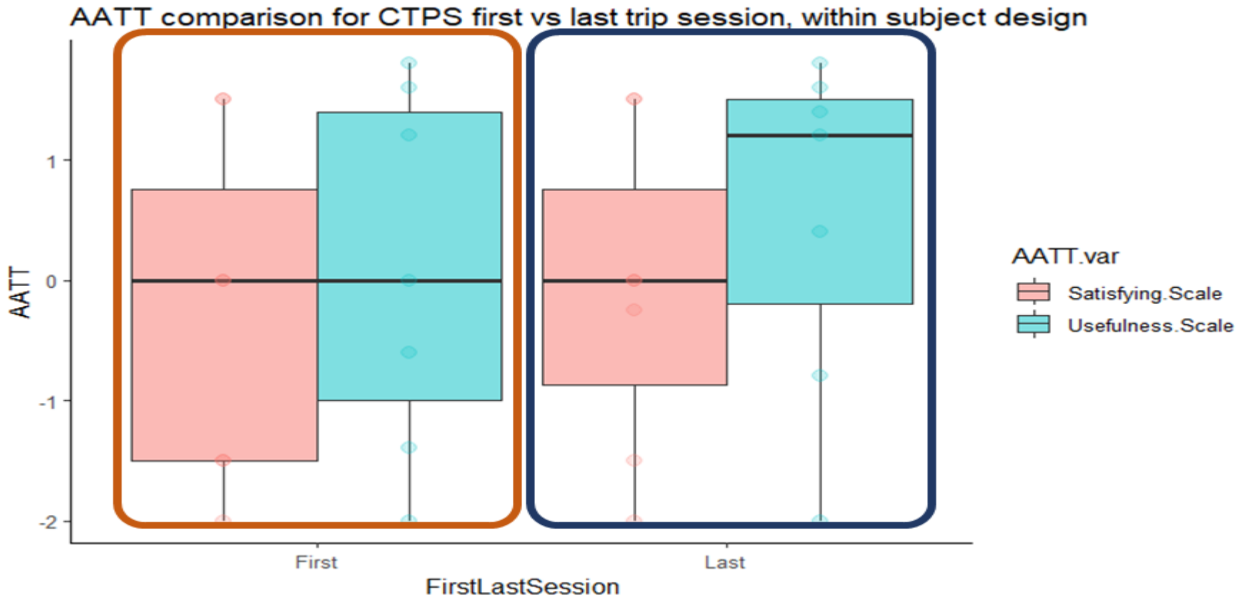


Figure 14: AATT scores comparing AB1 and AB2 for all CTPS trips and drivers

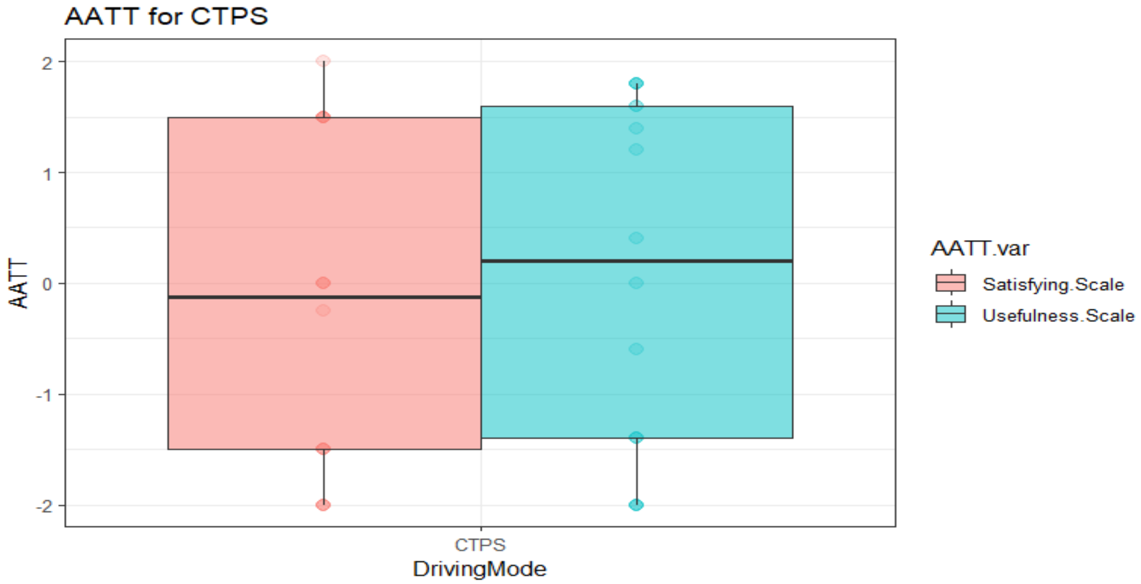


Figure 15: Aggregate AATT scores for all CTPS trips and drivers

5.3.5 Equipment and Comfort (E&C)

Drivers completed an Equipment and Comfort Questionnaire on their final drive. By comparing their normal comfort/discomfort levels associated with operating their regular trucks, to those reported after their on-road experience, it allowed for the identification of acute pain onset associated with driving the platooning trucks. Acute pain/discomfort can explain an increased stress response due to neuroinflammation, independent of other task related factors.

Drivers noted no discernible differences between the features of the lead and follower trucks, and a review of the equipment specs indicated the same cab design and seat features in each of these trucks. Thus, it is assumed that the change in comfort levels based on their individual responses were reflective of the differences between their regular truck and the Pronto equipment.

The two trucks utilized in this study contained passive suspension seats, as do the regular trucks that the Bison drivers have. However, it is theorized that the drivers are conditioned to their regular truck seats (e.g. seat pan and back rest comfort, reach distance to pedals, steering wheel hand heights, armrests, etc.) and introducing a new seat of any kind could result in acute discomfort and subsequently, an elevated stress response as the body adjusts to the new positioning.

Using a comparison scale of increased comfort, less comfort, or no change in comfort, Figure 16 demonstrates increases in subjective discomfort across all regions of the body associated with the Pronto trucks. In particular, shoulder and back pain noted the highest discomfort responses. A comparison of specific cab features relative to driver comfort (Figure 17) indicates that the likely cause of the increase in back discomfort was the seat design itself. However, there were initial concerns raised by the drivers during the on-road trials that their assigned driving days were longer than their regular trips and compounded due to the amount of measurement protocols that needed to be completed during Pre Trip, Post Trip 1, and Post Trip 2. The prevalence for the onset of musculoskeletal discomfort, especially low back pain, is significantly influenced by daily duration exposures to sustained driving postures and vibration levels (Andrusaitis et al., 2006; Lee et al., 2020; Miyamoto et al., 2000; Robb & Mansfield, 2007). During the administration of questionnaires and TBCT protocols during Post Trip 1 the drivers remained seated in their trucks and the trucks remained running, providing ongoing vibration primarily from the transmission, increasing exposure levels compared to their regular working days.

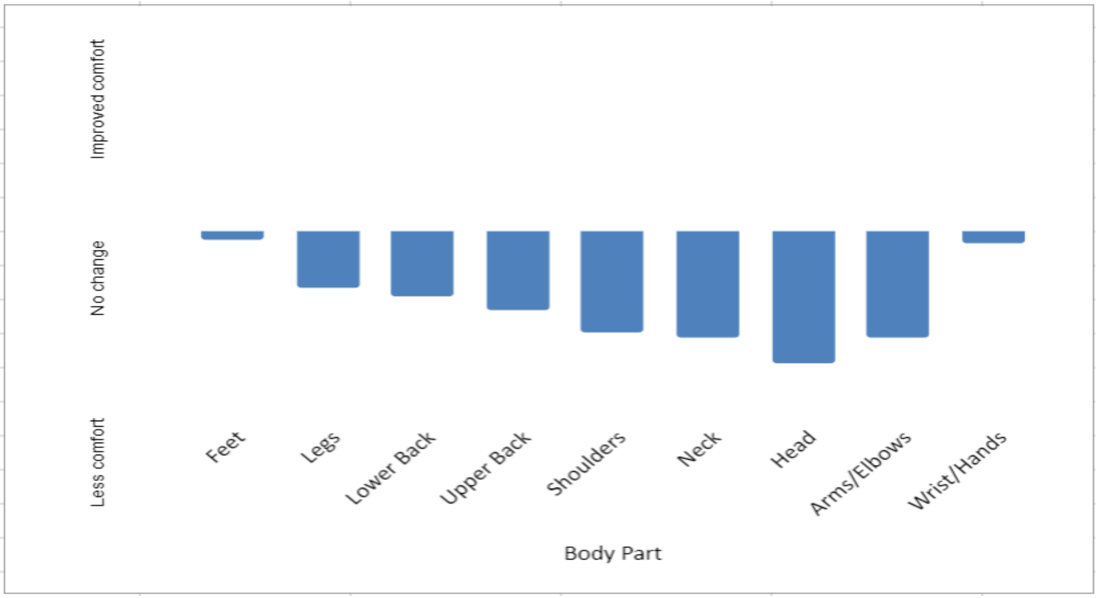


Figure 16: Equipment and comfort for all trips (body parts)

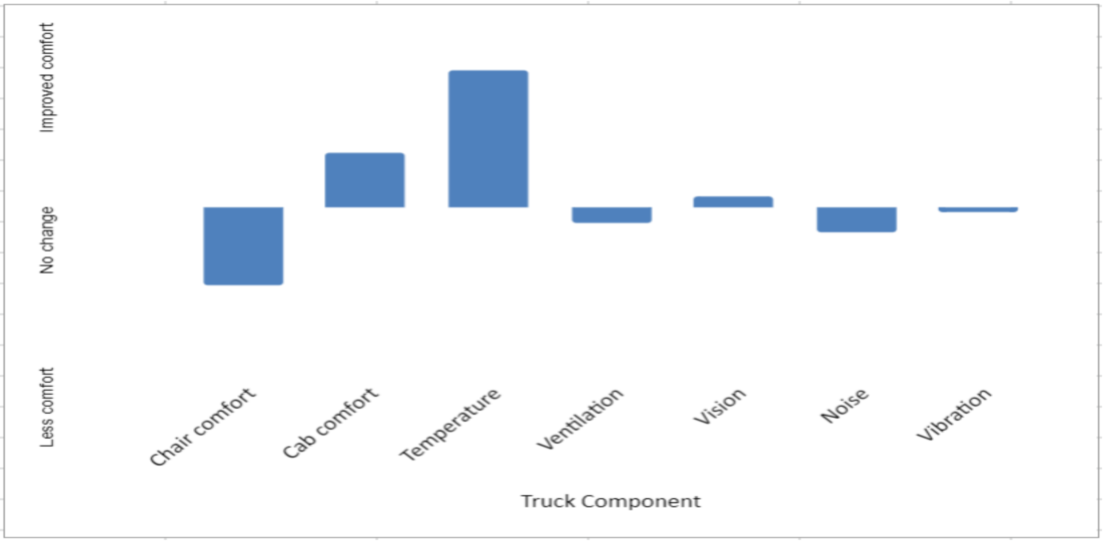


Figure 17: Equipment and comfort for all trips (truck components)

5.4 Conclusion

The questionnaires utilized in this study were selected to support the original methodology and hypotheses which sought to explore the effects of ADS and platooning on drivers' active and passive fatigue. The MRPQ and DSI were selected to determine risk propensity and personality traits that could affect the stress response and fatigue vulnerability. The TiA and AATT were included to further probe how trust and acceptance of the ADS platooning system would affect driver engagement, alertness, and vigilance and thus, their potential for fatigue and/or complacency. The KSS and E&C scales were incorporated to explore the role of pre-existing fatigue (poor sleep, shortened sleep) and of any fatigue induced by the driving environment itself (musculoskeletal fatigue, noise and vibration). Researchers included the DALI scales to capture driver perceived workload during the various driving conditions.

As the subjective nature of self-reporting tools may not accurately reflect the drivers' internal states (e.g., alertness, mental workload, engagement, vigilance, stress, etc.), physiological measures (Fitbit, Muse headband) were included in the study design to provide more objective measurements.

Unfortunately, due to the low number of participants and other limitations of this pilot study, researchers were unable to explore direct correlations between subjective, self-reporting measures and objective, physiological measures. As such, the value of the information derived from the questionnaires and their impact on driver performance in this ADS platooning system pilot study is limited.

As standalone measures, the MRPQ and DSI identified risk perception, willingness to engage in risky behaviors, and personality traits that affect risk tolerance. Although supported by research, in the absence of corroborative and objective physiological data it is difficult to derive a link between these scales and correlates to stress and fatigue vulnerability, in addition to user acceptance and trust in an ADS system.

The TiA, AATT, DALI, and MRPQ scales provided insight into driver trust, behavior and acceptance of the ADS platooning system that was in line with research. The literature shows that driver trust and risk attitude are related to acceptance of ADS but literature about driver trust and acceptance of platooning in real word driving is lacking; this was the first study attempt to do so. Drivers in this pilot study demonstrated stereotypical levels of trust, risk attitudes, and acceptance.

The DALI scores also reflect general driver perceptions of workload in ADS, and the peculiarities of the CTPS driving experience (unpredictable or erratic system behavior) likely influenced the change in their perceived vs experienced workload, and also prevented the onset of passive fatigue.

The E&C results suggest that there is preliminary work to be done prior to introducing the drivers to the ADS. Recognizing the impact of acute discomfort on the stress response and how it may in turn affect driver performance from a fatigue and/or distraction perspective, future studies should consider the provision of driver training in how to optimize the cab environment, particularly how to adjust unfamiliar seat features to reduce musculoskeletal discomfort and distraction, and ongoing alterations to their seated position to reduce the effects of whole body vibration and subsequent discomfort levels. Consideration should also be given to the total amount of driving exposure with the truck prior to data capture, as well as the length of time in the truck compared to normal daily exposures, as these were both noted to negatively influence the overall driver platooning experience, impacting time on task factors as well as time away from home factors.

6 TABLET-BASED COGNITIVE TASKS (TBCT)

6.1 Overview

The tablet-based cognitive tasks (TBCT) consist of four discrete tasks. They are designed to test for simple reaction time (RT), judgement, memory, and bi-manual perceptual motor (control) performance. This continuous battery of cognitive tasks was originally designed, developed, and validated by Impirica Tech in conjunction with the University of Alberta as an in-office screen to discriminate between safe and unsafe drivers in several contexts. An earlier version of this assessment, consisting of a touchscreen and a 3-button base was successful in distinguishing between safe and unsafe drivers in a stroke population (Choi et al., 2015), in cognitively impaired drivers (Dobbs & Résumé, 2013), and in drivers who were referred for a driving assessment (Korner-Bitensky & Sofer, 2009). The current version of this assessment, which is completed on a mounted iPad with a trained examiner, was successful in predicting on-road performance in a mixed (healthy and at-risk for cognitive impairment) older adult population, using machine learning (Bakhtiari et al., 2020). We have also used the current assessment to detect specific cognitive deficits in those who are under the influence of different drugs (e.g., cannabis and cocaine) (Tomzcak et al., 2022).

Low and high workloads can produce higher and lower amounts of passive fatigue, respectively, while driving. (May & Baldwin, 2009) proposed three sources that contribute to fatigue in drivers: sleep deprivation or untreated sleep disorders, performing a monotonous task (passive fatigue), or excessive depletion of cognitive resources during load or multitasking such as when driving (active fatigue). It is well-known fatigue can negatively impact cognition, and fatigue occurring from highway driving can impact safety through impaired cognition (Ting et al., 2008) Kontotasios (2020) suggests active fatigue seems to deplete individuals' cognitive resources such that fatigue is increased in subsequent driving tasks, and while passive fatigue also depletes individuals' cognitive resources, fatigue induction in subsequent driving tasks may be less than with prior active fatigue). The purpose of the TBCT in this study was to assess if, and in what way, cognition is impacted after driving with CTPS and to what extent fatigue and perceived task load influences cognition.

6.1.1 Reaction Time (RT)

For the RT task, drivers hold a button until the light at the top of the screen changes from red to green. When this happens, the box in the middle of the screen will move either left or right. The driver must press the stop button on the same side the box is moving. Furthermore, the location of the green light on the screen (either left or right) usually provides a hint (Figure 18) and creates two-levels of difficulty. In congruent trials (Figure 18.a), the visual cue is the same as the object's moving direction (easy), while in the incongruent trials (Figure 18.b), the visual cue and the object's moving direction are not the same (hard). The incongruent trials are more difficult because they require attention to be shifted from one side of space to the other. A short video of the task is available at http://tiny.cc/TBCTReaction.

The goal of the task is to assess the driver's attention, reaction time, and accuracy. Additionally, it examines how well drivers can wait for a cue and respond to miscues quickly and accurately. It also reveals differences in attention associated with spatial congruity. If drivers are cognitively impaired, they are more likely to react slowly and are more easily confused by the incongruent miscues. Thus, it simulates performance requirements for safe real-world driving in several ways.

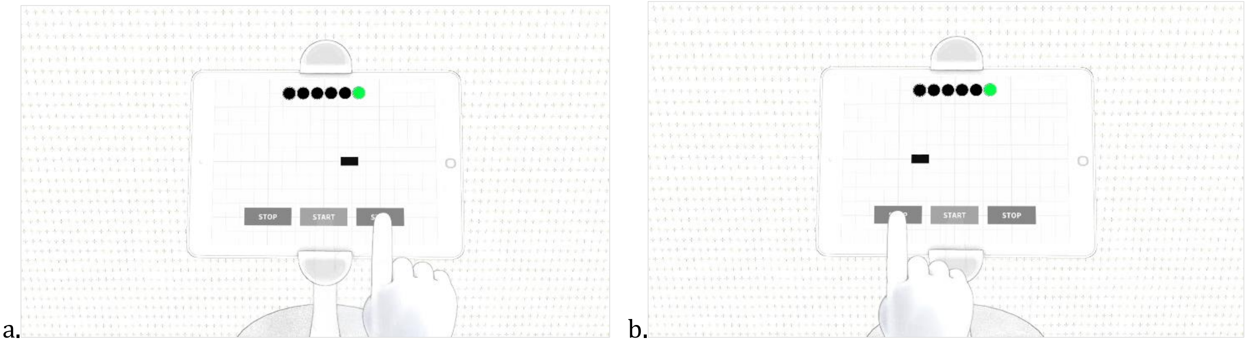


Figure 18: Reaction Time Task. Participants must hold a button until the light at the top of the screen changes from red to green. When this happens, the box in the middle of the screen will move either left or right. The driver must press the stop button on the same side the box is moving. In congruent trials (a), the visual cue is the same as the object's moving direction, unlike the incongruent trials (b).

6.1.2 Judgment Task

For the judgment task, drivers asked to control a box through a set of moving lines. The first step is to hold a button until the light turns green. Once it is green, they are to click the 'Go' button when they feel the path is clear (Figure 19). Drivers can press the Stop button to stop the box and press 'Go' again when it is safe. There are two stages with this task. In stage 1, there is only one set of moving lines, whereas in stage 2, two sets of moving lines exist. A short video of the task is available at http://tiny.cc/TBCTJudgment.

This task aims to assess drivers' judgment, reaction time, attention, and inhibition. When drivers are cognitively impaired, they have difficulty judging the appropriate time to move the box. Because regular judgment makes safer decisions on how to maneuver, this task relates to safe driving because stopping and starting in dynamic situations is a key factor on the road.

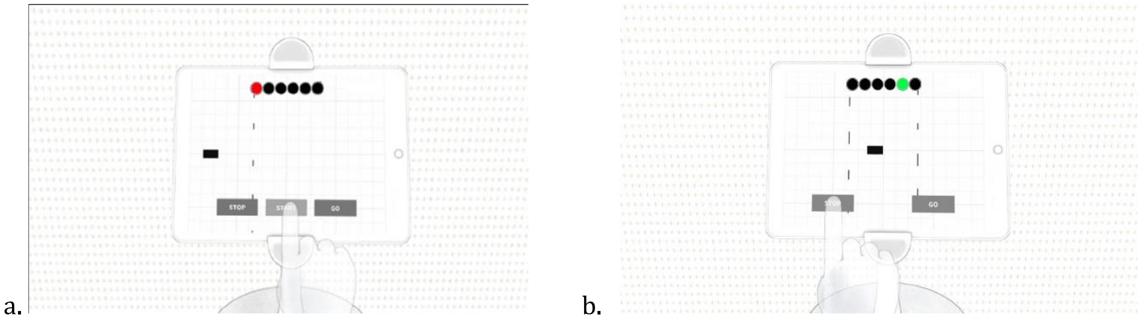


Figure 19: Judgement Task. Drivers were required to control a box through a set of moving lines. The first step is to hold a button until the light turns green. Once it is green, they are to click the 'Go' button when they feel the path is clear. In stage 1 (a.), there is only one set of moving lines, whereas in stage 2 (b.), two sets of moving lines exist.

6.1.3 Memory Task

For the memory task, the participants were required to draw a set of shapes from memory. First, the test shows drivers a set of shapes followed by a brief delay with distraction. They then must recreate these shapes (Figure 20). In the initial 8 trials, one shape is presented. In the last 8 trials, two shapes are shown, and the driver redraws both figures after the distraction period. A short video of the task is available at http://tiny.cc/TBCTMemory.

This task assesses drivers' ability to engage the input, storage, and retrieval aspects of short-term memory reflected by the task completion time and shape recall accuracy. If drivers are cognitively impaired, they will likely struggle to recreate the shapes accurately and will take longer. As such, this reflects real-world driving, because memory is a basic component of all skilled performance including safe driving.

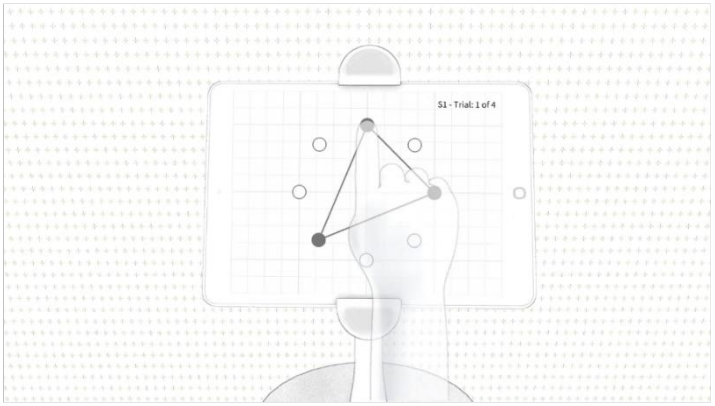


Figure 20: Memory Task. Participants were required to draw a set of shapes from memory. First, the test shows drivers a set of shapes followed by a brief distraction. They then must recreate these shapes.

6.1.4 Control Task

For the bi-manual perceptual motor (control) task, the participants were required to keep a virtual ball inside a target circle while it moves down a track. On the path, stationary and moving hazards are presented simultaneously. The driver must avoid these hazards while keeping the ball within the circle for as long as possible (Figure 21). A short video of the task is available at http://tiny.cc/TBCTControl.

The task examines drivers' ability to dynamically track a moving target, multi-task, and avoid unexpected hazards. Cognitively impaired drivers often have difficulty performing this task. It mimics driving because the rolling ball position is determined by "steering" the iPad with two hands as one would with a steering wheel. It is also a continuous and dynamic task with varied levels of difficulty as in the case of real-world driving.

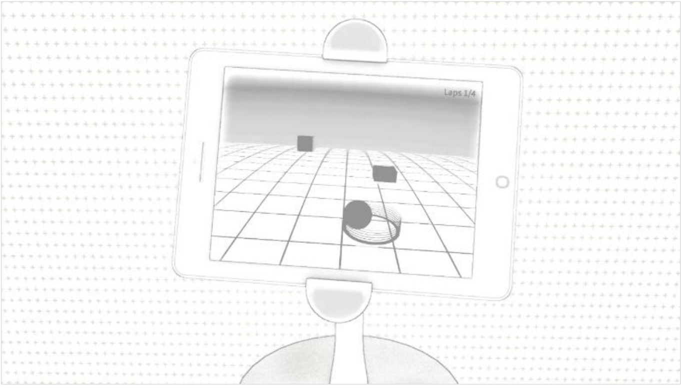


Figure 21: Control Task. The participants were required to keep a ball inside a circle while it moves down a track. On the path, stationary and moving hazards will be present simultaneously. The driver must avoid these hazards while keeping the ball within the circle.

6.2 Collected Data

Based on the Initial methodology, we planned to collect TBCT data before and after each trip, which means Pre Trip in Calgary, Post Trip 1 in Edmonton, and finally Post Trip 2 in Calgary. However, due to the long working day, and at the request of the drivers, we dropped the midpoint data collection from TBCT in Edmonton. A total of 65 dependent measures were extracted from all the tasks. Appendix E provides an explanation of each extracted measure.

6.3 Analysis

As a result of incomplete data collection, some of the initial hypotheses could not be evaluated, including the effect of the platooning interval on drivers' performance (Hypothesis 5, and Hypothesis 6). In this section, drivers' performance on TBCT between Pre Trip and Post Trip 2 for CTPS trips has been compared. As we were underpowered, investigating other hypotheses was not possible.

6.3.1 Reaction Time (RT)

Figure 22 shows the performance of drivers for both congruent (cue and response direction are the same) and incongruent (cue and response direction are not the same) trials. In both conditions, and for both pre and post trip, drivers were able to perform the task with almost 100% success rate (Figure 22). Figure 23 shows drivers' reaction time in these tasks. As we expected, responding to incongruent trials takes longer, which is likely related to greater effort and processing time needed in this task to inhibit the response to the incorrect cue, and shift attention to the target location. When we compare the difference between congruent and incongruent trials across two time points (Pre Trip vs. Post Trip) we see a trend-level increase for the second time point (df = 6, p-val = .089, diff = -0.035, t-test= -2.03). While not statistically significant, this finding likely reflects the impact of fatigue and cognitive load associated with the drivers' long workday. That is, the attentional load effect of incongruent trials is exacerbated by driving the two trips. Furthermore, since the was no difference in accuracy in this task over time, it may suggest that the RT differences reflect cognitive workload rather than a shift in speed-accuracy tradeoff strategies. This would be an interesting aspect of the drivers' cognition to probe further in future well-powered studies.

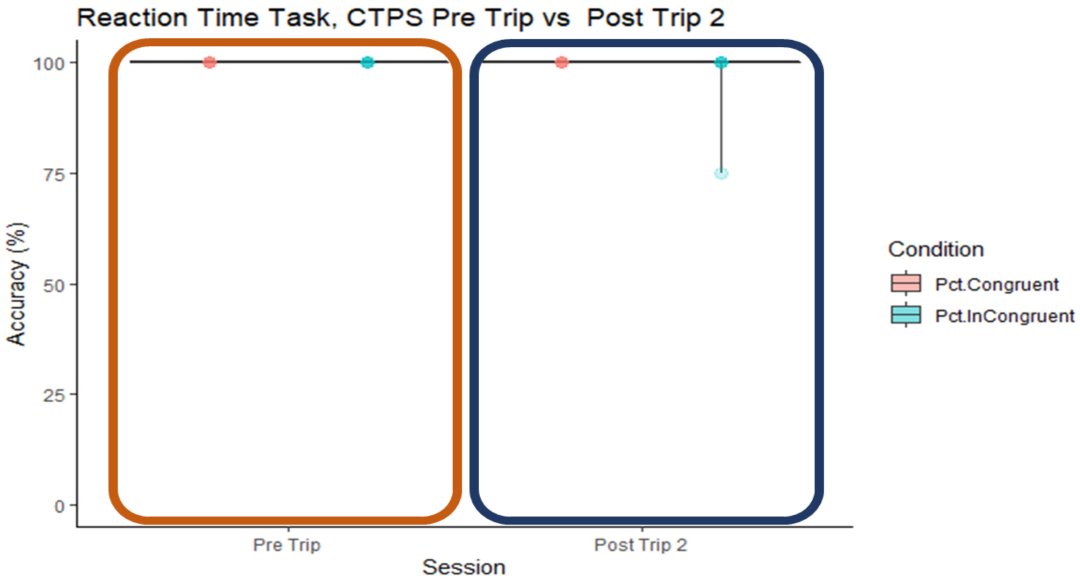


Figure 22: Reaction Time Task (Accuracy), CTPS Pre Trip vs Post Trip 2. Drivers successfully responded to this task in both Pre Trip and Post Trip 2 sessions.

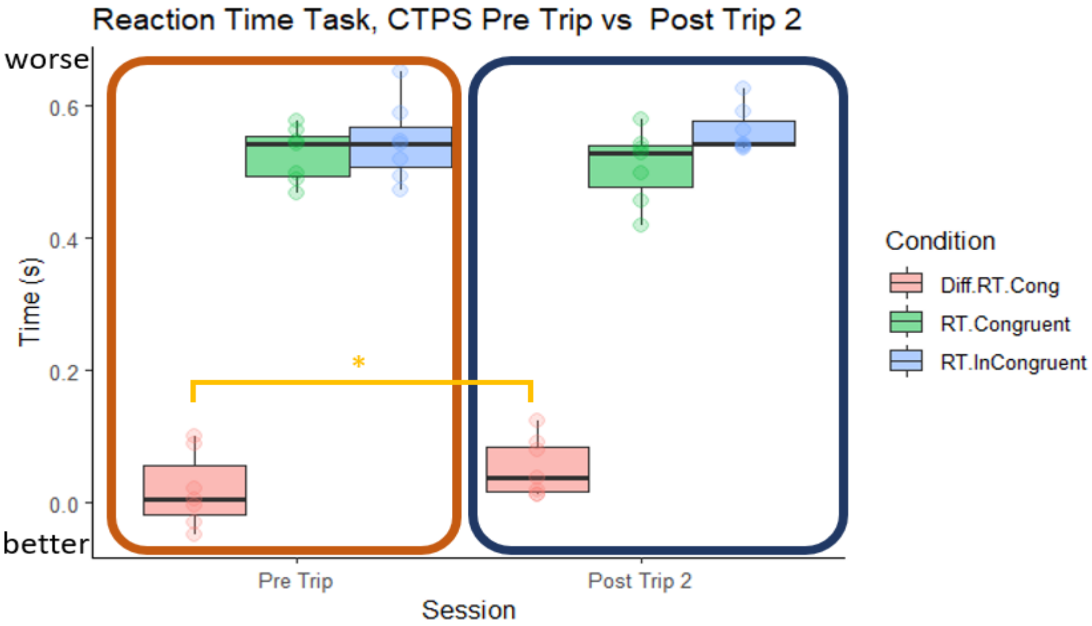


Figure 23: Reaction Time Task (Reaction Time). A trend for increase in the difference between the reaction time for incongruent and congruent trials in Post Trip 2 compared to Pre Trip (df = 6, p-val = .089, diff = -0.035, t-test)

6.3.2 Judgment Task

Drivers' accuracy in the judgment tasks was high and did not change from the Pre Trip to Post Trip 2 session (Figure 24). Comparing reaction time showed that drivers were slightly faster to press the start button in the second stage of the task, for Post Trip 2 (df = 6, p-val = .068, diff = 0.45, t-test = 2.23)

(Figure 25). This is interesting as we expected drivers to become more fatigued and slower to respond at the end of the day. One possible explanation is that due to the fatigue they became more impatient to finish their tasks and rushed into starting. However, this effect was also only at the trend-level, and overall, the drivers performed this task well.

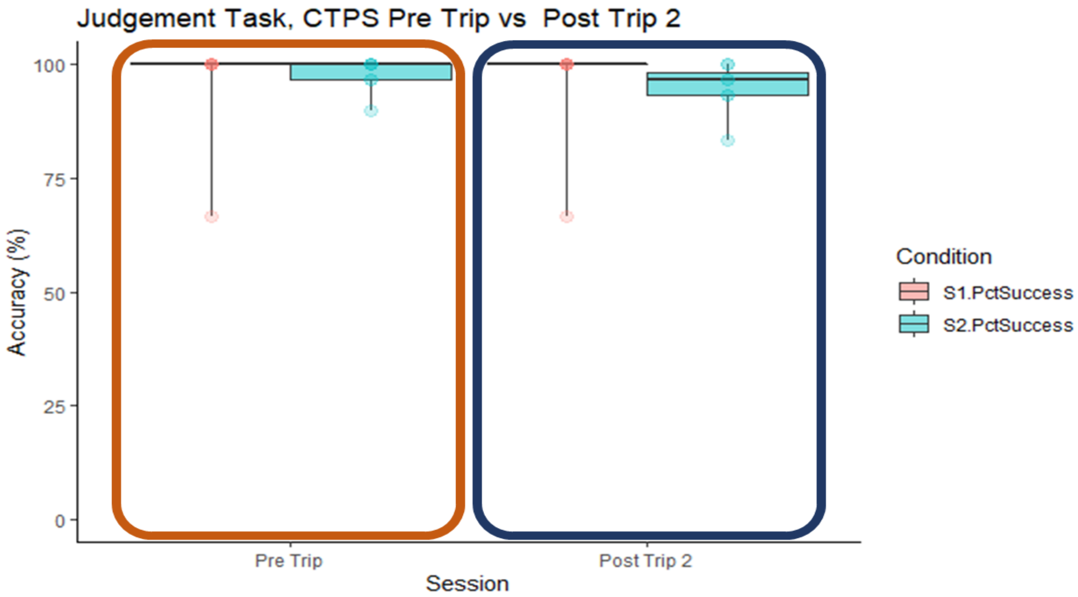


Figure 24: Judgment Task, Accuracy, CTPS Pre Trip vs Post Trip 2. No significant difference.

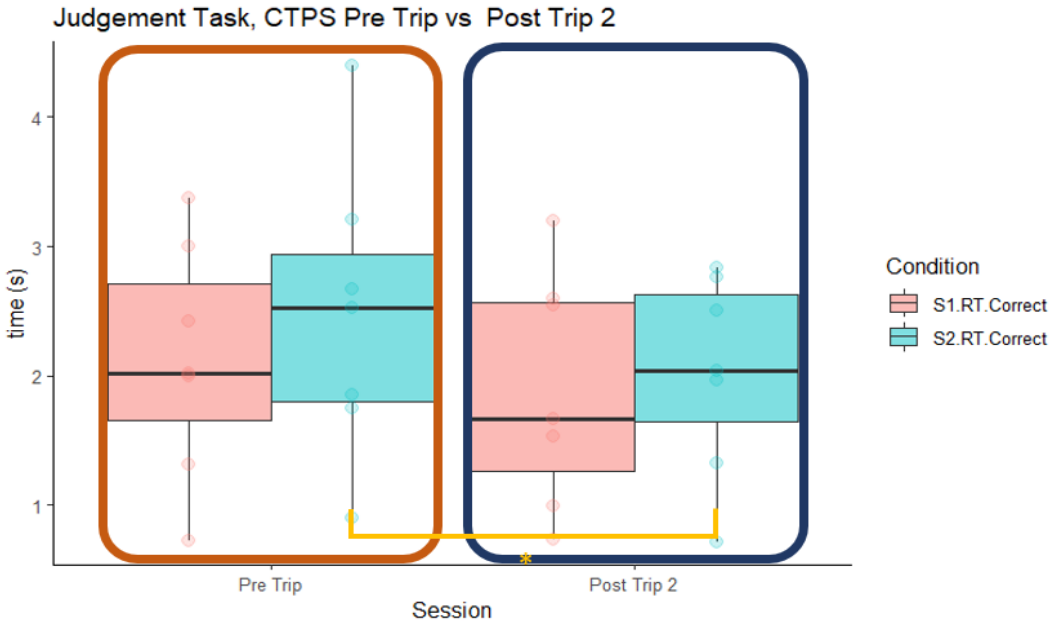


Figure 25: Judgment Task, CTPS Pre Trip vs Post Trip 2. Drivers were faster in Post Trip 2 session (df = 6, p-val = .068, diff = 0.45, t-test = 2.23).

6.3.3 Memory Task

We assessed performance on the memory task by comparing the two levels of difficulty of the task: drawing one shape, vs. drawing two shapes. Drawing two shapes is more complex, and as expected the accuracy is less compared to drawing one shape (Figure 26). However, there is not a statistically significant difference. There were no differences in the time to draw the correct shapes between Pre and Post Trip 2 sessions (Figure 27). This suggests that driving for long hours does not impact memory performance. This is not surprising given that memory is a robust cognitive behavior in healthy adults that is typically only impaired by high-levels of stress, workload, and fatigue.

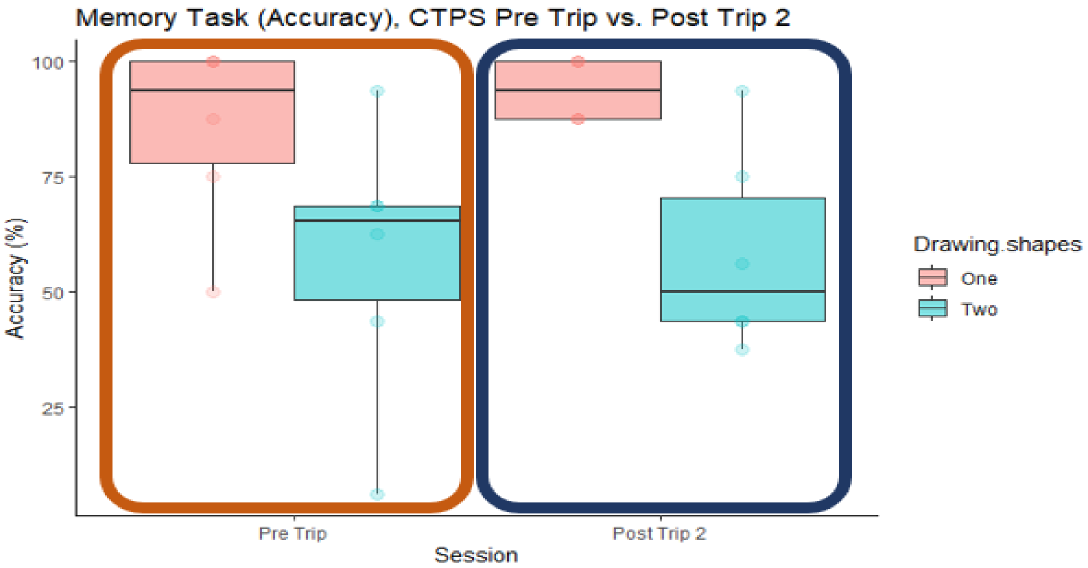


Figure 26: Memory Task (Accuracy), No difference between drivers' performance before and after trip.

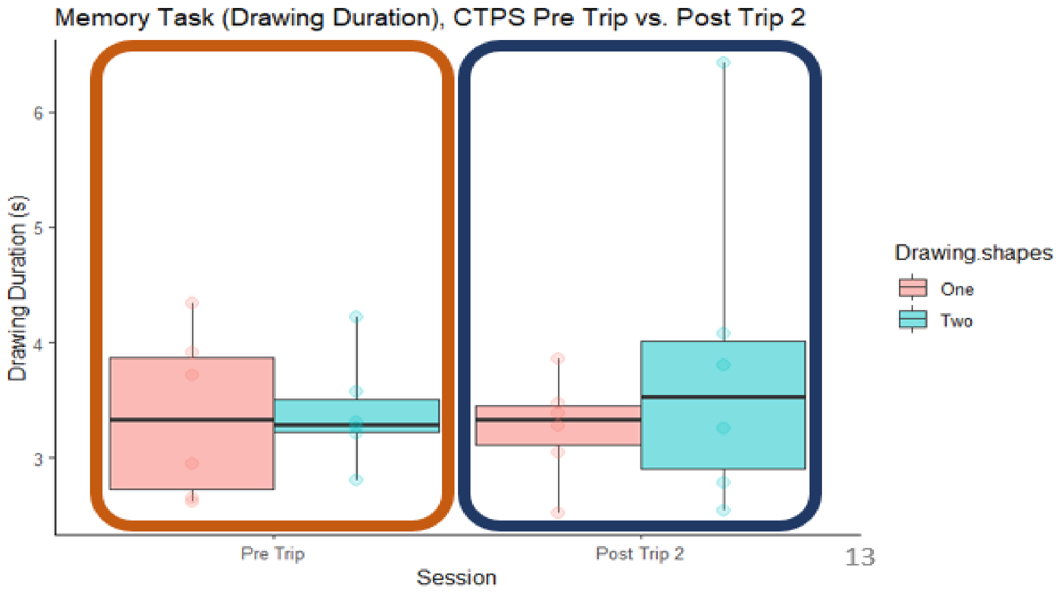


Figure 27: Memory Task (Drawing Duration), No difference between drivers' drawing speed before and after trip

6.3.4 Control Task

There were no effects in the Control task. Drivers could track the target ring similarly in both the Pre Trip and Post Trip 2 sessions. They were able to avoid both fixed and surprised obstacles with high accuracy (Figure 28). For some reason, drivers spent more time on the left edge of the screen during the Pre Trip session (df = 6, p-val = .044, diff = 0.41, t-test = 2.54) (Figure 29). It is unclear why this was the case, and it will require further investigation.

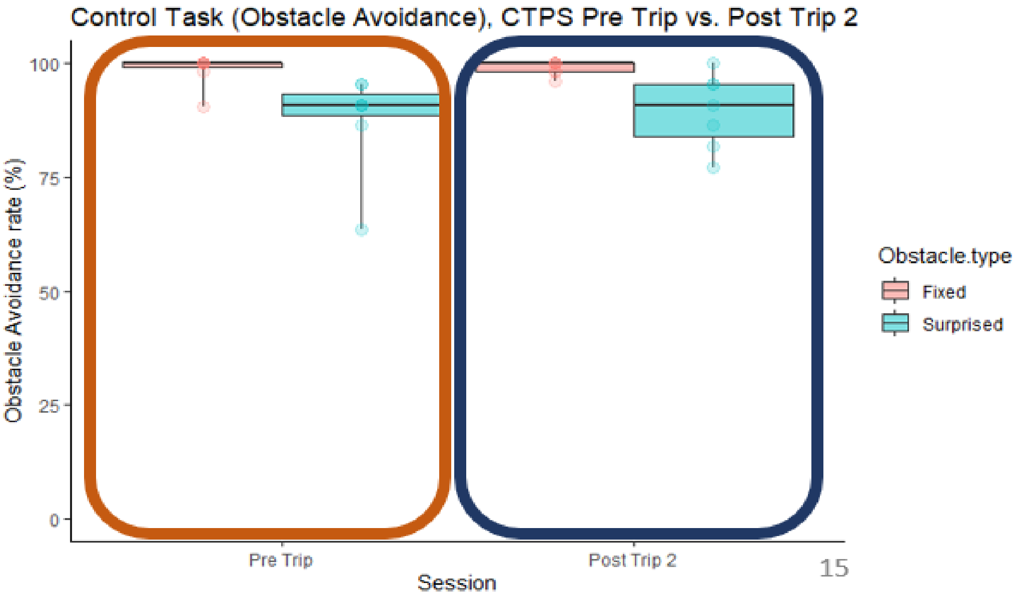


Figure 28: Control Task (Obstacle Avoidance). No difference between drivers' performance before and after trip.

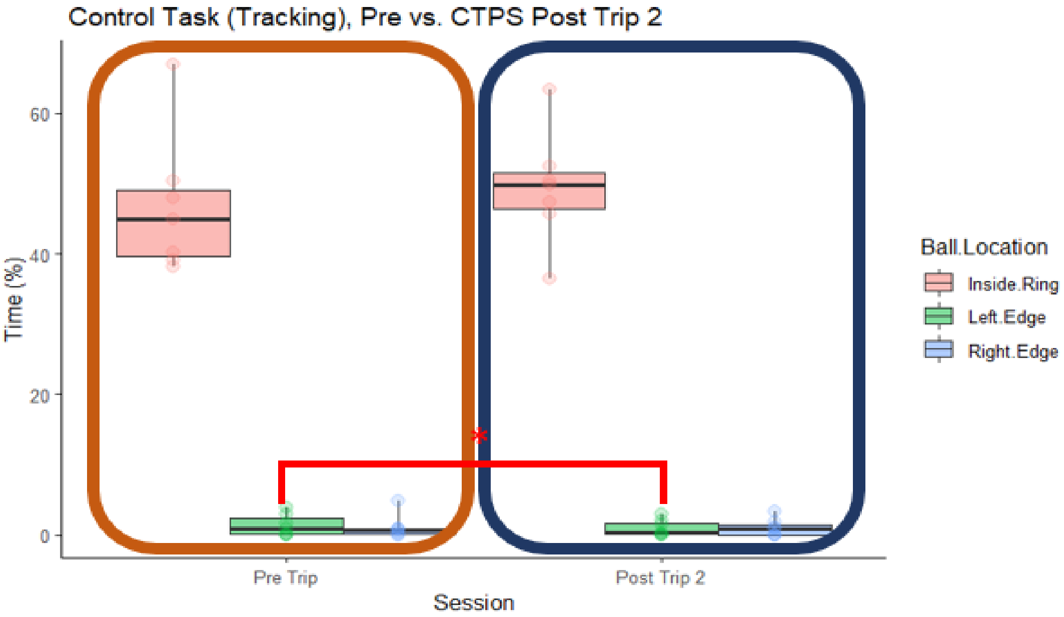


Figure 29: Control Task (Tracking). Less time on the left edge in the Post Trip 2 session (df = 6, p-val = .044, diff = 0.41, t-test = 2.54).

6.4 Conclusion

The TBCT showed some expected results. The drivers performed well, and overall, there were only minimal effects associated with driving between the Pre Trip and Post Trip tests. With the reduced scope of the study, we lost the ability to examine the various aspects of CTPS, and not being able to test at three time points. As such, we are cautious to interpret these results in the context of CTPS. However, we are encouraged that discrete cognitive testing could be important to examine in future studies of this nature.

Although the reaction time to the congruent and incongruent trials were not significantly different in the Post Trip 2 compared to the Pre Trip, the difference between these two types of trials was significantly greater during the Post Trip 2 session. This difference effect likely provides some evidence of the impact of driving-related fatigue on attention.

During the judgement task, the drivers' accuracy was similar before and after the full trip. This suggests the judgment task is not as sensitive to the same effects of fatigue as the incongruent RT effect above.

The finding that drivers were faster to begin the second stage of the judgement task is somewhat puzzling. One possibility is that it reflects the drivers' desire to finish the tasks as quickly as possible at the end of a long workday.

The drivers' performance during the memory and control tasks was unaffected in this study. This is potentially important because it shows differential sensitivity to driving related fatigue between the RT, judgement, memory, and control tasks. The spatial incongruity effect in the RT task is likely the lowest-level effect on cognition in the TBCT. The other tasks all require high-level cognition under more conscious control, which can be compensated for because they tap into a larger pool of cognitive resources. This worthwhile investigating further, across more interesting aspects of CTPS such as comparing the lead versus the follower truck and across different following distances.

7 EYE TRACKER DATA

7.1 Overview

Eye-tracking technology measures and analyzes the movements of one's eyes through detecting the infrared light reflected from the cornea (corneal reflection) to track where one is looking, what one is looking at, and how long one's gaze rests on a particular spot (Kapitaniak, et al., 2015). The Smart Eye Pro eye-tracking system was used to collect real-time driver gaze and pupil data during driving. In addition to corneal reflection, the system also collects information about the pupil, canthus (the outer and inner corner of the eye), and eyelid behavior (e.g., blinks, PERCLOS [average time the eyelids occlude the pupil]). While many eye tracking systems are marketed, the Smart Eye Pro has a higher successful calibration rate and includes features not found in other systems (e.g., tracking of eyelid behavior) (Funke et al., 2016). In relation to driving, the Smart Eye Pro has been used to detect cognitive distraction in drivers, in which eye movements become more abnormal as drivers become cognitively distracted (Le et al., 2020). It was found distraction resulted in less time gazing at off-road targets, decreased blink rate, and increased saccades (Yuen et al., 2021).

The Smart Eye Pro has previously been utilized in assessing the effects of driving automation. While investigating driver-automation with a driving simulator, Wang et al., (2019) tracked eye behavior and steering performance through all five levels of automation (no automation to hands free automated driving). Findings identified that eye gaze and correlated steering movements decreased as automation increased. The system allowed Hayashi et al., (2021) to create a situational awareness estimation model for unscheduled steering takeovers with semi-autonomous level 3 simulator. It has also been used in conjunction with EEG to assess driver drowsiness using PERCLOS and pupil diameter to create EEG-based driver drowsiness detection systems (Arefnezhad et al., 2022).

Eye tracking has been frequently used to detect fatigue in drivers, and eyelid behavior has been shown to be a reliable indicator of fatigue (Arefnezhad et al., 2022; Chang & Chen, 2014; Zhou et al., 2021). The disappearance of blinks, mini-blinks, and relative quiescence are reliable early signs of drowsiness (Larue et al., 2011), while PERCLOS is a scientifically established standard of driver drowsiness (Hanowski et al. 2008). Korber et al. (2015) investigated vigilance and passive fatigue in automated simulated highway driving and found increased driving time produced eye tracking indicators of fatigue and decreased vigilance (pupil diameter, blink frequency, blink duration), but had no effect on reaction time. Similarly, Buld et al. (2005) found eye closure increased with fatigue when using an automated driving system (Buld et al., 2005). Partially automated driving has been shown to lead to longer blink durations and decreased pupil diameter. Higher levels of subjective sleepiness were reported by truck drivers in a partially automated truck platooning simulator compared with baseline (Hjälmdahl et al., 2017). The purpose of the Smart Eye Pro system in this study was to assess the presence and degree of passive fatigue while using the CTPS with a focus on blink duration and frequency, as well as pupil diameter.

7.2 Data Acquisition

Eye data was collected using the 2-camera Smart Eye Pro system. Smart Eye Pro is available with three different types of cameras, and as the drivers' field of view during driving is mainly on the road, we chose a two-camera system in both trucks. The system installed on AB1 was borrowed from Transport Canada and had the Smart Eye Pro 1.3MP camera system. The system installed on AB2 was purchased with the Smart Eye Pro Dx camera. In Appendix F the specifications of the Smart Eye cameras and the comparison between the two camera systems are given. An industrial PC was used for collecting eye-tracker data in each truck.

7.3 Collected Data and Preprocessing

Nine drivers' data were utilized. Data for 17 trips was successfully collected for AB1, five morning trips, and 12 afternoon trips. Data from 24 trips was collected for AB2, 13 morning, and 12 afternoon trips. The trips' duration ranged from 2 hours to 3 hours 45 minutes. As previously explained in Figure 3, the collected data during each trip has been split into three sections: 1: the data within the first 45 minutes, 3: the data within the last 45 minutes, 2: the data between these two. Within each section, we tried to find 10 minutes of continuous clean data.

The study ruled out segments featuring disconnections or excessive noise. For a consistent range, a 10-minute time frame of clean data was chosen for each segment. The selected segment of the data was analyzed with MATLAB to ensure that they had a minimum gaze direction quality of 0.6 and 10 minutes of clean data. In cases where the 10-minute time frame did not yield 10 minutes of clean data, it was extended up to 15 minutes. In addition, the portions chosen for each segment were distinctly arranged in time to ensure that they were unique pieces of information. Two of the trips from AB2 had two segments instead of three. There were 11 trips that had at least one segment that did not yield ten minutes of clean data.

For the eye-tracker, clean data is defined by the portion of data for which the value of the Gaze Direction Quality is above 0.6. This variable that changes between 0 and 1, shows the quality of the collected data, with 1 representing excellent data quality and 0 representing poor data quality. Researchers visually inspected each data set and chose a segment of data with the best quality as representative of that section Figure 30.

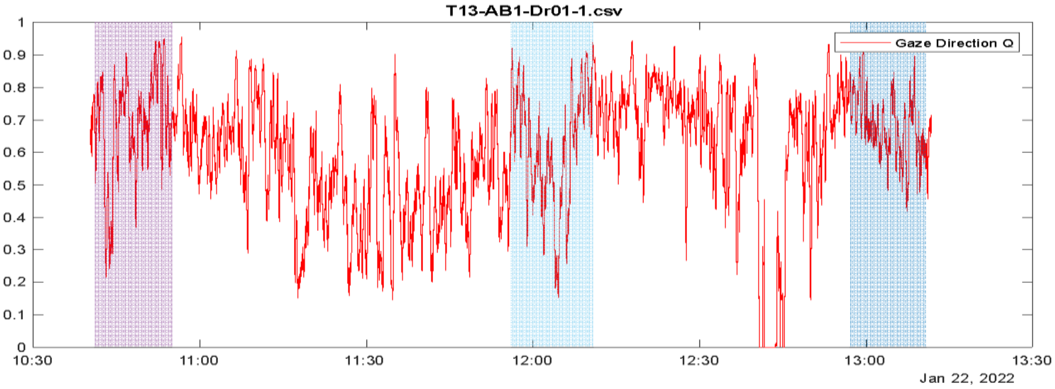


Figure 30: Gaze Direction Quality for one of the trips. Within each section, a segment of 10-15 minutes of clean data was chosen where the Gaze Direction Quality has a high and stable value.

This measure is dependent on many factors such as the location of the gaze, blink, and eye closure. Therefore, even in ideal in-office testing when the participants are required to look at two fixed points, there is lots of variation in this measure as can be seen in Figure 31. Importantly, the dash-mounted eye-tracking data in the present study looks very much like the data collected in a laboratory on a stationary table, suggesting the dash-mounted system is a viable option for future studies.

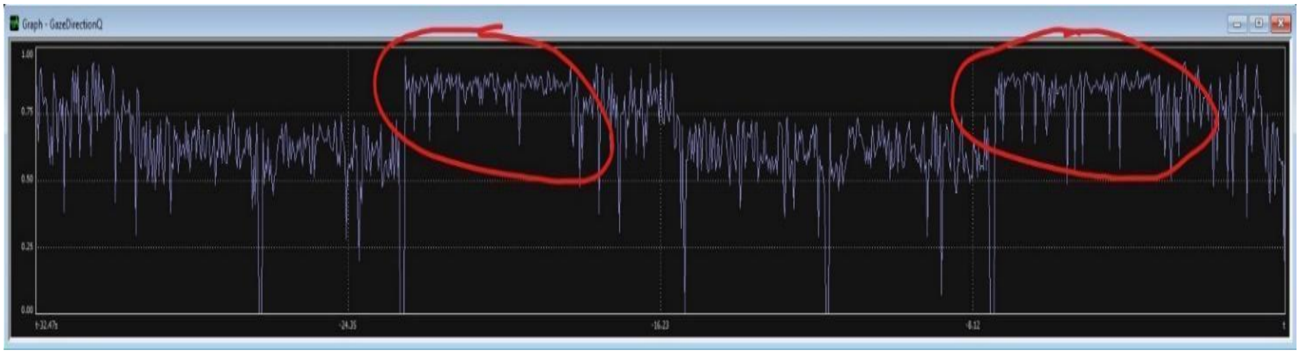


Figure 31: Gaze Direction Quality in a laboratory on a stationary table, when participants were asked to gaze at two fixed points.

7.4 Analysis

Smart Eye Pro provides various measures of eye behavior with a sampling frequency of 60 Hz. Previous studies have shown that pupil diameter, eye opening, blink, saccade, fixation, and the pattern of scanning the environment is associated with fatigue. In Appendix F, a list of measures that were extracted from the Smart Eye output and its relevance to fatigue is given.

In this Section, we first compared the quality of the collected time across time. For this purpose, we compared Gaze Direction Quality between Seg 1 and Seg 3 for AB1 and AB2 separately. For this analysis, we used all the collected data, and did not limit the analysis to one session from each driver unlike other analyses. Next, the comparison between various conditions based on the hypotheses was performed. For investigating the impact of driving condition (e.g., Platooning vs single truck), the data in the last segment (Seg 3) was considered representative of that trip.

7.4.1 Equipment validation and comparison

There was no significant difference between the two eye-tracker systems (Figure 32). However, this value declines in Seg 3 compared to Seg 1 for both AB1 (system borrowed from Transport Canada, df = 67, Wilcox p-val = .046, diff = 0.041, V = 1500), and AB2 (the new system, (df = 43, Wilcox p-val = .014, diff = 0.065, V = 703). Although this may be interpreted as decline in system's quality throughout the session, even under controlled in-office conditions the value of this signal is not the same within a few minutes' interval when the participants were asked to fix their gaze at two differing points Figure 29. Many factors including blinking, eye and head movement, and direction of gaze change the value of this signal.

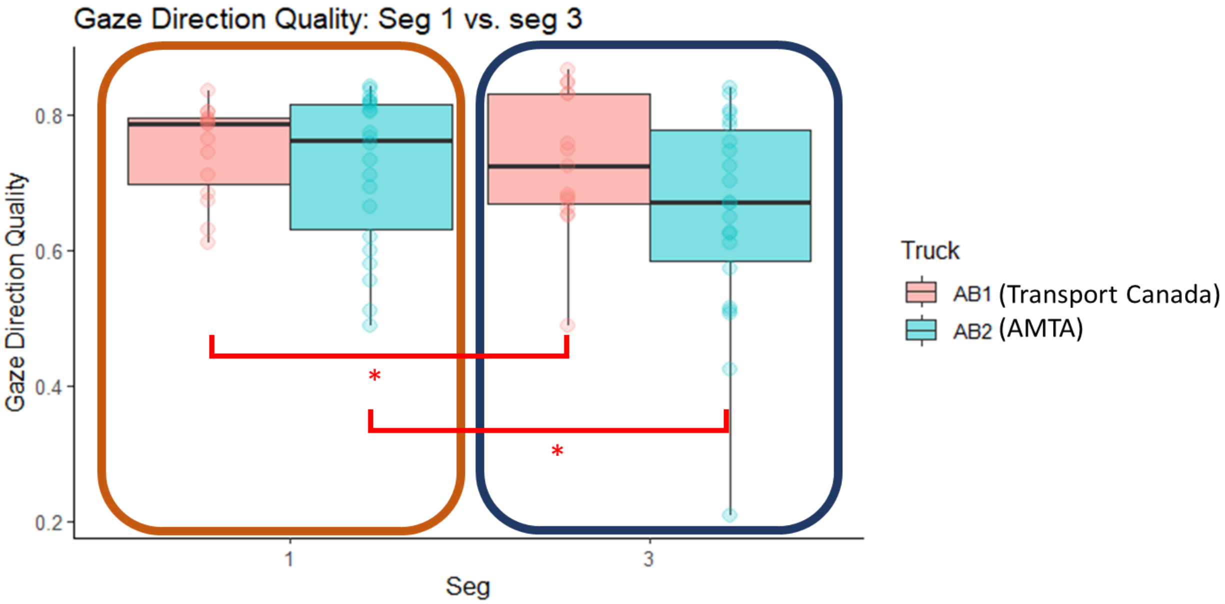


Figure 32: Gaze Direction Quality in Seg 1 is greater than Seg 3 in both AB1 (df = 67, Wilcox p-val = .046, diff = 0.041, V = 1500), and AB2 (df = 43, Wilcox p-val = .014, diff = 0.065, V = 703).

7.4.2 The effect of the ambient light on pupil diameter

Figure 33 shows the pupil diameter for platooning trips across time. The left side panel shows data for Trip 1, and the right-side panel shows data for Trip 2, with Seg 1 and Seg 3 shown in pink and turquoise respectively. The estimated timing of each segment is shown in the x-axis. For Trip 1, pupil diameter is significantly greater in Seg 1 compared to Seg 3 (df = 7, p-val = 0.049, diff = 0.00016, t-test = 2.38). For this trip, Seg 1 occurs around 8:30 AM, which is close to sunrise time in January in Calgary, while Seg 3 occurs around 11:00 AM, very close to mid-day. The light difference between these two segments is likely the reason for these pupil diameter effects.

For Trip 2, pupil diameter is significantly smaller in Seg 1 compared to Seg 3 (df = 7, p-val = .027, diff = -0.81, t-test = -2.78). For this trip, Seg 1 occurs around 15:30PM, while Seg 3 occurs around 17:00 PM after sunset in Calgary. Again, the light difference between these two segments is likely the main reason for these pupil diameter effects. There was also a trend for increases in the pupil diameter from Trip 1, Seg 3 to Trip 2 Seg 3, which also likely reflects darker ambient light in the evening time compared to the daytime (df = 6, p-val = .068, diff = -0.83, t-test = -2.22).

Research has shown that there is a strong relationship between sleep deprivation, pupil size, and pupil stability. Both pupil size and stability can objectively identify sleepiness and sleep deprivation. A well-rested individual can maintain constant pupil size in darkness for 15 minutes. The more sleep deprived the participants are, the less stable the pupil size will become. It fluctuates, becoming subtly bigger and smaller rather than maintaining its size. Moreover, the pupils' overall size will shrink, perhaps reflecting fatigue in the task of maintaining the larger size. The muscles themselves may tire and the ability to keep the pupil open may fade.

The normal pupil size in adults varies from 2 to 4 mm in diameter in bright light to 4 to 8 mm in the dark, which is in line with our observation in this study.

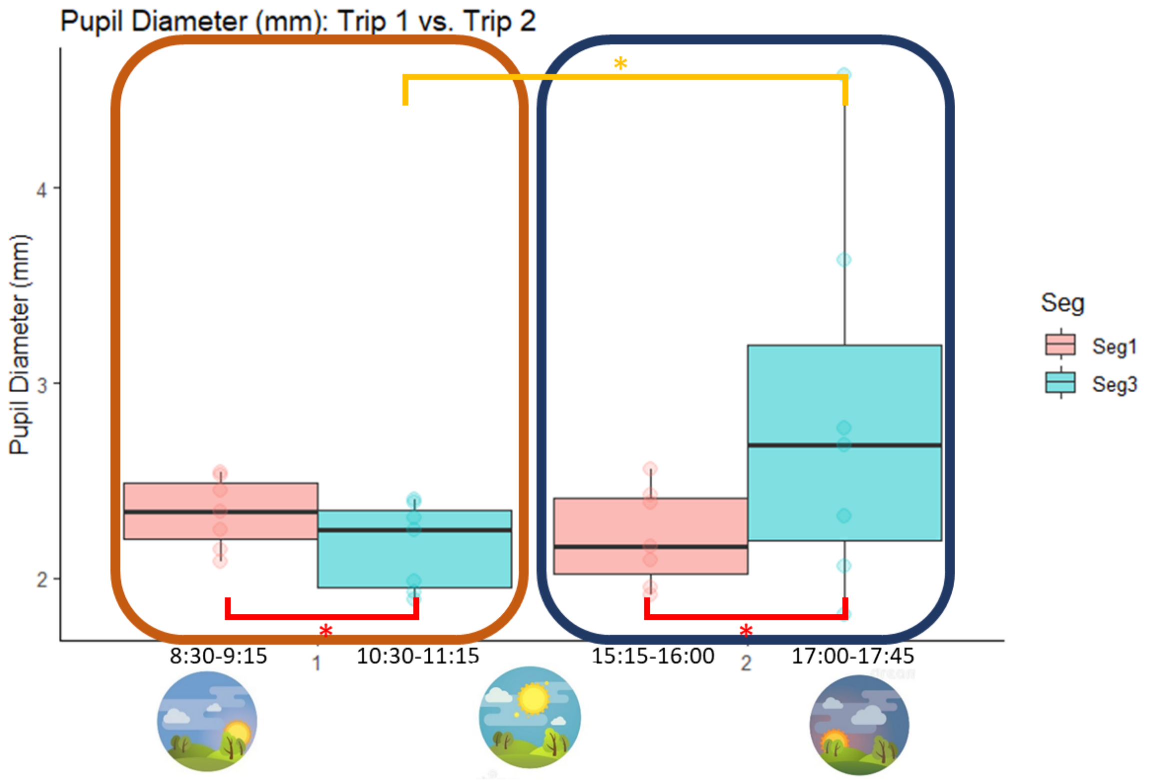


Figure 33: Pupil Diameter (mm) for Trip 1 (left panel) and Trip 2 (right panel). Pupil diameter decreases during mid-day compared to morning, and then increases again at the end of Trip 2

7.5 Conclusion

In this Chapter, the validity of using Smart Eye tracking system during on-road trials with two Smart Eye camera systems (Eye Pro Dx and Smart Eye Pro 1.3MP camera systems) was investigated first. Based on gaze direction quality measure, the average quality of the collected data was within acceptable range (>.3). However, this measure declined at the last segments of the trips for both camera systems. Although considering this decline the quality of data was still in the acceptable range for 4-5 hours driving session, for longer duration, more analysis is required.

Hypothesis 8 was designed to examine the impact of the time of day and driving hours on the drivers' experience. While we could not answer our original questions, the results show the ability of the eye-tracking system to successfully track changes in pupil diameter related to ambient light. We think this is important because future fully powered and balanced studies will likely be interested in examining this measure closely. Our preliminary results show that dash mounted eye-tracing will provide good quality data and has the sensitivity to observe differences in pupil size while driving. Unfortunately, comparisons of the other eye behavior measures (e.g., blinks, saccades) were not possible.

8 BRAIN DATA

8.1 Overview

Electroencephalography (EEG) is a technique that records the electrical activity of the brain traditionally using electrodes that are affixed to the scalp. These electrical signals can be used to infer and predict brain processes that occur during specific tasks or are associated with human behavior. Recent advances in technology have allowed portable electrode arrays to be readily available. In this study, we used a 5-channel Muse S headband to measure real time brain electrical signals during driving. Typical EEGs will have 32, 64, 128, or 256 channels with a higher number of channels increasing data quality, capture, and resolution. However, we used a 5-channel headband to maximize driver comfort and reduce the intrusiveness of the equipment. The Muse has been validated to be used in event related potential (ERP) research (Krigolson et al., 2017) and monitoring one's attention during meditation (Lutz et al., 2008; Wolkin, 2015), among other areas.

There are three types of waves that have been most widely associated with fatigue: alpha, beta, and theta rhythms. Each wave is associated with different types of cognition, likely originating from different brain circuits, and can be independently observable. Alpha waves likely come from the thalamus (Halgren et al., 2019) and the parietal cortex (Bagherzadeh et al., 2020). Alpha waves oscillate in the range of 7 to 14 Hz and likely play a part in filtering out information in order to help us focus and attend (Bagherzadeh et al., 2020). In a study conducted at MIT, researchers found that participants were able to better attend to objects when suppressing alpha waves in the opposite parietal cortex (Bagherzadeh et al., 2020). As such, a decrease in alpha waves signifies enhanced attention, while an increase in alpha waves signifies enhanced fatigue (Tran et al., 2020). Alpha waves are considered to be the most sensitive indicator of brain fatigue (Li et al., 2020).

Beta waves (15-29 Hz) likely come mainly from the thalamus (Sherman et al., 2016). The relationship between beta waves and fatigue is unclear, as some studies have found an increase in beta waves when participants are fatigued, while others have found a decrease or no change (Dissanayake et al., 2022). One theory is that beta waves may increase when participants are tired due to an attempt to maintain vigilance when fatigued (Craig et al., 2012).

Theta waves are associated with two main brain circuits. They likely come from the hippocampal-entorhinal system (Nuñez & Buño, 2021), but scalp recorded theta waves likely come from larger cortical regions including the frontal lobes (Cruikshank et al., 2012). These waves are strong during activities such as meditation, prayer, daydreaming, and sleep (Craig et al., 2012). They occur in the 3 to 8 Hz range and are often seen to increase when fatigued (Li et al., 2020; Tran et al., 2020).

The Muse has been utilized in numerous studies, across a variety of fields. However, only one study has used the Muse headband to directly analyze fatigue. Krigolson et al. (2021) investigated cognitive fatigue while participants completed a traditional visual attention task on an iPad. This study found increased fatigue was correlated with a decrease in P300 amplitude and an increase in P300 latency. Additionally, fatigue was associated with an increase in frontal delta and frontal theta power, which were in line with previous studies. However, contrary to previous studies there was a decrease in posterior alpha power with increased fatigue, suggesting an increased demand on attentional resources when one is fatigued.

Previous studies have utilized other EEG systems to analyze brain activity and fatigue and/or stress while driving. A critical review concluded that theta and delta activity increased when drivers transitioned to a fatigue state (Lal & Craig, 2001). While using a driving simulator, Jagannath & Balasubramanian (2014) observed that alpha power increased, and theta power decreased as driver fatigue increased. Park et al. (2018) investigated stress while driving an autonomous vehicle simulator and found the beta wave to alpha wave power ratio increased as driver stress increased. In another study, alpha wave power increased under monotonous driving conditions in an autonomous driving simulator while reaction time became slower (Zhang et al., 2021). The purpose of the Muse S EEG headband in this study was to assess how brain waves, specifically alpha, beta, and theta oscillations, are affected during the presence of driver passive fatigue while using the CTPS.

8.2 Data Acquisition

Electroencephalographic data were recorded from a Muse S (Gen 2) Brain sensing headband sampling at 256 Hz. The Muse EEG system has electrodes located analogous to Fpz, AF7, AF8, TP9, and TP10 with electrode Fpz utilized as the reference electrode during recording. These electrodes are placed over the forehead, and collect brain data from multiple parts of the brain. The Mind Monitor App was used to stream EEG data via Bluetooth from the Muse EEG system to an Apple iPhone 8 Smartphone, 2GB RAM, Apple A11 Bionic chip.

8.3 Collected Data and Preprocessing

We collected data from all nine drivers. For the drivers of the AB1 truck, there were 21 trips collected, 10 morning trips, and 11 afternoon trips. AB2 had 20 trips, 14 in the morning, and six in the afternoon. Similar to the eye tracker data, the collected data during each trip has been split into three sections: 1: the data within the first 45 minutes, 3: the data within the last 45 minutes, 2: the data between these two. Within each section, we tried to find 10 minutes of continuous clean data. In cases where the 10-minute time frame did not yield 10 minutes of clean data, it was extended up to 15 minutes. One trip from AB1 had two segments instead of three. There were 14 trips in AB1 (AB2) that had at least one (17) segment that did not yield ten minutes of clean data for the average of the lateral electrodes.

The Muse headband software records skin to electrode connectivity (resistance) for each electrode, while data is being collected. We used this information to define the portion of clean data. The connectivity measure changes between 1 (good connection), 2 (weak connection), and 4 (no connection). By visually inspecting the collected data, the segment of data with the best connectivity measures for all electrodes is chosen as a representative of that section (Figure 34).

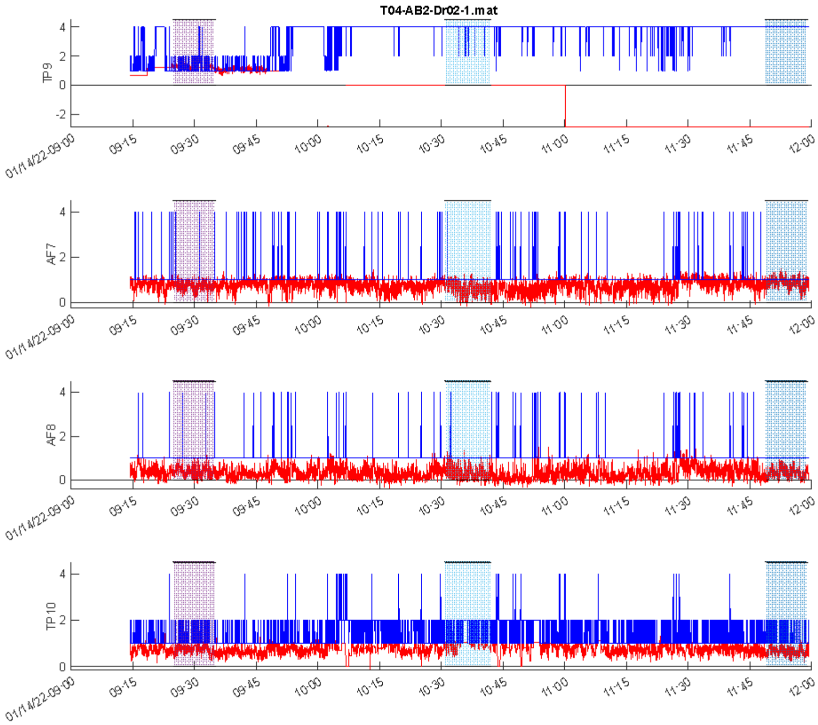


Figure 34: Brain oscillation from each electrode (red), and electrode-to-skin connectivity in blue. Portion of data with the best connectivity is chosen.

Mind Monitor calculates the absolute power in various frequency bands (the summation of all the power values of a signal within a frequency range.), including theta (3-8 Hz), alpha (7.5-13 Hz), and beta (15-29 Hz). We averaged absolute power for each frequency band over the signals from the two central electrodes (AF7 and AF8), and two lateral electrodes (TP9 and TP10) and investigated the changes of brain oscillations in these sites across different conditions (Figure 35).

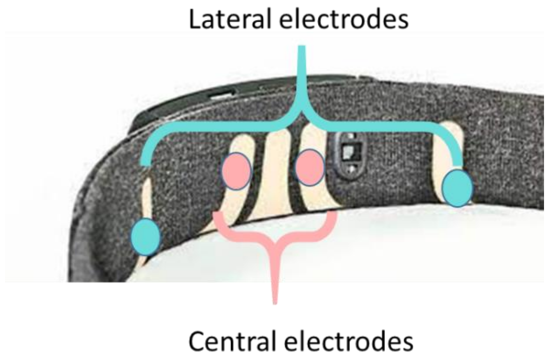


Figure 35: Electrodes on the Muse head band. The signals from the two central electrodes (AF7 and AF8 shown in pink), and two lateral electrodes (TP9 and TP10 shown in turquoise) are averaged.

8.4 Analysis

For the purposes of this set of analyses, we used the proprietary software from Muse (Mind Monitor). The Muse sensors collect EEG with sampling frequency of 250 Hz. We used the calculated frequency bands provided by the Mind Monitor software and focused on alpha, beta, and theta. These are the frequencies known to be associated with fatigue. The software also identifies various sources of noise (such as jaw clenching) and removes them from data using a correction algorithm.

In the next section, we compared the quality of the data across time. For this purpose, we compared the electrode-to-skin connectivity between Seg 1 and Seg 3 for the central and lateral electrodes separately. For this analysis, we used all the data, and did not limit the analysis to one session from each driver unlike other analysis. Next, the comparison between various conditions based on the hypotheses was performed. For investigating the impact of driving condition (i.e. Platooning vs Single truck), the data in the last segment (Seg 3) was considered representative of that trip and used to perform the statistical analyses.

8.4.1 Equipment validation and comparison

In Figure 36 the percentage of clean data from each electrode in each segment is shown. The central electrodes have a better connection compared to the lateral electrodes, possibly due to the shape and rigidity of the headband and its inability to provide a good connection in lateral sites due to variability of the skull shapes. However, both electrodes maintained their connection to the skin throughout the session and there was not a significant change across time.

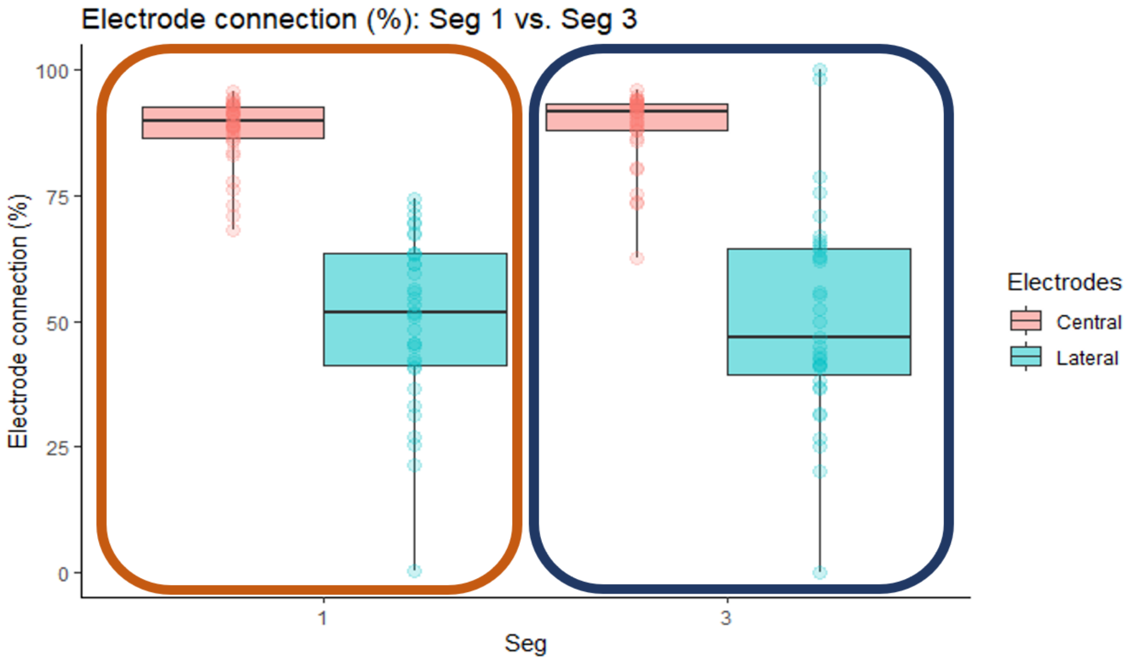


Figure 36: Percentage of time with good electrode-to-skin connection for central (pink) and lateral (turquoise) electrodes in Seg 1 and Seg 3 of Trip 1 (left panel) and Trip 2 (right panel): There is no difference from the beginning to the end of the trip.

8.4.2 The effect of time of day on brain oscillations

As with other aspects of this study, we took an exploratory approach with the analyses, and we are cautious in our interpretation due to the low statistical power. To investigate the impact of time of the day on drivers' fatigue related brain oscillations, we compared Seg 1 and Seg 3 within Trip 1 and 2, and then Seg 3 of Trip 1 and 2 for lateral (Figure 37) and central electrodes.

The brain oscillations were compared for platooning Trips 1 from the beginning of the trip (Seg 1) to the end of the trip (Seg 3). Based on the normality test (Shapiro test) data is normally distributed for both alpha and beta frequency band oscillations in Trip 1 (p-val >.05). The t-test shows that in Trip 1 there is a trend that the absolute band power of oscillations in the alpha frequency band is less in Seg 3 compared to Seg 1 (df = 6, p-val = .051, diff = 0.048, t-test = 2.43), and oscillations in the beta frequency band is significantly less in Seg 3 compared to Seg 1 (df = 6, p-val = .019, diff = 0.16, t-test = 3.19). As Seg 3 was collected around 11 AM, and Seg 1 is collected around sunrise, the decrease in alpha in Seg 3 compared to Seg 1, may suggest an increase in the drivers' vigilance as the circadian low passes. The increase in beta during the morning session may be due to the drivers' attempting to maintain vigilance while experiencing fatigue.

Moreover, by comparing the brain oscillations at the end (Seg 3) of Trip 1 to Trip 2, it was found that there is a trend for the absolute band power of the alpha oscillations to be greater at the end of Trip 2 compared to the end of Trip 1 (df = 5, p-val = .083, diff = -0.09, t-test = -2.17). The increase in alpha oscillations may suggest an increase in drivers' fatigue at the end of the long working day.

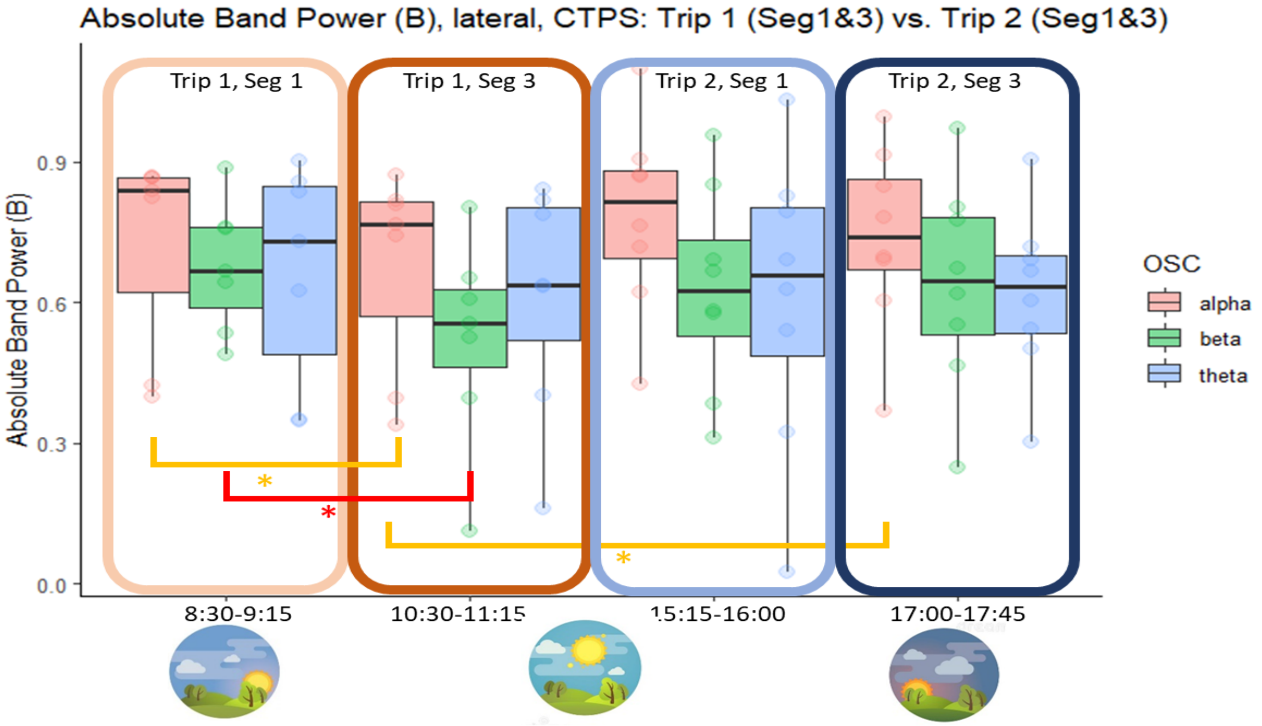


Figure 37: Changes in the brain oscillations from the lateral electrode. The absolute band power of beta oscillations is significantly less at the end of the Trip 1, and there is a trend for the absolute band power of alpha oscillations decreasing at the end of Trip 1 compared to the beginning and increasing at the end of Trip 2 compared to the end of Trip 1.

In the central electrodes (Figure 38), during Trip 1, the absolute band power of theta oscillations is significantly less at the end of the Trip compared to the beginning (df = 7, p-val = .047, diff = 0.064, t-test = 2.41). The decrease in theta may reflect fluctuations in fatigue during midday.

Comparing the oscillations at the end of Trip 2 vs. the end of Trip 1 shows that beta oscillations are increased (df = 7, p-val = .045, diff = -0.13, t-test = -2.44), while there is trend for a decrease in theta (df = 7, p-val = .062, diff = 0.05, t-test = 2.22). The increase in beta at the end of Trip 2 may suggest that drivers increased effort to maintain their vigilance while being tired at the end of the long day. There is also a trend level decrease in theta oscillations.

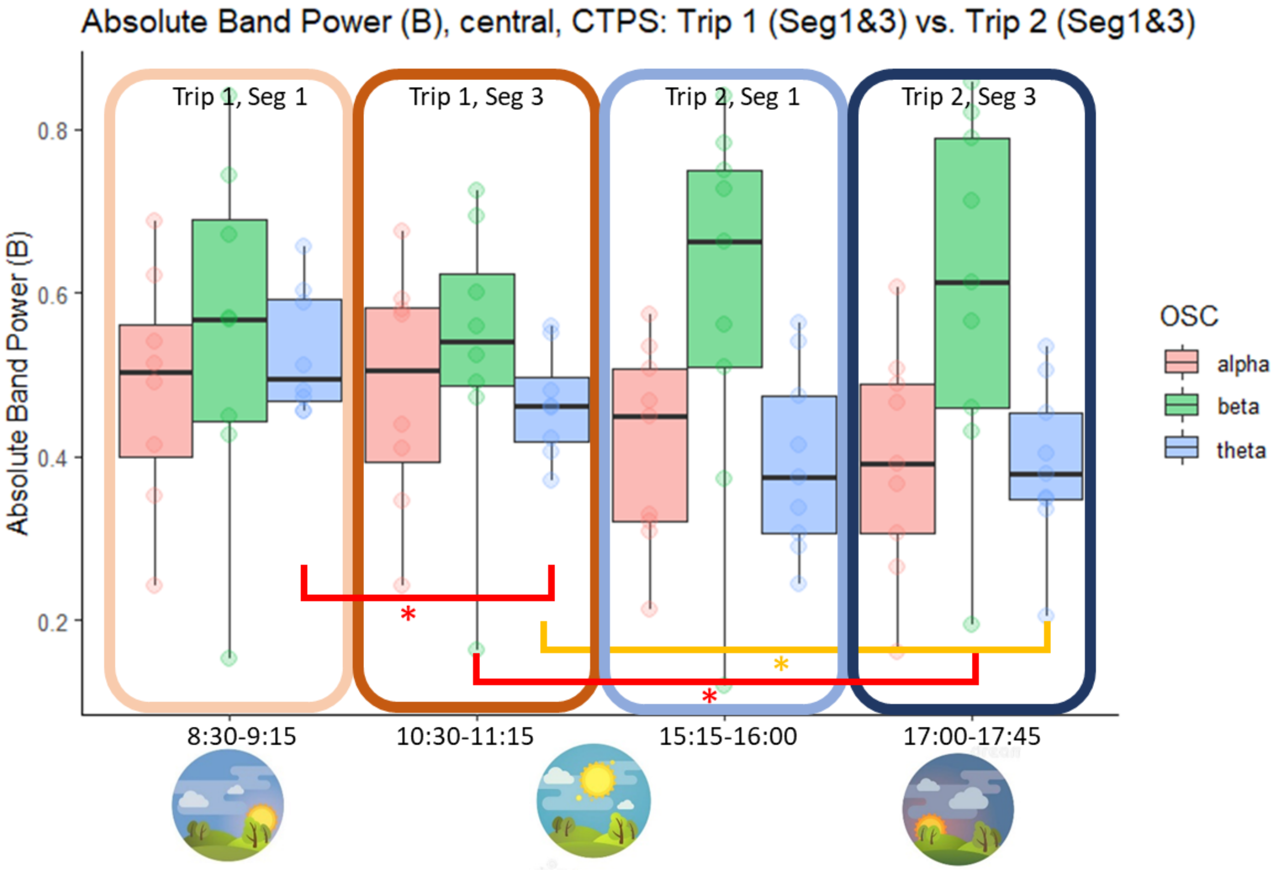


Figure 38: Changes in the brain oscillations from the central electrode. The absolute band power of theta oscillations in theta is significantly less at the end of Trip 1 compared to the beginning. At the end of Trip 2 compared to the end of Trip 1, the absolute band power of beta oscillations was increased significantly and there was a trend for decrease in theta.

8.5 Conclusion

The results of the EEG part of this study strongly suggest that a portable EEG system would be valuable to collect brain activity data in a real-world driving study. The electrode array could collect good quality data that was suitable for further analyses of frequency bands of interest to a fatigue, vigilance, and workload study.

As described above, and due to the low statistical power in this study, we are cautious in our interpretation of the EEG results. We were successful in collecting usable data in the alpha, beta, and theta frequency bands. The preliminary results show the absolute band power of alpha, and beta decreased during midday compared to the beginning of the trip. During the last part of Trip 2, these oscillations decreased. The theta oscillations followed a similar pattern, except in one instance. This is in line with previous studies showing that increasing fatigue is associated with an increase in alpha and theta, perhaps due to the participants attempting to stay vigilant when fatigued.

9 MULTI MODAL ANALYSIS

This Chapter will examine the drivers' pupil diameter and EEG data during two types of road events (1) vehicle cut ins and (2) hard breaking events.

9.1 Cut-ins Events

As cut offs are experienced daily by commercial truck drivers, we do not expect cut offs to impact drivers significantly. However, if there would be an impact, we would expect the cut-ins events to impact the follower truck driver the most. Based on the data provided by the traffic team, we identified the timing of cut-ins on the road. For each case (we investigated 15 events in total), we plotted the eye tracker and brain data, 10 minutes before each event to have a baseline of drivers' psychophysiological data, and 5 minutes after each event to investigate the impact of the event. Our visual inspection did not reveal any differences between data collected before and after the cut-in events. Figure 39 and Figure 40 provide a sample of collected data, before and after a cut-in event on 20 January 2022, 10:19 AM. No changes are observed in this data or in the other events that were reviewed.

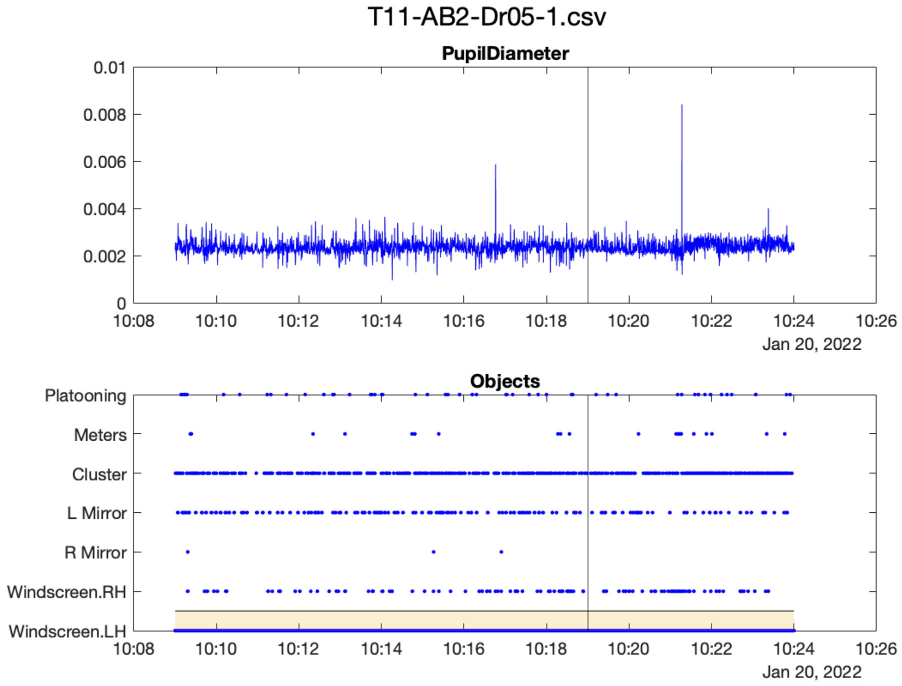


Figure 39: Pupil diameter, and location the driver was looking at while driving, before and after cut-in event at 20 January 2022 ,10:19 AM. No changes are observed on this data

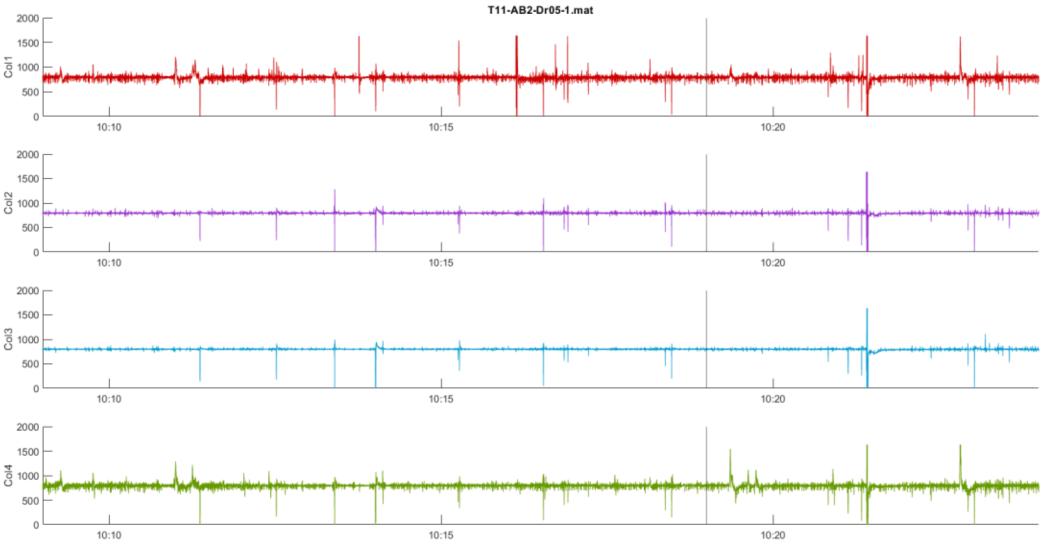


Figure 40: Raw brain data signals from each of the four Muse head band electrodes, before and after cut-in event at 20 January 2022, 10:19 AM. No changes are observed on this data

9.2 Hard Braking Events

Hard braking occurs when a driver presses the pedal quickly in response to an emergency (Ju et al., 2022). A study by Hernandez et al. (2018) has indicated possibilities of early detection of hard braking. One such way is using electroencephalography (EEG) signals which occurs early in the brain processes of triggering a hard brake (Ju et al., 2022). This can then be used in automated driving assistance systems to avoid accidents (Hernández et al., 2018). Furthermore, hard braking involves a complex series of cognitive and peripheral processes before using the physical pedal brake (Hernández et al., 2018). Thus, driving assistance systems can help create a faster braking response than drivers can make on their own (Hernandez et al., 2018). Despite the benefits, further research is still needed as EEG-based systems have their own short-comings such as low stability (Ju et al., 2022).

Moreover, hard braking is mostly associated with cognitive distractions (Khan & Lee, 2019). Studies that looked at cognitive distractions in driving did not find any increase or decrease in the frequency and duration of eye fixations (Khan & Lee, 2019). This may be the reason why the use of eye movements has not been incorporated to detect drivers' hard brakes (Hernández et al., 2018). However, it is important to consider these measurements as it seems to be correlated with fatigue or drowsiness which can potentially impair the early detection of braking (Hernández et al., 2018).

Hard brakes in the platooning system can be initiated either by drivers or the copilot system. In this section, the impact of the hard brakes by the copilot systems on the drivers will be investigated. The threshold of $a<-1.5$ m^2/s has been chosen for defining the hard brakes. Based on the data provided by the analytic team, we identified the timing of hard brakes initiated by the copilot system in either the lead or the follower truck. The eye tracker and brain data obtained 10 minutes before each event was considered a baseline of drivers' psychophysiological data. Changes in data was analyzed at the 5-minute interval after each hard brake event. The visual inspection did not reveal any differences between data collected before and after the hard brake events. In Figure 41 and Figure 42 a sample of collected data before and after a hard brake event on January 23, 2022, 15:38AM is shown. No changes are observed in this data or in the other events that were reviewed.

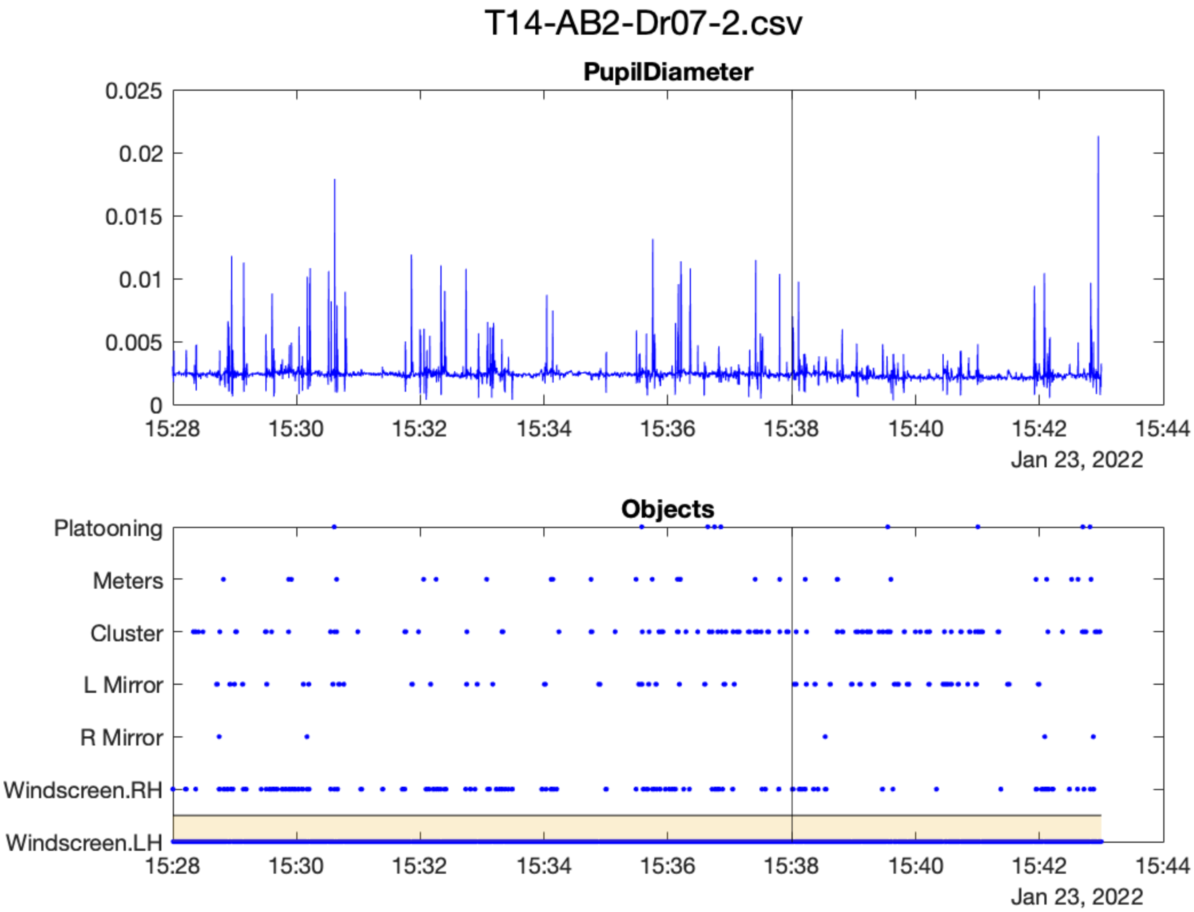


Figure 41: Pupil diameter, and location the driver was looking at while driving, before and after cut-in event on 23 January 2022, 15:38 AM. No changes are observed on this data

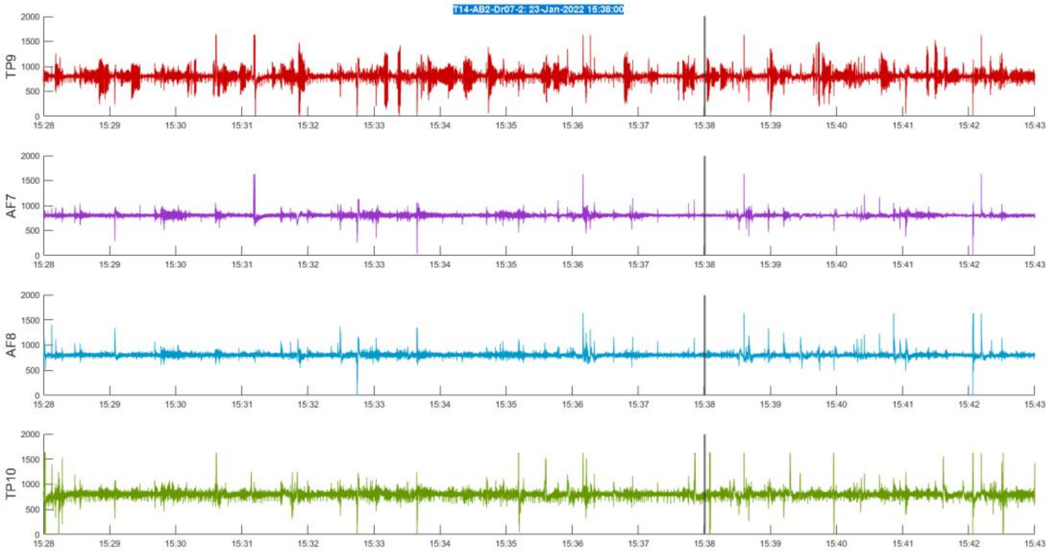


Figure 42: Raw brain data signals from each of the four Muse head band electrodes, before and after cut-in event on 22 January 2022, 11:56 AM. No changes are observed on this data

10 OUTLIER ANALYSIS

In this Chapter, two outlier cases during on-road trials: i) low platooning engagement, and ii) the Alberta trucker convoy protest were investigated.

10.1 Low Platooning Engagement

During the on-road trials, for two trips, the minimum threshold of 30 minutes for platooning was not met. These trips are classified as outliers, and in this Section, the confounding factors that may lead to low platooning engagement are examined. Outlier trips occurred over two days, that will be referred to as Day 1 and Day 2 to maintain anonymity of the drivers. The required platooning condition for Day 1 was the 5s platooning distance, and for Day 2 it was 3s and 5s platooning distances for Trip 1 and Trip 2, respectively. Table 7 gives a summary of the copilot and platooning engagement in each trip and the weather and road conditions.

During Trip 1 of the first day Driver B experienced low copilot engagement. On Day 2, both Trip 1 and Trip 2, Driver C activated the copilot system for less than 30 minutes.

Additional analysis was completed for the Trust in Automation (TiA), the Acceptance Scale for Advanced Transport Telematics (AATT), and the Driver Activity Load Index (DALI) questionnaires, (Modified Risk Perception Questionnaire (MRPQ), and Driver Stress Inventory (DSI) for each of these drivers to explore what factors may have contributed to low copilot engagement during these trips.

Table 7: Copilot and CTPS engaged durations and distances, disengagement error codes, weather, and road conditions for each outlier trip.

	Day 1				**Day 2**			
Trip	**Trip 1 (5s)**		**Trip 2 (3s)**		**Trip 1 (5s)**		**Trip 2 (3s)**	
Truck	**AB1**	**AB2**	**AB1**	**AB2**	**AB1**	**AB2**	**AB1**	**AB2**
Driver	**Driver A**	**Driver B**	**Driver A**	**Driver B**	**Driver B**	**Driver C**	**Driver B**	**Driver C**
Copilot engaged duration	01:15:02	00:08:23	00:48:18	00:41:50	01:51:24	00:18:17	00:44:12	00:21:48
CTPS engaged duration	00:03:55		00:33:29		00:03:01		00:02:15	
Disengagement error codes and numbers for AB2 Copilot	2 N/A 7 Steer mismatch 3 Poor lane markings		2 N/A 4 Steer mismatch 1 Poor lane markings		3 N/A 8 Steer mismatch		5 N/A 3 Steer mismatch	
Disengagement error codes and numbers for AB2 Platooning	4 N/A 4 Steer mismatch 1 Poor lane markings		28 N/A 4 Steer mismatch 1 Poor lane markings		8 N/A 3 Steer mismatch		2 N/A 1 Steer mismatch	
Weather	Overcast		Sunny		Overcast		Light snow	
Road conditions	Partly covered snow		Shoulder ice/snow		Bare dry		Bare wet	

10.1.1 The effect of time of day on brain oscillations

During Day 1 the road conditions were partly covered by snow/ice and the shoulder had ice/snow for Trip 1 and 2, respectively (Table 7). On Day 1 it was noted by the researchers that platooning was only possible during the first 45 minutes of the first trip due to bad road conditions. As the AB2 copilot and platoon engagement will not engage when there are bad road conditions (e.g., faded lane markings, shadows, etc.), this may have accounted for the low copilot and platoon engagement duration during the first trip of Day 1, when the road was partly covered with snow. There was only one other day with poor road conditions (partly covered in snow), but data for copilot and CTPS engagement is not available on that day. Therefore, the poor road conditions on Day 1 are a likely reason for the low platooning engagement during this trip. The weather and road conditions were unremarkable for Day 2, except for scattered flurries.

10.1.2 Modified Risk Profile Questionnaire

All drivers, except one, who did it before the training session, completed the MRPQ questionnaires at the introduction session, which was held in September 2021. Both Driver B and Driver C were tied for the second lowest score for the MRPQ, indicating both drivers had a very low tolerance for risk compared with the other drivers (Table 8, Figure 43).

Table 8: Modified RPQ (Risk Profile Questionnaire) questionnaire data for Driver's B, C, and average of all drivers. Lower scores on the RPQ mean less tolerance for risk.

	MRPQ
Driver B	0.18
Driver C	0.18
Average All	0.53

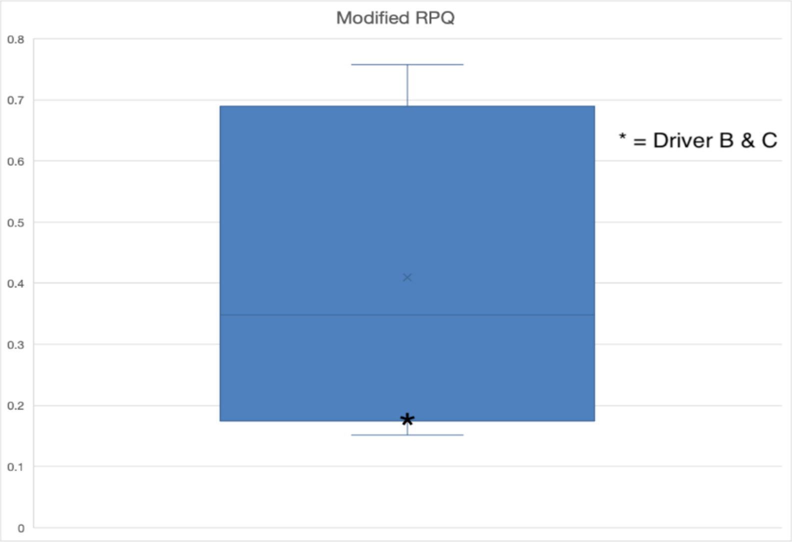


*Figure 43: Boxplot of the Modified MRPQ for all drivers' first platooning trip. Driver B and Driver C's scores are indicated with *.*

10.1.3 Driver Stress Inventory (DSI)

The Driver Stress Inventory (DSI) was collected during the introduction session (for all drivers except one), and before the training session. For the DSI, Driver B had higher scores in aggression, dislike of driving, and hazard monitoring compared to the other drivers indicating a higher stress level while driving (Table 9, Figure 44); however, Driver B did not answer 15/40 questions). Additionally, when asked “Do you worry when driving in bad weather” on the DSI, Driver B answered, “Very much” (the maximum), which likely contributed to his reluctance to engage copilot when driving on Day 1. Driver C was at approximately the average of other drivers for stress.

Table 9: DSI (Driver Stress Index) questionnaire data for Driver’s B, C, and average of all drivers. Higher scores on the DSI mean more stress while driving. Note: Driver B’s scores for Thrill seeking are not available as >30% of questions were not answered.

	Aggression	Dislike of driving	Hazard monitoring	Thrill seeking
Driver B	83	75	63	NA
Driver C	80	38	48	46
Average All	70.6	47.5	47.6	42.5

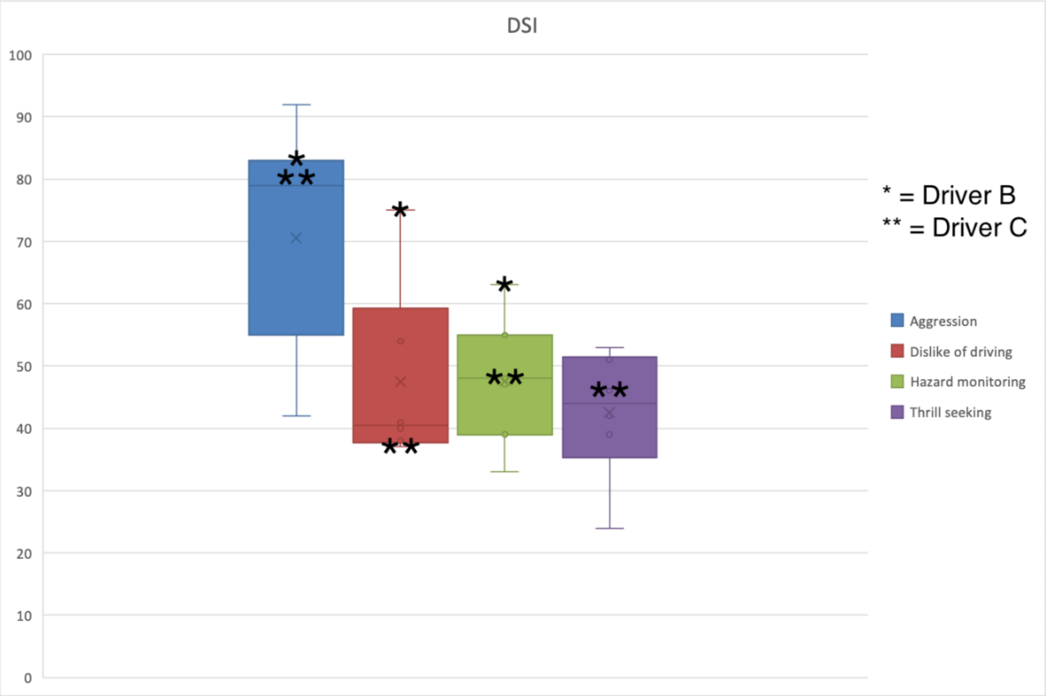


*Figure 44: Boxplot of the DSI for all drivers’ first platooning trip. Driver B and Driver C’s scores are indicated with *. Note: Driver B’s scores for Thrill seeking are not available as >30% of questions were not answered.*

10.1.4 Sleep quality and alertness

For both days, Drivers B and C reported their previous night’s sleep quality to be either good or very good and their alertness at the start of the trip to be either very alert or extremely alert based on KSS values (Table 10). This reduces the chance that acute fatigue played a role in their driving performance and decision making on these days.

Table 10: Pre Trip and Post Trip Sleep Quality and Karolinska Sleepiness Questionnaire (KSS).

	Day 1	Day 2	
	Driver B (AB2)	**Driver B (AB1)**	**Driver C (AB2)**
Pre Trip			
Sleep quality	Very good	Very good	Good
KSS	Extremely alert	Extremely alert	Very alert
Post Trip 1			
Sleep quality	Not available	Not available	Not available
KSS	Extremely alert	Extremely alert	Very alert

10.1.5 Trust in Automation (TiA)

Drivers completed the Trust in Automation (TiA) questionnaire during Post Trip 2 sessions. Driver B only completed one question on the TiA questionnaire on Day 1 due to an anomaly of the survey monkey website, they were able to skip answering all the questions in a section by answering only one question. Therefore, it is not possible to know if trust played a role on that specific day. Driver B's TiA scores were averaged for the other four days he drove and were compared with the average of all other drivers' scores. A two-tailed t-test (assuming unequal variances) was conducted for each of the six TiA categories. Driver B had significantly lower scores in the Reliability/Competence and the Understanding/Predictability sections of the TiA compared to the other drivers (Table 11, Figure 45). These are two of three factors of perceived trustworthiness (Körber, 2019), and low average scores in these areas suggest Driver B may have had a lower degree of trust than the other drivers in the platooning technology. Interestingly, when all factors were considered, Driver B's total average score on the TiA scale was not lower than other drivers.

There was no significant difference between scores of Driver C and other drivers.

Table 11: Trust in Automation (TiA) questionnaire data for each driver. The higher the score, the more trust one is perceived to have.

	Reliability/ Competence	Understanding /Predictability	Familiarity	Intention of Developers	Propensity to Trust	Trust in Automation
Day 1 Driver B	NA	NA	NA	NA	NA	NA
Day 2 Driver B	9	8	4	8	6	6
Day 2 Driver C	24	13	NA	8	9	6
All drivers except Driver B (Average ±SD)	19.20 ± 5.21	19.20 ± 5.21	19.20 ± 5.21	19.20 ± 5.21	19.20 ± 5.21	19.20 ± 5.21
Average Driver B	11.5	9.25	3.5	8	6	5.25
Average Driver C	22.8	14	3.6	8	9.4	7.8
p-val (t-test Driver B and other drivers)	<.01	<.001	0.24	0.38	0.22	0.41

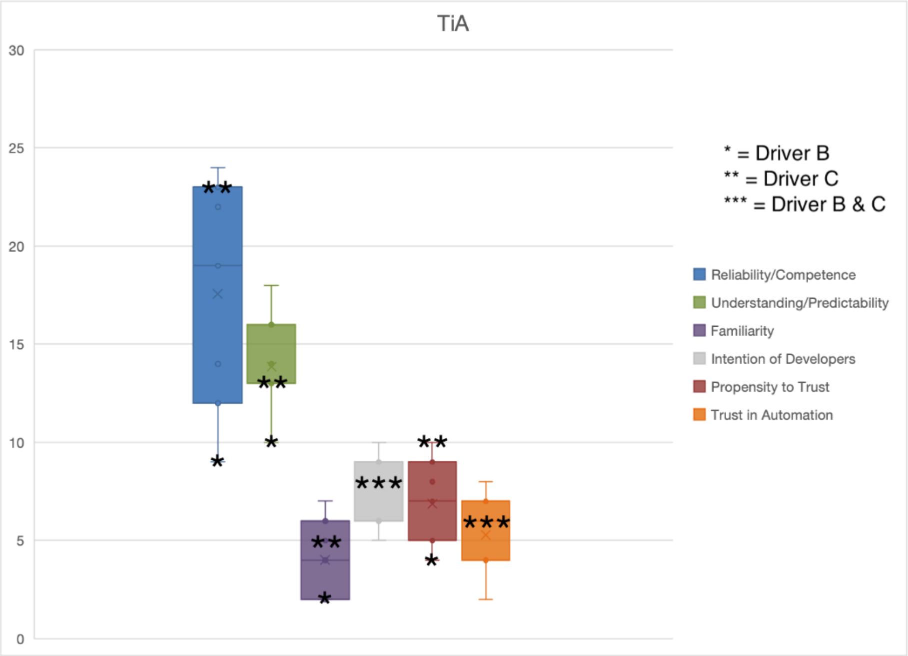


*Figure 45: Boxplot of TiA for all drivers' first platooning trip. Driver B and Driver C's scores are indicated with *.*

10.1.6 Acceptance Scale for Advanced Transport Telematics (AATT)

All drivers were instructed to fill out the Acceptance Scale for Advanced Transport Telematics (AATT) questionnaire during Post Trip 2 sessions. Driver B had very low average scores in the AATT questionnaire compared to the combined average scores for other drivers, suggesting Driver B did not perceive using the platooning technology as useful or satisfying (Table 12, Figure 46). There was no significant difference between scores of Driver C and other drivers.

Table 12: Acceptance Scale for Advanced Transport Telematics (AATT) questionnaire data for each driver. The more negative a score the less useful and/or satisfying the technology was perceived to be, the more positive a score the more useful and/or satisfying the technology was perceived to be.

	Usefulness	**Satisfying**
Day 1 Driver B	-2	-2
Day 2 Driver B	-2	-2
Day 2 Driver C	1.8	1.5
All drivers except Driver B (Average ±SD)	0.58 ±1.25	0.175 ± 1.29
Average Driver B	-2	-2
Average Driver C	1.8	1.6
T-test (b/w Driver B and others)	<.05	<.05

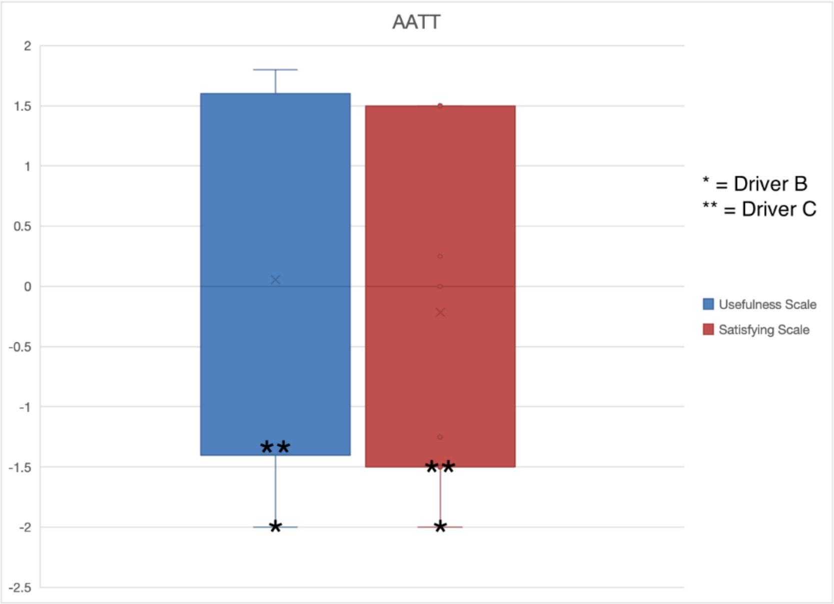


*Figure 46: Boxplot of AATT for all drivers' first platooning trip. Driver B and Driver C's scores are indicated with *.*

10.1.7 Driving Activity Load Index (DALI)

After each trip (Post Trip 1 and Post Trip 2 sessions), driver responses to the Driving Activity Load Index (DALI) questionnaire were acquired. Driver B reported high levels of stress and attentional demand during all drives (Table 13, Figure 47). A two-tailed t-test (assuming unequal variances) was conducted for each of the six DALI categories, and Driver B's scores were significantly higher than the averages of all other drivers There was not a significant difference between the scores of Driver C and other drivers.

Table 13: Driver Activity Load Index (DALI) questionnaire data for each driver. Higher scores mean more effort/stress was perceived.

	Effort of Attention	Visual Demand	Auditory Demand	Temporal Demand	Interference	Situational Stress
Day 1 Driver B	10	8	6	10	10	7
Day 2 Driver B	10	10	6	10	6	10
Day 2 Driver C	5	6	4	4	3	6
All drivers except Driver B (Average ±SD)	5.3 ± 2.72	5.25 ± 2.73	3.65 ± 1.81	3.65 ± 2.32	3.7 ± 2.13	4.5 ± 2.28
Average Driver B	10	9.6	7.6	10	6.8	9.4
Average Driver C	4.4	5	2.4	4.2	2.4	4.2
T-test (b/w Driver B and All - Driver B)	<.001	<.001	<.01	<.001	<.05	<.001

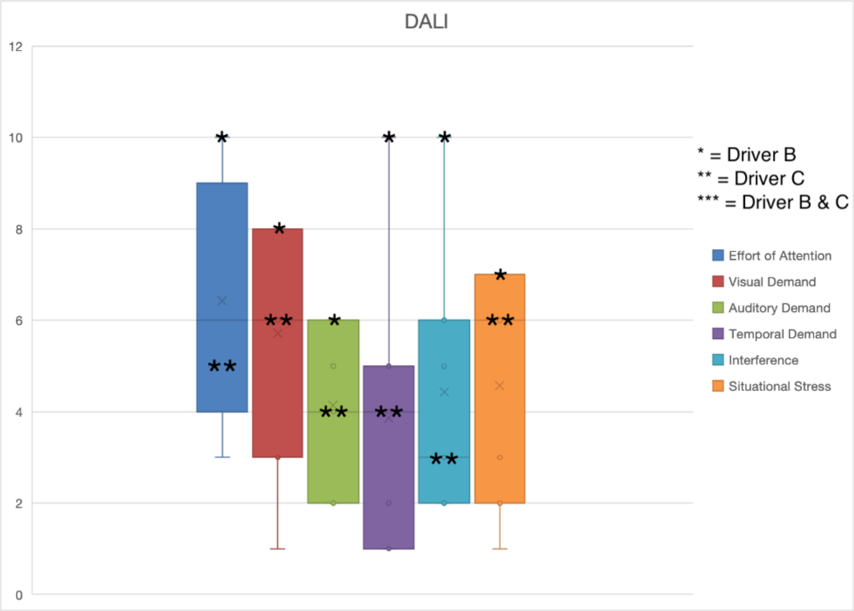


*Figure 47: Boxplot of DALI for all drivers' first platooning trip. Driver B and Driver C's scores are indicated with *.*

10.1.8 Conclusion

Table 14 shows a summary of Driver B's and C's responses on various questionnaires compared to the other drivers. It should be noted that for both Driver B and C, Day 1 and Day 2 respectively, were the first days experiencing platooning.

On Day 1, the combination of poor road conditions, Driver B's decreased trust in the platooning technology, personality (low risk tolerance, increased stress while driving), and the fact it was Driver B's first trip likely contributed to the low copilot and CTPS durations during Trip 1 (5s interval), with poor road conditions likely being the most impactful. Taken together, Driver B had a more stressful experience, which may have led to a more negative perception of using the technology.

On Day 2, Driver C's scores on the TiA, AATT, and DALI for that day and on average for Driver C were unremarkable when compared with the average for all drivers. The road conditions and weather were also unremarkable except for a few scattered snow flurries which are not unusual events for these drivers at that time of year. However, Driver C's risk tolerance was lower compared to the other drivers. Driver C also observed the steering in AB2 pulled to the right, so this may have influenced his comfortability in having copilot engaged.

Drivers B and C both engaged the copilot and platooning systems more and similar to other drivers in other days. It can be concluded that these drivers needed more time to get used to the platooning technology and particularly for Driver B, how it functions under different conditions e.g., deteriorating road conditions due to weather factors.

Table 14: Summary of drivers' B and C scores compared to the group in various categories

Responses	Driver B	Driver C
Sleep quality	Good to very good sleep	
Alertness	Very alert to extremely alert	
Trust in Automation (TiA)	Compared to the other drivers, had significantly lower scores in ● Reliability/Competence ● Understanding/Predictability	No significant difference compared to other drivers
Acceptance for Advanced Transport Telematics (AATT)	Compared to other drivers, had significantly lower scores when rating the automation as either ● Useful ● Satisfactory	No significant difference compared to other drivers
Driver Activity Load Index (DALI)	During all trips, reported high levels of ● Stress ● Attentional demand	No high levels recorded
Driver Stress Index (DSI)	Compared to other drivers, ● Recorded highest score for stress while driving ● Youngest and least experienced driver When asked, "Do you worry when driving in bad weather", indicated ● Very Much	Average stress responses in line with other drivers
Modified Risk Profile Questionnaire (MRPQ)	Both Driver B and Driver C tied for the second lowest score for the RPQ, indicating both drivers had a very low tolerance for risk compared with the other drivers.	

10.2 Alberta Trucker Convoy Protest

On January 29, 2022, the Alberta trucker convoy protest affected the ability for the drivers to complete their scheduled delivery. Around Didsbury Alberta, heavy road traffic began to cumulate along Highway II. By early morning, a heavy traffic warning and travel not recommended advisory was issued by 511 Alberta. The increased traffic and heightened emotions of the surrounding drivers on the Queen Elizabeth II highway northbound resulted in the driver's inability to complete the planned platooning trip on this day. Both drivers executed their right to remove themselves from dangerous road conditions and returned to Calgary in advance of reaching the midpoint of the trip.

Heavy traffic and heightened emotions of surrounding drivers enabled the research team to see if data displayed a difference in the way the two drivers (Dr07 and Dr09) responded during this trip through differences reported on the Drivers' Active Load Index (DALI) questionnaire after this trip compared to other trips.

The results of the Driver Activity Load Index (DALI) questionnaire for this baseline trip, and all baseline trips that were conducted during the on-road trials are presented in Figure 48. Overall, based on the scores from the DALI, it appears that Dr09 experienced an elevated level of temporal demand, interference, and situational stress during that day.

As the number of baseline trips in the platooning trials was limited, we also compared the drivers DALI performance on this day with all CTPS trips (Figure 49). When analyzing responses to the interference question "To what level were you distracted by tasks other than driving such as responding to alarms, communications, road construction, etc...?", we found that Dr09 rated significantly higher levels of interference compared to the other drivers during the CTPS trips (his score was above 1.96 standard deviations from the mean, 7.84). As such, we conclude that the convoy protest day event induced a greater need to check and use communication devices. Of note, the drivers were willing to cancel the trip and get off the road on that day, which likely reflects their increased level of discomfort.

To determine the impact of this day on the drivers, we sought to determine whether the two participants (Dr07 and Dr09) responded differently to the Drivers' Active Load Index (DALI) questionnaires compared to other days.

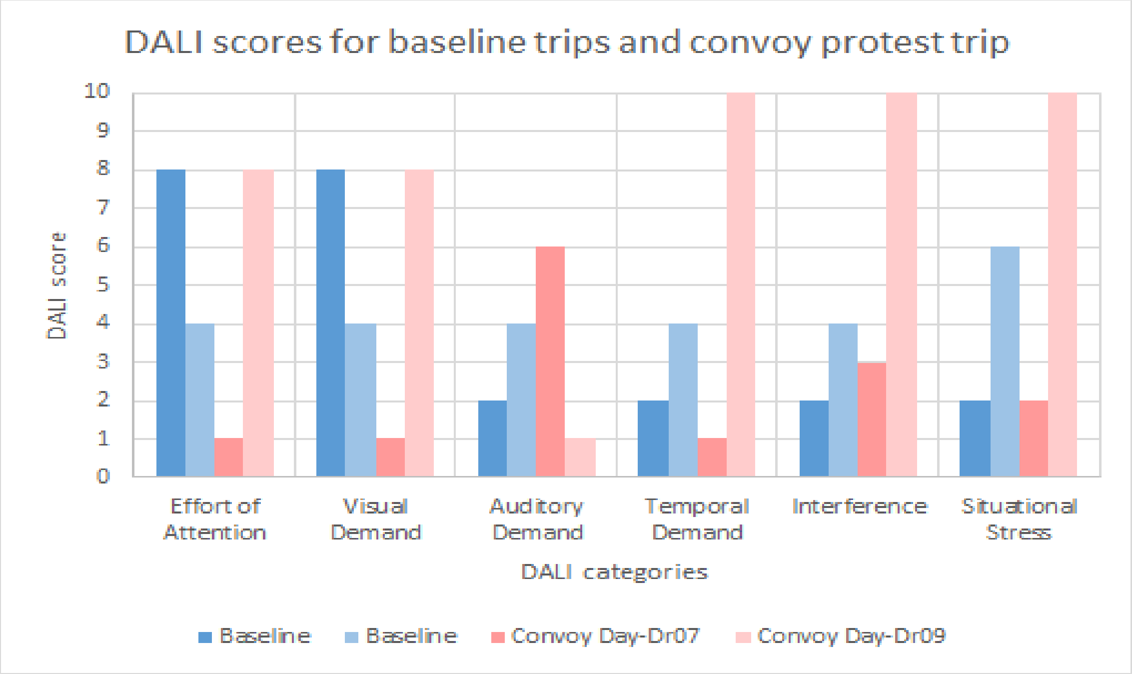


Figure 48: DALI scores for all baseline trips in blue shades, and Dr07 and Dr09 from truck drivers convoy protest day in pink shades. Higher scores indicate higher levels of stress/demand in each category.

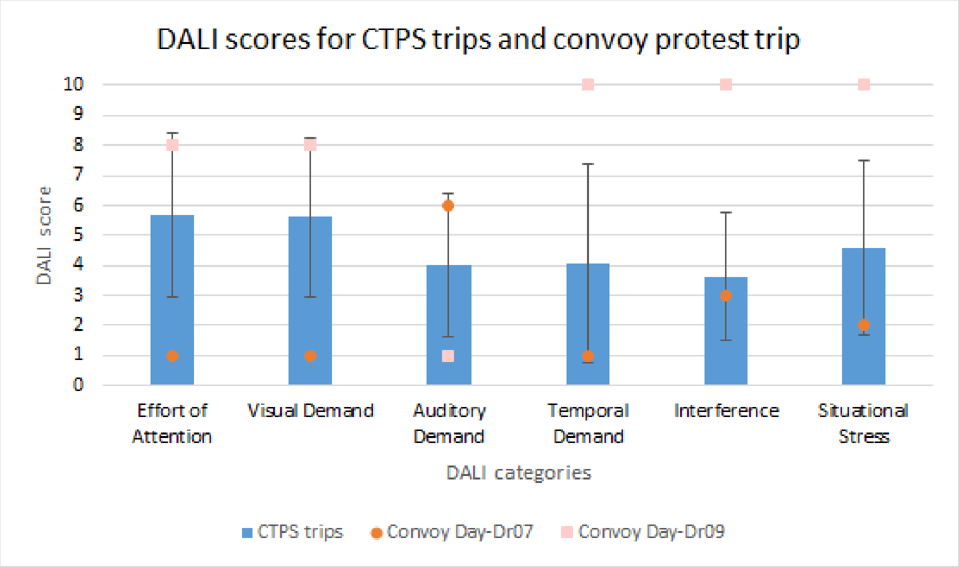


Figure 49: DALI scores for all CTPS trips, with Dr07 and Dr09 indicated circle and square markers. Error bars represent 1 standard deviation, and higher scores indicate higher levels of stress/demand in each category.

11 SUMMARY

This study was successful in showing that real-time psychophysiological data could be collected from professional drivers while conducting real-world, on-road trucking operations on a major highway. This is an important outcome of our study. Further, the administration of discretely collected questionnaire data and tablet-based cognitive task data was possible at two locations at different times in the day. Many logistical challenges presented due to the trucking schedules and the normally high driver workload. Nevertheless, after some initial trips, our team was able to conduct a smooth study that accommodated the needs of the participants. As described previously, this study became exploratory in nature, and it was not able to fully address the original hypotheses about fatigue, vigilance, and workload during on-road CTPS driving. To somewhat address the 8th original hypothesis, we compared Trip 1 and Trip 2, and within each of these trips to examine the impact of time of day and working hours on drivers. Table 15 contains the summary of the multi-modal findings in this report.

Overall, we think the results of our study failed to capture the effects of platooning on the drivers' experience. The inclusion of an instructor/trainer in each truck during the trials added an unexpected challenge. This was done to increase safety and facilitate a greater sense of familiarity with the platooning system and accommodate drivers' concerns about not having enough training time in the trucks. While necessary, it altered our study because it is known that passengers in vehicles change the nature of the driver's attentional focus. This project scope alteration could certainly impact the vigilance of the driver as well as cognitive workload factors, and the driver's strategies. We cannot differentiate across these things from our data because there was a passenger present in each of the trucks for every trip. These challenges notwithstanding, we successfully adjusted and fine-tuned a complex methodology that could be valuable for future real-world commercial driving studies interested in examining psychological measures and human factors. Below are the main conclusions from each of our primary dependent measures.

- **Questionnaires**
  Except for E&C measurement tool, which was designed specifically for application within this study, all the questionnaires and accompanying scales used in the study show high validity and reliability. Correlation with physiological measures in future studies will greatly enhance their applicability in helping researchers understand drivers' cognitive engagement, alertness, and passive vs active fatigue in CTPS. Insight from these subjective measures also identified the need for greater consideration of macro ergonomics prior to soliciting and engaging drivers (e.g., schedule design, length of research days, impact on worker and home life, driver training requirements, researcher training requirements, communication structures, etc.) and micro ergonomics (e.g., cab design, seat features, workloads, etc.).

- **TBCT**
  The TBCT consisted of a battery of 4 cognitive tests that have been extensively used to assess fitness to drive in the context of aging, drug use, and various workplace stressors. We chose to include these measures in the present study to complement the continuous monitoring data and the questionnaire data. The main idea was to use the TBCT as probe tasks at three time points during each return trip to assess the overall cognitive functioning of the drivers and to view the impact of all the (original) conditions on cognitive performance. For example, if the full return trip of platooning affected one aspect of cognitive testing more than half the return

trip, or compared to a different condition, then we could infer that the effects of that platooning condition had some specific impact on cognition. It would have been an indirect measure of the effects of platooning on specific aspects of cognition, with the ability to observe a range of cognitive behaviors. Based on the length of the daily activities and driver advice, we decided to drop one of the data collection point. We proceeded with TBCT testing at the beginning of the trip in Calgary and again at the end of the return trip back in Calgary. The mid-point in Edmonton was dropped to reduce the increased workload and time added to the drivers already full day. A future study might attempt to shorten a cognitive test battery to accommodate a third testing time without adding substantial time to the drivers' workday. The TBCT tested for simple reaction time, spatial attention and attention switching, memory, and visual-motor control. All these aspects of cognition are widely studied and are of great interest to the cognitive science community. We were unable to examine the effects of the different platooning conditions due to being underpowered and having an unbalanced design. The main preliminary finding was that there was sensitivity in the spatial-attention RT task to the long workday, but the other measures were predominantly insensitive. This was an important finding at this stage because it could have implications for future studies.

The attention task examined low-level cognitive behavior related to spatial attention shifting, and the other tasks examined higher-level, more consciously controlled aspects of cognition such as explicit memory, decision making, and action. We offer two perspectives on these results while keeping in mind that we did not have the right statistical power to make firm conclusions. The first perspective is that the driving experience, collapsed across the different platooning conditions affected low-level (yet critical) cognition more than it did higher-level cognition. One interpretation is that the drivers could exert more conscious effort to maintain performance on the higher-level cognition tests than they could on the low-level tests that are difficult for participants to know are happening. Thus, a future study might choose to further probe this phenomenon with a finer set of tests including other low-level tasks that are more properly timed for the drivers to feel comfortable to perform. The second perspective is that as the literature suggests, the workload associated with driving impacts consciously controlled cognition such as working memory and selective attention. This perspective goes with the idea that the study was simply not strong enough to make the first conclusion, and it remains to be seen if different platooning conditions would significantly affect other aspects of performance on tablet-based cognitive tasks. In another study we have observed that the motor-control task is most predictive of on-road performance for commercial drivers (unpublished data). It is important to note that the driving scenario in the referenced study was different than the SAE level 2 driving scenario in the present study.

- **Eye-tracking Data**
  Real-time eye tracking data was continuously collected as a measure of driver fatigue and vigilance. In reviewing a large body of evidence, we considered eye-tracking to be a robust measure of the drivers' experience. As will other psychophysiological measures, eye-tracking data requires stronger statistical power than we were afforded. However, we were able to make two interesting observations from the eye tracking data.

First, the dash mounted trackers worked adequately. This is important because it suggests that good quality eye-tracking can be collected in a real-world on road trucking study. Our interpretation of the quality of the data was that it is as good as data collected in the laboratory on a stationary table. There seemed to be some loss in the quality of gaze direction data from Seg 1 to Seg 3, but it remained measurable.

Second, the eye tracking data showed a clear sensitivity to the amount of ambient sunlight in the truck. That is, the diameter of the drivers' pupils was sensitive to the low light of early morning (larger diameter) and the brighter light of later in the day (smaller diameter). This is encouraging because pupil diameter is well known to correlate with aspects of fatigue and vigilance. Thus, our findings suggest that dash-mounted eye tracking has the strong potential to be a valuable experimental tool in future studies of this nature. We could not make strong inferences about the effects of platooning on eye movement behavior or pupillometry. However as further discussed in the conclusion section, we see value in extending the exploratory analyses to probe for some non-statistical phenomena to potentially guide future research.

- **EEG Data**
  Comfortable ambulatory EEG technology is relatively new and has not been extensively tested in real-world skilled performance scenarios like commercial trucking over many hours. The EEG measures that we collected are the measures in this study that require the highest number of participants. Laboratory EEG studies typically have more than 20 participants to obtain statistical significance, and in many cases the number can exceed 50 participants. However, an important EEG finding from this study was that a portable EEG system is potentially viable for future studies.

  We showed that good quality data can be collected in a real-world on-road trucking study, and that oscillation frequencies known to be associated with fatigue and vigilance can be collected and quantified. The electrode array was sufficient to collect the data over the length of the trial, but the lateral electrodes showed less connection with slightly more loss of signal over time compared to the central electrodes.

  We observed some interesting alpha, beta, and theta results when comparing early and later parts of the trips. Most of the findings are in line with what would be expected. Particularly beta oscillation power was significantly less after driving for 2 hours compared to the early in the trip. This was also a trend-level finding for alpha. For the most part, the theta oscillations followed a similar pattern. Interestingly, we observed a decrease in alpha and theta during Trip 2 compared to Trip 1, but an increase in beta. Since beta increased, the decrease in the other two frequencies is unlikely related to some systemic problem with the technology. Rather, the EEG system was sensitive to recording differential effects in the EEG. Overall, these observations are in line with previous studies showing that fatigue can be associated with changes in alpha, beta, and theta, perhaps due to changes in arousal, or the participants attempting to stay vigilant when fatigued. We are encouraged by these results and think it would be valuable to study fatigue in the context of EEG while driving more thoroughly.

Table 15: Summary of the findings

Data	Task	Measure	Comparison	N	p-val	Change
Questionnaire	DALI	Distraction	Post Trip 1 vs Post Trip 2	6	.042	Increased
Questionnaire	TiA	Understanding/Predictability	Post First Trip vs. Post Last trip	7	.089	Decreased
Questionnaire	TiA	Familiarity	Post First Trip vs. Post Last trip	7	.098	Increased
TBCT	Reaction time	Reaction time	Pre Trip vs. CTPS Post Trip 2	7	.067	Increased
TBCT	Judgement	Reaction time for stage 2	Pre Trip vs. CTPS Post Trip 2	7	.068	Decrease
TBCT	Motor Control	Time on the left edge	Pre Trip vs. CTPS Post Trip 2	7	.044	Decreased
Eye Tracker		Gaze Direction Quality	AB1, Seg 1 vs Seg 3	68	.046	Decreased
Eye Tracker		Gaze Direction Quality	AB2, Seg 1 vs Seg 3	44	.014	Decreased
Eye Tracker		Pupil Diameter	Trip 1 Seg 1 vs Trip1, Seg 3	8	.049	Decreased
Eye Tracker		Pupil Diameter	Trip 2 Seg 1 vs Trip1, Seg 3	8	.027	Increased
Eye Tracker		Pupil Diameter	Trip 1 Seg 3 vs Trip 2, Seg 3	7	.068	Increased
Brain data		Good electrode-to-skin connection	Seg 1 vs Seg 3	35		No difference
Brain data		alpha Oscillations (lateral)	Trip 1 Seg 1 vs Trip1, Seg 3	7	.051	Decreased
Brain data		betta oscillations (lateral)	Trip 1 Seg 1 vs Trip1, Seg 3	7	.019	Decreased
Brain data		alpha Oscillations (lateral)	Trip 1 Seg 3 vs Trip 2, Seg 3	6	.083	Increased
Brain data		theta Oscillations (central)	Trip 1 Seg 1 vs Trip1, Seg 3	8	.047	Decreased
Brain data		betta oscillations (central)	Trip 1 Seg 3 vs Trip 2, Seg 3	6	.045	Increased
Brain data		theta Oscillations (central)	Trip 1 Seg 3 vs Trip 2, Seg 3	6	.062	Decreased

12 CONCLUSION, LESSONS LEARNED, & FUTURE WORK:

This was an appropriately complex and ambitious human factors study designed to test several important hypotheses about commercial driver fatigue during various conditions associated within a cooperative truck platooning system. The multi-modal approach was the first of its kind in a real-world driving scenario on a major Canadian highway route. We were not able to systematically address the originally planned main hypotheses, however we gained value through significant accomplishments that evaluating the driver experience during this shortened and underpowered exploratory study. There is valuable future work to be done with the current data sets through additional studies. A higher powered st`udy with additional participants will be required to conduct a complexed human factors study. Below we outline some important lessons that were learned and describe future work to be undertaken.

12.1 Lessons Learned and Areas for Improvement

It is exciting to be a significant part of a larger collaborative project. Diligence is required to effectively manage the many competing interests of numerous research teams and project partners. We are fortunate that the larger team collaborated well and communicated effectively. The lessons we learned are not intended to be highly critical of any aspect of this complex and challenging study. Rather, we offer them to provide value beyond what our scientific results show.

- It would have been advantageous to the entire team if the human factors team had been more involved during the pre-planning and project planning stages of the project's life cycle. Additional focus on critical health, safety, and performance consequences could be built into the project design. These additions may have resulted in better communication and integration of project teams with less revisions, adaptations, interruptions and re-work due to unanticipated system and process flaws.

- There were system reliability factors identified by the drivers that were classified as "training" issues. Training, while necessary, is the lowest form of control to prevent human error. A human factors evaluation of the trucks and equipment prior to driver testing could have identified critical training and other mitigation strategies. A summary of noted human-machine interface factors that complicated this project can be found in Appendix A.

- Driver considerations for additional pilot testing:
    - We would have benefitted from further discussions with them about the importance of confidentiality and how sharing their experience with other drivers could negatively bias the study. Driver concerns might create bias impacting other drivers' views going into the project. For example, in some cases, the drivers talked with other drivers about their experience in the trials, which may have influenced the other drivers.

    - When highly skilled participants are involved in a study, we will often recommend that the trial participants to be involved in study planning. Participants were introduced to the researchers at the early recruitment sessions and drivers did not participate in the study design.

  - Talking and learning from the drivers often ended up being a part of data collection, which was not ideal. We learned about their reservations and comfort levels while we were engaged in data collection. This resulted in last minute changes mid-stream, led to inconsistent trials between trips. Having more time to work with some drivers ahead of recruitment, and additional pilot testing on-road would have helped fine-tune things considerably.
  - The impact on the drivers' personal lives, home and lifestyle could have been more thoroughly considered. Additional time was added to an already full day and in some cases, drivers participating the trials were uncomfortable with the increased length of their workday.
- We learned that the shake-down trials were a very critical stage, and we would have benefited from more time to analyze the data to catch glitches and issues as listed below:
  - Occasionally drivers' responses to some questions were not collected for the Pre Trip or Post Trip questionnaires on Survey Monkey. We learned it was possible to skip questions in a block matrix by answering only one question in that block matrix.
  - Some drivers moved the sliding bar for the AATT and put it back to the initial position, which was coded as “no response”.
  - The EEG electrode conductance was inadequate for some of the initial trips. We learned to improve this by cleaning the skin more, as well as cleaning the headband with a water based facial cleanser, and more finely adjusting the headband or changing the headband altogether.
  - The on-road trials took longer than we originally discussed with drivers. This resulted in the drivers being somewhat dissatisfied and wanting to drop one of the TBCT sessions. More shakedown trips would have informed us more about these issues earlier so we could address them before the full study trial began
- Researcher training: Pre and Post driving sessions involved many steps, including administering cognitive tasks, design of questionnaires, implementation of an online survey collection tool, power up/down checklist equipment setup, calibration of equipment, and monitoring of driver compliance e.g., wearing the Fitbit. This project would have benefitted from additional training time with the researchers, ideally during the shakedown trials. We also would have benefitted from developing a systematic quality control process to ensure all activities met necessary standards.
- Scope changes and project amendments that occurred late December resulted in a condensed trial that started in early January, ahead of the anticipated schedule.
  - Fitbits were given to the drivers during their driver training. Drivers were to be notified 2 weeks before their first driving session, reminding them to wear their Fitbits each day before and during the trials. Often driver schedules were confirmed the day before they participated in their driving sessions, and it was not realized until they arrived that they

were not wearing their Fitbits. This resulted in losing important sleep schedules data to support data collection.

- Heart-rate variability data from the Movisens was not collected as planned. The originally identified and tested equipment became unusable when the manufacturer filed for bankruptcy in late 2021. Replacement equipment was ordered but unfortunately not received in time for the rescheduled trials.

- COVID illnesses affected student data collection. Full data sets were to be collected in both Edmonton and Calgary. Calgary coverage was reduced due to COVID and conflicting personal schedules restricted the available students' ability to "cover" absent staffs data collection. This was resolved as soon as possible.

12.2 Future Work

The work accomplished in this study suggests that future multi-modal human factors studies are feasible in real-world, on-road environments. Moreover, based on the work done to date, we are planning three future activities.

1) We will continue to analyze all the data collected in this study. Currently we have focused on data quality and examining each data modality in isolation. We have the opportunity to examine correlations between modalities, and in the context of data from the other research teams. We think it will be interesting to further examine the drivers' responses to on-road events, such as cut-ins. This type of inquiry would be more of a single trial analysis that would not reach statistical significance but could provide meaningful practical significance. For example, we know from visual inspection that the EEG frequencies (alpha, beta, & theta) were not influenced by sudden on-road events. However, we have not examined the overall EEG signal for spikes, which can result from the occurrence of unexpected events. At this stage we were not able to identify the presence or absence of specific EEG responses (e.g., P300) with a visual inspection, but this would be a focus in future studies and analyses. The eye movement data will also be of interest to more closely inspect in these kinds of situations. From visual inspection there does not seem to be novel eye movements occurring from the on-road events. Our team has experience with machine learning approaches to analyze multi-modal data, which can sometimes provide insights into relationships between variables that traditional analyses cannot yield. Extensive collected data will be suitable for such an analysis.

2) Although we were not able to conduct the full study as originally planned, what we did accomplish has scientific value as a proof-of-concept study. Because this study was the first of its kind and involved real-world highway commercial driving, other transportation and human factors researchers will be interested to know about our methods. As such, we are planning to write and submit a methodological paper about this study for publication in a peer-reviewed journal. We hope to further collaborate with Transport Canada on this work.

3) We are planning a validation study of the EEG data. Thus far our analyses of the EEG data made use of the proprietary software from Muse to visualize the data. Our laboratory has its own internally developed EEG software based in MATLAB that we will use to further analyze the data. It will be interesting to compare the results across the two analysis systems. More importantly, it will be valuable to compare the data collected via the Muse with data from a larger, more robust EEG system. We plan to conduct an in-lab study with our driving simulator in conjunction with our 256 electrode EEG system that has been used to identify the components of interest in numerous peer-reviewed publications. While simulated driving is not the same as on-road driving for many reasons, examining EEG in a long duration simulated driving condition could yield data that could be compared against the Muse data to further understand the on-road data outcomes. We may even be able to reproduce the initial experimental structure in the simulator and assess the impact of platooning conditions (e.g. platooning interval, driving role, automation level), and also the impact of sudden road events (e.g. cut-ins and hard brakes) to address some questions related to fatigue and vigilance in that proposed study with appropriate sample size.

Appendix A: Human-Machine Interface Challenges

The following notes are based on driver interviews and personal observations by the human factors team. While software updates were implemented after these observations were reported, the majority of issues were not resolved and driver perceptions did not improve.

System Reliability

- Harsh braking occurs in the follower truck when the lead and following vehicles are too close upon Copilot engagement and when close cut-ins and cut-offs occur. Concerns have been voiced about potential hard braking when snow or ice are encountered on the road surface.
- Cut-ins can cause large gaps between the lead and follower trucks reducing potential for successful platooning. A yo-yo effect occurs when vehicles enter and leave the space between the vehicles. The follower vehicle decelerates when the cut in occurs and speeds up to resume the set time gap when the vehicle leaves.
- Excessive platooning disengagements were experienced on the Strathmore route. During the Edmonton to Calgary route the platooning seemed stable.

Driver Distractions

- There is a general lack of familiarity with location of the vehicle buttons (e.g., 4-way flashers) requiring increased active driver engagement. Unfamiliarity with the positioning of vehicle controls can affect the resting state and driver's comfort level with the truck.
- The Copilot engagement button is located in a position that requires the driver to look around the steering wheel when engaging the system. Poor ergonomics in terms of user interface are an identified safety concern, as the eyes are no longer looking at the road.
- Drivers are concerned that when searching for the Copilot engagement button they may set off the Bison Guardian system that will place a visual distraction mark on the driver's record.
- Divers are uncomfortable with the inward cameras and find these and the tracking systems to be intimidating and intrusive. Driver behavior is affected as they are worried about other's perception of their driving, thereby reducing their comfort level with the system.

Driver Trust

- When the driver presses the button to engage the system there is a lag in the audio communication. This lag creates a decreased level of feedback on the status of system.
- The Copilot system actively centers the vehicle. Centering movements by the steering wheel can visually distract the driver.
- To slow down or decrease the vehicle speed drivers often remove their foot from the accelerator and allow the vehicle to naturally decelerate. Copilot technology slows the vehicle down by braking and each braking event activates the vehicles brake lights. There is concern that the continuous use of the brake pedal (for a cut-in/cut-off/hill deceleration) and corresponding brake lights may affect the trust and behavior of traffic surrounding the

platooning vehicles and consequently affect their driving behavior in response.

- When the driver adjusts the vehicle speed up or down there is no immediate visual, auditory, or tactile response. This lack of communication can lead to driving distractions and reduced trust in the system.
- Drivers who drive the Calgary to Edmonton run are paid by the kilometer, not by the hour. This payment schedule causes the drivers to evaluate the system from the perspective of how it will impact them both financially and timewise. Several items have been voiced by the drivers that may affect their stress levels and views of the system, including:
    - Waiting for a second vehicle to complete their drop offs and pickups
    - Traveling at a slower pace (95km in the lead vehicle) to ensure the following vehicle can catch up after cut ins
    - Vehicle maintenance issues affecting both drivers' timelines
    - Delays for one driver affecting both drivers' schedules including time on the road and time away from home and/or family obligations.

Appendix B: On-Road Trials Trip Configurations

Trip Configuration based on Duration of the Copilot and Platooning.

- Trip Code is coded as: TripNo:(T02-T21)-Truck (AB1, AB2)-DriverID (Dr01-Dr09)
- Duration is given in h:mm format.
- Trip Configuration is coded as: Hwy2 (Highway 2)-Trip-mode: (short: platooning trip with less than 30 minutes platooning engagement -CTPS: platooning trip, baseline: single truck, Copilot not engaged, copilot: single truck, copilot engaged), the number after short and CTPS trip shows platooning distance.

	Trip Code				Duration (h:mm)			Trip Configuration	
Date	AB1	AB2	Time	Trip	Copilot AB1	Copilot AB2	Platooning	AB1	AB2
Jan 13	T02-AB1-Dr01	T02-AB2-Dr02	Day	1	0:37	0:37	0:20	Hwy2-short- 5s	Hwy2-short- 5s
Jan 13	T02-AB1-Dr01	T02-AB2-Dr02	Day	2	1:50	1:50	1:06	Hwy2-CTPS-5s	Hwy2-CTPS-5s
Jan 13	T03-AB1-Dr05	T03-AB2-Dr09	Night	1	0:00	0:00	0:00	Hwy2-baseline	Hwy2-baseline
Jan 13	T03-AB1-Dr05	T03-AB2-Dr09	Night	2	0:00	0:00	0:00	Hwy2-baseline	Hwy2-baseline
Jan 14	T04-AB1-Dr01	T04-AB2-Dr02	Day	1	2:12	2:12	1:24	Hwy2-CTPS-5s	Hwy2-CTPS-5s
Jan 14	T04-AB1-Dr01	T04-AB2-Dr02	Day	2	1:52	1:52	1:11	Hwy2-CTPS-3s	Hwy2-CTPS-3s
Jan 14	T05-AB1-Dr09	T05-AB2-Dr05	Night	1	0:00	0:00	0:00	Hwy2-baseline	Hwy2-baseline
Jan 14	T05-AB1-Dr09	T05-AB2-Dr05	Night	2	0:00	0:00	0:00	Hwy2-baseline	Hwy2-baseline
Jan 15	T06-AB1-Dr02	T06-AB2-Dr03	Day	1	2:11	2:11	1:56	Hwy2-CTPS-3s	Hwy2-CTPS-3s
Jan 15	T06-AB1-Dr02	T06-AB2-Dr03	Day	2	2:07	2:07	1:40	Hwy2-CTPS-3s	Hwy2-CTPS-3s
Jan 16	T07-AB1-Dr03	T07-AB2-Dr04	Day	1	2:03	2:03	1:20	Hwy2-CTPS-5s	Hwy2-CTPS-5s
Jan 16	T07-AB1-Dr03	T07-AB2-Dr04	Day	2	1:57	1:57	0:53	Hwy2-CTPS-3s	Hwy2-CTPS-3s
Jan 17	T08-AB1-Dr04	T08-AB2-Dr06	Day	1	0:00	0:00	0:00	Hwy2-baseline	Hwy2-baseline
Jan 17	T08-AB1-Dr04	T08-AB2-Dr06	Day	2	0:40	0:40	0:00	Hwy2-baseline	Hwy2-baseline
Jan 18	T09-AB1-Dr05	T09-AB2-Dr06	Day	1	0:08	0:08	0:03	Hwy2-short-5s	Hwy2-short-5s
Jan 18	T09-AB1-Dr05	T09-AB2-Dr06	Day	2	0:42	0:42	0:33	Hwy2-CTPS-3s	Hwy2-CTPS-3s
Jan 19	T10-AB1-Dr05	T10-AB2-Dr06	Day	1	NA	NA	NA	Hwy2-Copilot	Cancelled
Jan 19	T10-AB1-Dr05	T10-AB2-Dr06	Day	2	NA	NA	NA	Hwy2-Copilot	Cancelled
Jan 20	T11-AB1-Dr04	T11-AB2-Dr05	Day	1	1:22	1:22	0:58	Hwy2-CTPS-3s	Hwy2-CTPS-3s
Jan 20	T11-AB1-Dr04	T11-AB2-Dr05	Day	2	1:31	1:31	0:50	Hwy2-CTPS-3s	Hwy2-CTPS-3s
Jan 21	T12-AB1-Dr09	T12-AB2-Dr01	Day	1	NA	NA	NA	Cancelled	Hwy2-Copilot
Jan 21	T12-AB1-Dr09	T12-AB2-Dr01	Day	2	NA	NA	NA	Cancelled	Hwy2-Copilot
Jan 22	T13-AB1-Dr01	T13-AB2-Dr06	Day	1	1:35	1:35	1:13	Hwy2-CTPS-3s	Hwy2-CTPS-3s
Jan 22	T13-AB1-Dr01	T13-AB2-Dr06	Day	2	2:13	2:13	1:40	Hwy2-CTPS-4s	Hwy2-CTPS-4s
Jan 23	T14-AB1-Dr06	T14-AB2-Dr07	Day	1	2:14	2:14	1:45	Hwy2-CTPS-3s	Hwy2-CTPS-3s
Jan 23	T14-AB1-Dr06	T14-AB2-Dr07	Day	2	2:22	2:22	0:55	Hwy2-CTPS-5s	Hwy2-CTPS-5s
Jan 24	T15-AB1-Dr06	T15-AB2-Dr08	Day	1	0:18	0:18	0:03	Hwy2-short-5s	Hwy2-short-5s
Jan 24	T15-AB1-Dr06	T15-AB2-Dr08	Day	2	0:21	0:21	0:02	Hwy2-short-3s	Hwy2-short-3s
Jan 25	T16-AB1-Dr08	T16-AB2-Dr06	NA	NA	NA	NA	NA	Hwy2-CTPS-5s	Hwy2-CTPS-5s
Jan 25	T16-AB1-Dr08	T16-AB2-Dr06	NA	NA	NA	NA	NA	Hwy2-CTPS-3s	Hwy2-CTPS-3s

Jan 26	T17-AB1-Dr09	T17-AB2-Dr08	Day	1	1:43	1:43	1:25	Hwy2-CTPS-3s	Hwy2-CTPS-3s
Jan 26	T17-AB1-Dr09	T17-AB2-Dr08	Day	2	0:37	0:37	0:30	Hwy2-CTPS-4s	Hwy2-CTPS-4s
Jan 27	T18-AB1-Dr08	T18-AB2-Dr09	Day	1	1:55	1:55	1:31	Hwy2-CTPS-3s	Hwy2-CTPS-3s
Jan 27	T18-AB1-Dr08	T18-AB2-Dr09	Day	2	1:29	1:29	1:15	Hwy2-CTPS-3s	Hwy2-CTPS-3s
Jan 28	T19-AB1-Dr09	T19-AB2-Dr08	Day	1	2:23	2:23	2:06	Hwy2-CTPS-3s	Hwy2-CTPS-3s
Jan 28	T19-AB1-Dr09	T19-AB2-Dr08	Day	2	NA	NA	NA	Hwy2-CTPS-5s	Hwy2-CTPS-5s
Jan 29	T20-AB1-Dr07	T20-AB2-Dr09	Day	1	0:00	0:00	0:00	Hwy2-baseline	Hwy2-baseline
Jan 29	T20-AB1-Dr07	T20-AB2-Dr09	Day	2	NA	NA	NA	Cancelled	Cancelled
Jan 30	T21-AB1-Dr04	T21-AB2-Dr07	Day	1	1:53	1:53	1:38	Hwy2-CTPS-5s	Hwy2-CTPS-5s
Jan 30	T21-AB1-Dr04	T21-AB2-Dr07	Day	2	1:26	1:26	0:53	Hwy2-CTPS-4s	Hwy2-CTPS-4s

Appendix C: Data Collection Protocol

COVID Protocol

1. Wear mask
2. Have sanitizer available for yourself and drivers
3. Make sure to sanitize all equipment that are shared with drivers (iPads and Tablets) are sanitized after being used
4. Make sure to use drivers' specific Muse headband

Steps Before Drivers Arrive

Tablet

1. If it is Pre Trip session, move the tablet to the office to do Researchers Questionnaire (http://tiny.cc/R_Q) in the office
2. Turn on the tablets
3. Verify they are fully charged
4. Verify the connection to the Wi-Fi is stablished
5. Verify the bookmark to the Power up (http://tiny.cc/powerup) and power down (http://tiny.cc/powerdown) checklist are ready
6. Verify the bookmark to the Researchers Questionnaire (http://tiny.cc/R_Q) is ready

iPads

1. Turn on the iPads
2. Verify they are fully charged
3. Verify the connection to the Wi-Fi is stablished
4. Verify the bookmarks to the SurveyMonkey questionnaires are ready
    a. http://tiny.cc/PreQ for Pre Trip session in Calgary,
    b. http://tiny.cc/PostQ1 for their stop at AMTA in Edmonton
    c. http://tiny.cc/PostQ2 for their final stop at Bison Transport, Calgary

Muse Pods

1. Verify Muse pods are fully charged (all 3 lights should turn on when you turn on the pods)

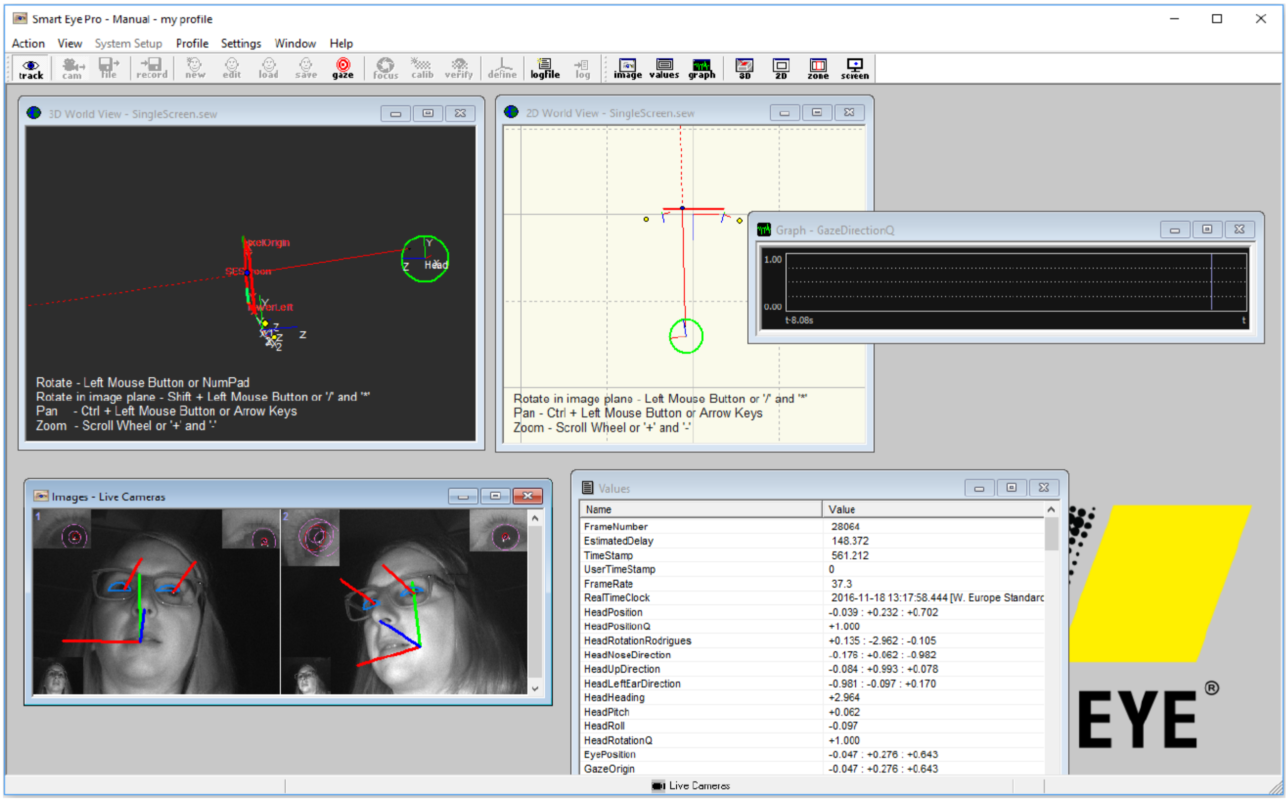
Setup Smart Eye

1. Press the orange button on the industrial PC and it will turn from orange to blue (which means that it is on).
2. The red LED on the Smart Eye exponator (the small box next to the industrial PC) should be turned on, and the yellow LED should be blinking.
3. The laptop is placed on top of the cabinet behind the driver's seat for both trucks. No password is needed. Turn it on.
4. On the task bar, there is a shortcut to Remote Desktop Connection. Open it (you may need to wait a minute or two for the industrial PC to boot).
5. Press Connect to get connected to the industrial PC:

6. After connecting to industrial PC, Open the Smart Eye program. The shortcut is on the desktop.

a. You should be able to see a similar view as that shown below.
b. If the zoom setting is not good enough, right click, and choose the zoom to Fit-to-window
c. By activating/deactivating track mode (track) on the upper left, the output will be toggled between the following two cases:

   i. Track is not activated



Rotate - Left Mouse Button or NumPad
Rotate in image plane - Shift + Left Mouse Button or '/' and '*'
Pan - Ctrl + Left Mouse Button or Arrow Keys
Zoom - Scroll Wheel or '+' and '-'

Rotate in image plane - Left Mouse Button or '/' and '*'
Pan - Ctrl + Left Mouse Button or Arrow Keys
Zoom - Scroll Wheel or '+' and '-'

Name	Value
FrameNumber	28064
EstimatedDelay	148.372
TimeStamp	561.212
UserTimeStamp	0
FrameRate	37.3
RealTimeClock	2016-11-18 13:17:58.444 [W. Europe Standarc
HeadPosition	-0.039 : +0.232 : +0.702
HeadPositionQ	+1.000
HeadRotationRodrigues	+0.135 : -2.962 : -0.105
HeadNoseDirection	-0.176 : +0.062 : -0.982
HeadUpDirection	-0.084 : +0.993 : +0.078
HeadLeftEarDirection	-0.981 : -0.097 : +0.170
HeadHeading	+2.964
HeadPitch	+0.062
HeadRoll	-0.097
HeadRotationQ	+1.000
EyePosition	-0.047 : +0.276 : +0.643
GazeOrigin	-0.047 : +0.276 : +0.643

   i. Track is activated

7. Grab the chessboard from the cabinet above the fridge (across from the laptop)

Muse Headband

1. Verify the Muse pod is fully charged and bring it to the truck.
2. Find drivers' specific muse headband (label them at the end of each trip and store in boxes or zip-loc bags)
3. Verify the iPhone is fully charged and bring it to the truck.
4. Connect iPhones to car USB charger

Steps upon Drivers Arrival

Welcome driver

1. Introduce yourself
2. Ask their names

3. If it is Pre Trip session explain their roles

Tablet tasks

1. Fill out the Researchers Questionnaire (http://tiny.cc/R_Q) on the tablet. These tasks will be done in office for Pre Trip session, and in-cab for Post Trip session 1 and 2.
2. (Pre Trip for their initial arrival at Bison Transport, Post Trip 1 for their arrival at AMTA in Edmonton, and Post Trip 2 for their arrival at Bison Transport, Calgary).
3. If it is Pre Trip session, then move back the tablet to the trucks to use them for power up/down procedure.

iPads Tasks

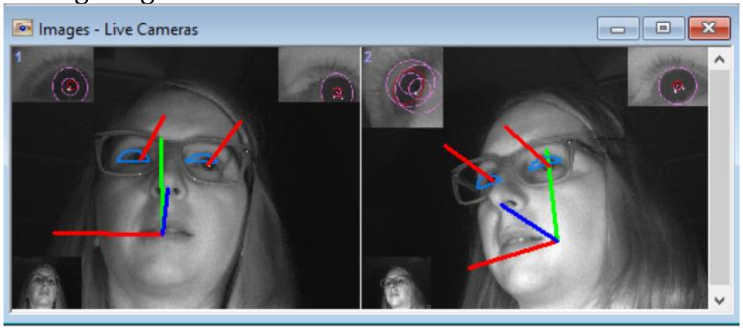
1. Perform the Impircia Vitals Mobile on the iPads
2. Open the SurveyMonkey link in Incontigo mode on the iPads and ask drivers to fill out the forms:
    a. http://tiny.cc/PreQ for Pre Trip session in Calgary,
    b. http://tiny.cc/PostQ1 for their stop at AMTA in Edmonton
    c. http://tiny.cc/PostQ2 for their final stop at Bison Transport, Calgary

Accompany Drivers to the trucks.

When Drivers Enter the Truck

Smart Eye

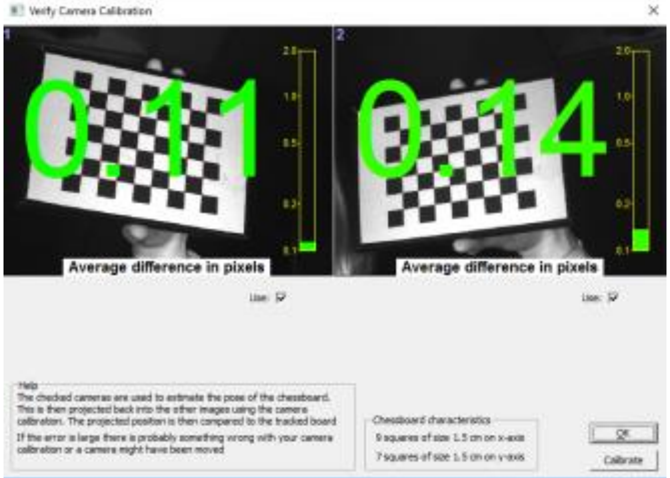
1. **Initial Setting**
    a. Ask driver to adjust their seat.
    b. In case the steering wheel is blocking their face (you can check it in Smart Eye), ask them to **momentarily** move the steering wheel down.
    c. Activate track mode (track) each time a participant comes into the truck
    d. Ask drivers to wear their sunglasses (if they use it during driving) to make sure their sunglasses are compatible with the system. The gaze direction lines should not change with wearing sunglasses.



If there was any reduction of difficulty for system to find their eyes or gaze, ask them to use our compatible sunglasses (found above the fridge).

2. **Eye Calibration**
    a. Align the middle of the chessboard with the driver's nose and move it around a little bit in a clockwise or counter-clockwise motion (this is to adjust the level of brightness)
    b. Deactivate the track mode
    c. From the menu select□ verify (verify)

- d. Place the chessboard in front of the driver's nose
- e. You will see numbers show up on the screen
- f. If the numbers are not good (meaning for small chess board, any numbers larger than 0.15 and for large chessboard, any numbers larger than 0.3), we need to do the calibration:
    - i. Press Calibrate
    - ii. Starting from drivers' nose, do the dolphin maneuver (moving forward tilt down and moving backward tilt up) **towards each camera** for a length of 15-20cm).
    - iii. After finishing the calibration, check the numbers while the chessboard is placed (stationary) in front of the driver's nose.
    - iv. If the numbers are still not good, then calibrate again.
    - v. Repeat steps i-iv two or three times or until you get acceptable calibration level



Help
The checked cameras are used to estimate the pose of the chessboard. This is then projected back into the other images using the camera calibration. The projected position is then compared to the tracked board
If the error is large there is probably something wrong with your camera calibration or a camera might have been moved

Chessboard characteristics
9 squares of size 1.5 cm on x-axis
7 squares of size 1.5 cm on y-axis

The Verify Camera Calibration Dialog. The error vectors are small, and the accuracy in the second camera is at 0.14 pixel

3. Once you have calibrated, load driver's profile:
    - a. We found each driver's specific eye model previously, and we just need to load that file for each driver.
    - b. Make sure track mode is not activated.
    - c. Open driver profile (load)
    - d. Ask driver to tell you their Bison code
    - e. Find and open it
4. Ready to record!
    - a. Activate track mode
    - b. Click (logfile)
    - c. This will open a window to define the name of the output file name. The default name is "date-time". Add '-Drivers' code' after this default name. Then press ok.
    - d. Press log (log) to start recording
5. Before you leave the trucks, ensure that both drivers know that if they are to turn the truck/inverter off for ANY reason (such as to get gas), that they need to click on the "log" button on the Smart Eye system before doing so in order to preserve our collected data.

Muse Headband

1. Have water and a napkin ready. If there is bad connection to the electrodes, use a damp cloth to gently wipe the driver's forehead area and the headband.
2. Watch the following video before the session:
3. https://www.youtube.com/watch?v=61nJZxXwbQc
4. Make sure iphone can connect to Wi-Fi
    a. You will need to be connected to Wi-Fi for both of the following devices. If you are not able to connect to the Bison yard (in Calgary) or the AMTA (in Edmonton) Wi-Fi, then you may be able to use the Wi-Fi inside the trucks by turning on the device on the left of the laptop in AB2.
5. Ask drivers to put their headband on before they put their mask on.
6. Attach the muse pod to the headband (can be done either before or after they put their headband on)
    a. The power button should always be on top.
    b. The muse logo on the headband should always be right way up, such that the brown part of the headband (ear sensors) should be resting on the driver's ears.
        i. You can ask the driver to move any hair that may be in the way.
7. Adjust the length of the strap so that it's snug but comfortable for the driver.
8. Press the power button on top of the pod (you should see orange lights light up on the bottom of the pod).
9. Open the Mind Monitor app on the iPhone
10. The ovals on the bottom left side of the screen shows connection strength of electrodes to skin. Having all the ovals be filled with solid colours (bottom left) means the electrodes are properly connected to the skin and are gathering data.
11. Once you get good signals, click the red "record" button, on the bottom of the screen.

Tablet

1. Do the power up procedure with drivers (http://tiny.cc/powerup)

End of trip

The end of trip is considered during the arrival to AMTA Edmonton or Bison Transport Calgary.

Smart Eye

1. Finish data collection session
    a. When drivers get back, press log (log) again to stop recording data
2. Back up data:
    a. Connect the external hard drive to USB port of the industrial PC
    b. Find the collected data in the following folder:
    c. Documents/Smart Eye/Smart Eye Pro 9.2 /Log Files
    d. You can sort files based on date to easily find the recently collected data
    e. Copy the files to the external hard drive:
    f. D:\Start_here_mac.app\CTPS\Smart Eye\Truck1 or Truck2
3. At the end of the day, to turn off the equipment:
    a. From Smart eye, press shut down from start menu. Shut down the industrial PC.

Muse Headband

1. Make sure you're connected to Wi-Fi
    a. You will need to be connected to Wi-Fi for both of the following devices. If you are not able to connect to the Bison yard (in Calgary) or the AMTA (in Edmonton) Wi-Fi, then you may be able to use the Wi-Fi inside the trucks by turning on the device on the left of the laptop in AB2.
2. click the white "stop" button (square),
3. Wait for the data to upload to the cloud.

Tablet

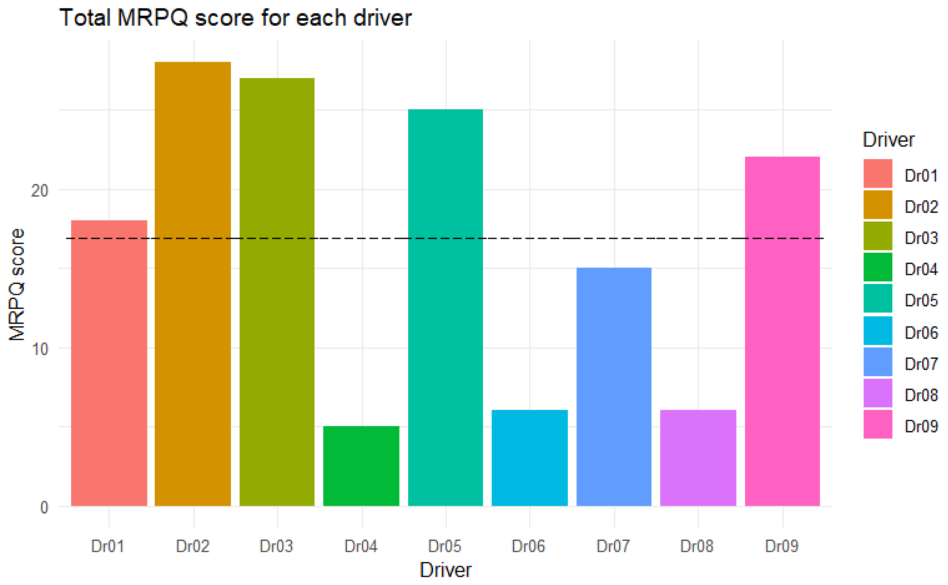
1. Do the power down procedure with drivers (http://tiny.cc/powerdown)

Charge devices

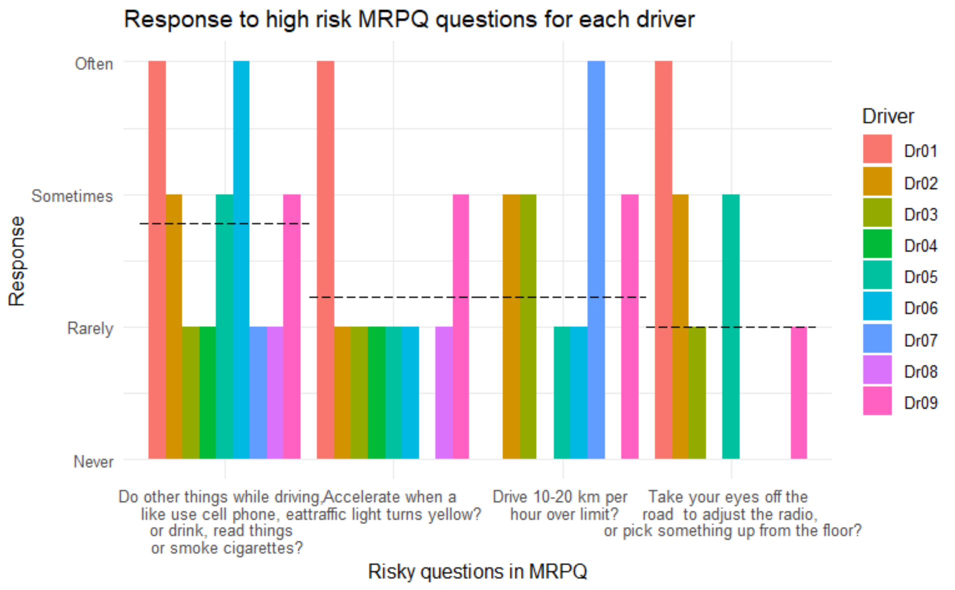
1. Bring Muse pod to the office to be charged
2. Bring Tablet to the office to be charged
3. If the iPhone is not charged, bring it to the office to be charged

Appendix D: Questionnaire Responses

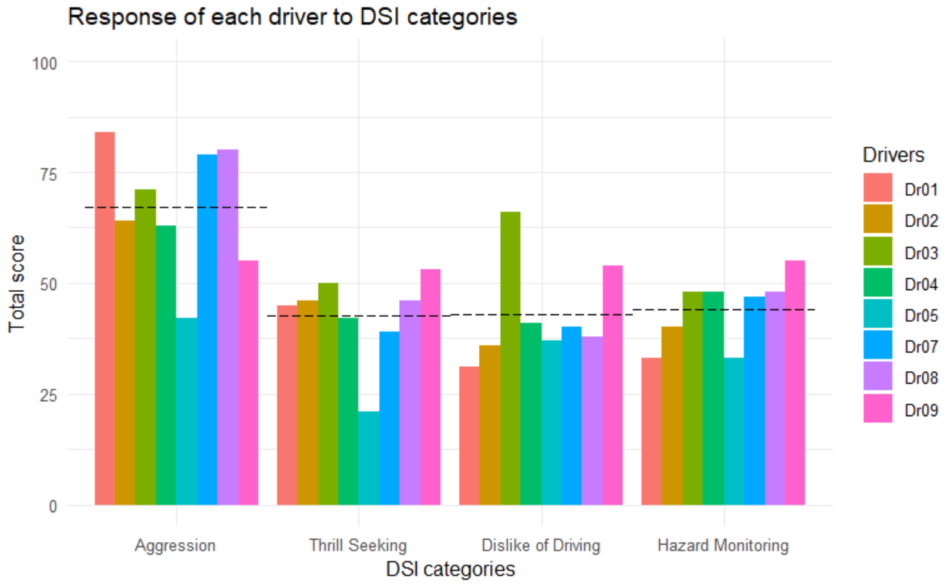
In this section drivers' individual responses to MRPQ, DSI, TiA, and AATT questionnaires are presented.



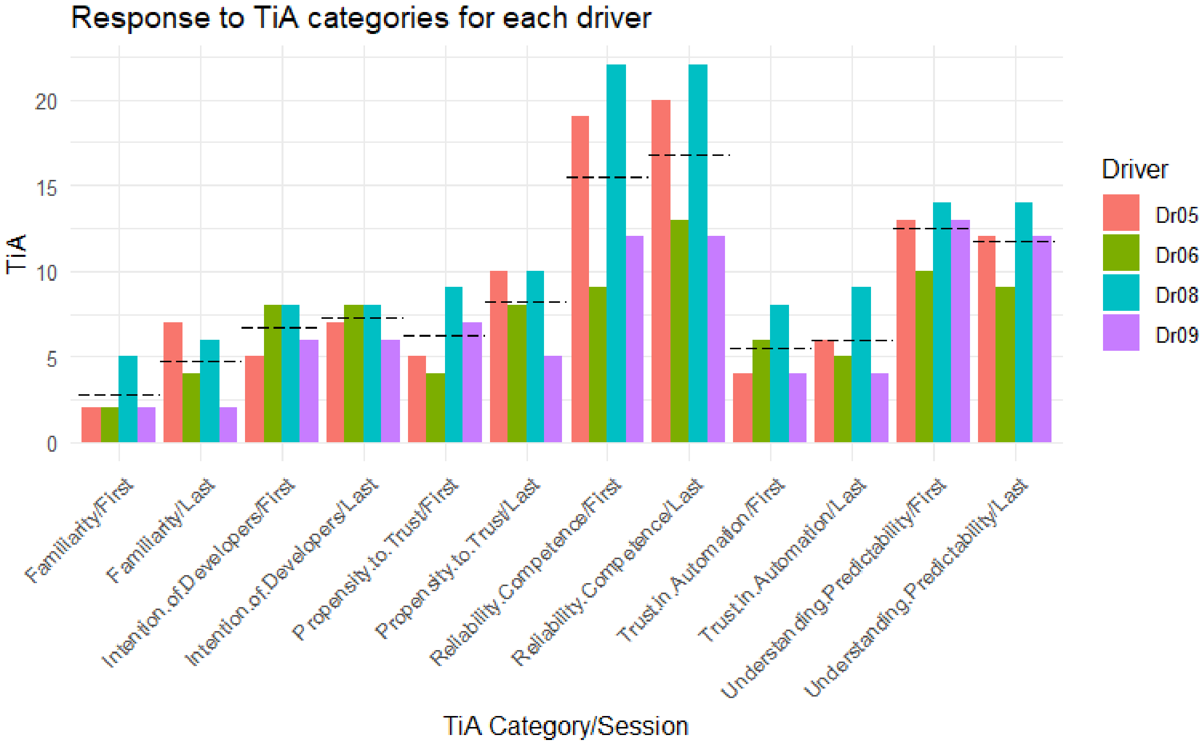
Total score of each drivers' response to MRPQ. The group average is shown in dashed line.



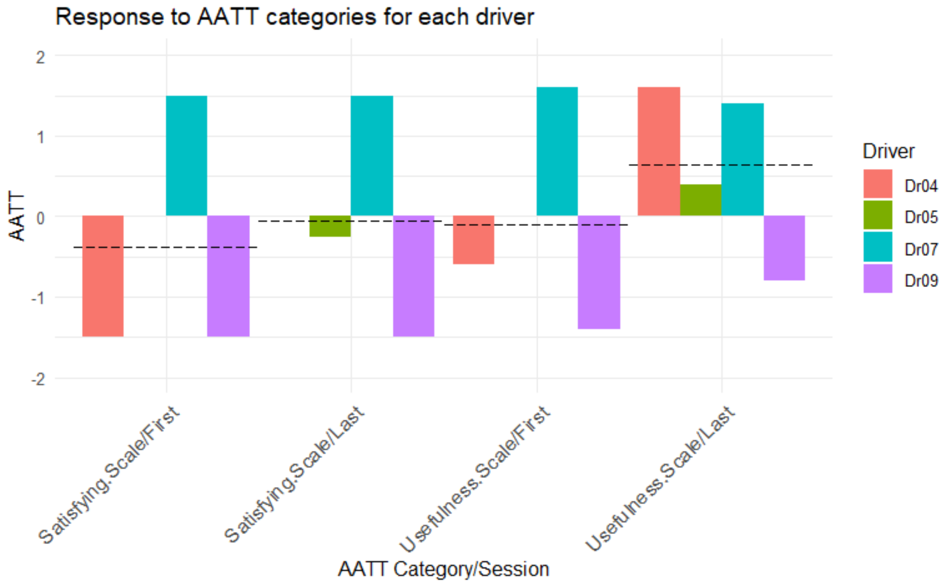
Response of each driver to the top four identified risk behaviors in MRPQ. The group average to each question is shown in dashed line



Total score of each driver's response to each category in DSI. The group average to each question is shown in dashed line



Total score of each driver's response to each category in TiA in the first vs. Last platooning session. The group average to each question is shown in dashed line



Total score of each driver's response to each category in AATT in the first vs. Last platooning session. The group average to each question is shown in dashed line

Appendix E: Tablet-Based Cognitive Tasks (TBCT)

In this section, extracted measures from each of the tasks in TBCT is introduced.

Reaction time task

Measure	Description
% Correct	Percentage of times when the object was correctly stopped by the correct stop button
% Incorrect	Percentage of times when the object was incorrectly stopped using the incorrect stop button
% Premature	Percentage of times when the participant had lifted their finger from the Start button before any visual cue was given
Reaction-Time (Correct)	Average reaction-time taken to correctly stop the moving object using the correct stop button
Reaction-Time (Incorrect)	Average reaction-time taken to stop the moving object using the incorrect stop button
% Correct (Congruent)	Percentage of times when the object was correctly stopped by the correct stop button for congruent trials
% Incorrect (Congruent)	Percentage of times when the object was stopped by the incorrect stop button for congruent trials
% Correct (Incongruent)	Percentage of times when the object was correctly stopped by the correct stop button for incongruent trials
% Incorrect (Incongruent)	Percentage of times when the object was stopped by the incorrect stop button for incongruent trials
Reaction-Time (Correct, Congruent)	Average reaction-time taken to correctly stop the moving object using the correct stop button for congruent trials
Reaction-Time (Incorrect, Congruent)	Average reaction-time taken to stop the moving object using the incorrect stop button for congruent trials
Reaction-Time (Correct, Incongruent)	Average reaction-time taken to correctly stop the moving object using the correct stop button for incongruent trials
Reaction-Time (Incorrect, Incongruent)	Average reaction-time taken to stop the moving object using the incorrect stop button for incongruent trials
% Timeout	Percentage of times the participant had timed out for the trial

Judgment task

Measure	Description
Reaction-Time (Correct)	Average reaction-time after the visual cue to correctly pass the lines
Reaction-Time (Incorrect)	Average reaction-time after the visual cue but failed to pass the first set of lines
Go Count	Average amount of times the ‘Go’ button was pressed after the visual cue in order to correctly pass the lines
% Premature go	Percentage of times the participant had released the Start button before visual cue was given
% Correct	Percentage the participant correctly passed all set of lines

% Early	Percentage the participant had collided with the lines from the frontside
% Late	Percentage the participant had collided with the lines other than from the frontside
% Timeout	Percentage of times the participant had timed out

Note. Each stage was measured with all listed measurements for this task

Memory task

Measure	Description
Overall Draw Time (Correct)	Average time it took to draw the shape correctly overall
Overall Draw Time (Incorrect)	Average time it took to draw the shape incorrectly overall
Draw Time (Correct, Stage 1 & 2)	Average time it took to draw one shape correctly for Stage 1 and 2
Draw Time (Correct, Stage 3 & 4)	Average time it took to draw two shapes correctly for Stage 3 and 4
Draw Time (Incorrect, Stage 1 & 2)	Average time it took to draw one shape incorrectly for Stage 1 and 2
Draw Time (Incorrect, Stage 3 & 4)	Average time it took to draw two shapes incorrectly for Stage 3 and 4
Line Efficiency Overall (Correct)	Average line efficiency when drawing the shape correctly
Line Efficiency Overall (Incorrect)	Average line efficiency when drawing the shape incorrectly
Line Efficiency (Correct, Stage 1 & 2)	Average line efficiency when drawing the shape correctly in Stage 1 and 2
Line Efficiency (Incorrect, Stage 1 & 2)	Average line efficiency when drawing the shape incorrectly in Stage 1 and 2
Line Efficiency (Correct, Stage 3 & 4)	Average line efficiency when drawing the two shapes correctly in Stage 3 and 4
Line Efficiency (Incorrect, Stage 3 & 4)	Average line efficiency when drawing the two shapes incorrectly in Stage 3 and 4
% Shapes Replicated	Percentage of shapes that were replicated overall
% Shapes Replicated (Stage 1 & 2)	Percentage of shapes that were replicated in Stage 1 and 2
% Shapes Replicated (Stage 3 & 4)	Percentage of shapes that were replicated in Stage 3 and 4
Total Touches Overall (Correct)	Count of how many touches the participant made for correct shapes overall
Total Touches (Correct, Stage 1 & 2)	Count of how many touches the participant made for correct shapes in Stage 1 and 2
Total Touches (Correct, Stage 3 & 4)	Count of how many touches the participant made for correct shapes in Stage 3 and 4
Total Touches (Only one shape correct, Stage 3 & 4)	Count of how many touches the participant made for only one of the shapes out of the two correctly drawn in Stage 3 and 4
Total Touches Overall (Incorrect)	Count of how many touches the participant made for incorrect shapes overall
Total Touches (Incorrect, Stage 1 & 2)	Count of how many touches the participant made for incorrect shapes in Stage 1 and 2
Total Touches (Incorrect, Stage 3 & 4)	Count of how many touches the participant made for incorrect shapes in Stage 3 and 4
Long Touches Overall (Correct)	Count of how many long touches the participant made for correct shapes overall

Long Touches (Correct, Stage 1 & 2)	Count of how many long touches the participant made for correct shapes in Stage 1 and 2
Long Touches (Correct, Stage 3 & 4)	Count of how many long touches the participant made for correct shapes in Stage 3 and 4
Long Touches (1 out 2 shapes correctly, Stage 3 & 4)	Count of how many long touches the participant made for only one of the shapes out of the two correctly drawn in Stage 3 and 4
Long Touches Overall (Incorrect)	Count of how many long touches the participant made for incorrect shapes overall
Long Touches (Incorrect, Stage 1 & 2)	Count of how many long touches the participant made for incorrect shapes in Stage 1 and 2
Long Touches (Incorrect, Stage 3 & 4)	Count of how many long touches the participant made for incorrect shapes in Stage 3 and 4

Control task

Measure	**Description**
% Inside Target	Percentage inside the target out of the duration of the trial
% Fixed Objects Avoided	Percentage of fixed objects avoided out of fixed objects presented
% Surprise Objects Avoided	Percentage of surprised objects avoided out of surprise objects presented
% Right Edge	Percentage on the right edge of the target out of the duration of the trial
% Left Edge	Percentage on the left edge of the target out of the duration of the trial
Duration	Duration of the trial

Appendix F: Eye Tracking System (Smart Eye)

This section presents the specification of Smart Eye Pro system and cameras.

General Specifications and features for Smart Eye Pro Cameras

- Free placements of cameras: Smart Eye Pro is a flexible system where cameras are located based on customer setup
- Optimal Camera – Eye Distance: 30 – 300 cm, adjustable with lenses and positioning of cameras.
- Field of view: 90° - 360° (depending on number of cameras)
- Tracking Accuracy:
  - Head: Rotation 0.5 degrees (under ideal conditions)
  - Gaze: 0.5 degrees (under ideal conditions)
- Output: TCP / UDP / CAN (optional) / text file
- Output Data (true 3D values): Head orientation (6DOF), Eye position, Eye gaze, Pupil diameter, Saccades, Fixations, Blinks, Eyelid opening etc. Quality values Data for left and right eye
- Recovery Time (Blink/Tracking Lost): Typ. 1 frame (16/8ms)
- Eyewear Compatibility: Glasses, contact lenses and non-IR-blocking sunglasses
- Calibration Mode: Flexible number of calibration points
- Light independent: Smart Eye Pro uses active IR illumination @ 850nm
- Possible to synch with other systems: It is possible to sync or de-synch Smart Eye Pro compared to other systems, for example if you want to use more than one system in the same room
- Integration support Remote control over Ethernet and programming guide to support integrations for example analysis tools, simulators, etc.

	SMART EYE PRO 1.3MP	SMART EYE PRO Dx
Sampling rate	60 and 120Hz	60Hz
Number of cameras	Up to 6 cam Both 60 and 120Hz	Up to 6 cam (Desktop) Up to 4 cam (Laptop)
Head box (Freedom of head movement)	Up to 90x75cm	Up to 110x85cm
Camera interface	Ethernet (60Hz) USB3 (120Hz	USB3
Camera size	109x28x28mm	17x31x31mm
Camera weight	80g	20g
Resolution	1.3MP	2MP
Mounting kit (parts)	Triangle foot	Small form factor 3 DOF
Size of applicable flashes	30x30x30mm	19x14x28mm
Lenses (focal length)	4.5, 6, 8, 12, 16	8, 12, 16
Maximum available cable length*	10/8m	8m

Comparison between different cameras used in each truck

Extracted Measures

In this section, a list of measures that were extracted from Smart Eye output and its relevance to fatigue is given.

Pupil Diameter

Changes in pupil diameter are believed to be indicative of the presence of neuromodulators such as noradrenaline (Larsen & Waters, 2018). These changes in pupil diameter are often accompanied by behavioral and brain changes (Larsen & Waters, 2018). The Iris sphincter is the main muscle that controls the diameter of the pupil and is controlled by motor neurons in the Edinger-Westphal nucleus (Larsen & Waters, 2018). The diameter of the pupil changes due to both changing light conditions (pupillary light response) and arousal levels due to the activation of the sympathetic and parasympathetic nervous system (Larsen & Waters, 2018). The pupillary light response refers to the dilation of the pupil in low light conditions and constriction of the pupil in high light conditions in order to optimize retinal illumination and visual perception (Larsen & Waters, 2018).

Furthermore, changes in pupil diameter are also correlated with changes in neuromodulatory signaling in the locus coeruleus (LC) which are activated in response to sleep-wake transitions, noxious stimuli, and new sensory experiences (spiking of LC tends to precede increases in pupil diameter by 335ms) (Larsen & Waters, 2018).
During each segment, the average of pupil diameter is calculated and compared across conditions.

Eyelid Opening

Studies have shown that drowsiness can be detected based on the movement of the eyelids (Liu et al., 2010). A study conducted by Han et al. (n.d) suggested that eyelid movement is one of the most reliable methods for detecting drowsiness in drivers. They explain that eyelid movement can be calculated using the percentage of eye closure (PERCLOS), as well as the closing/opening duration of the eyelids, the average closing and opening speed of the eyelids, and the blink rate.
During each segment, the average of eyelid opening is calculated and compared across conditions.

Blink

The Smart Eye system detects blinks by evaluating the eyelid opening samples over a period of 700 ms. Blinking consists of the rapid closing of the eyelid, with the average blink length 250 milliseconds (Noguchi et al., 2007; Soukupová, 2016). Blank waveform patterns (from electro-oculogram data) have been used previously to assess the arousal level of drivers (Noguchi et al., 2007).
During each segment, the average blink rate (the number of blinks in one minute) is calculated and compared across conditions.

Fixation

Fixations are defined as pauses that occur over informative regions of interest, however it is less clear where fixations start and where they end (Salvucci & Goldberg, 2000). Smaller eye movements that occur during fixations, such as tremors, drifts or flicks, often mean little in higher-level analysis (Salvucci & Goldberg, 2000). There are a variety of different fixation identification algorithms that can be used to analyze fixations (Salvucci & Goldberg, 2000).
During each segment, the average fixation rate (the number of blinks in one minute) is calculated and compared across conditions.

Saccade

A saccade is a rapid, ballistic movement of the eye that occurs between fixations (Salvucci & Goldberg, 2000; Spivey et al., 2004). They typically occur 3-4 times every second (Spivey et al., 2004). Common metrics used to measure saccades include: saccadic velocity, or saccadic amplitude (Salvucci & Goldberg, 2000). However, there is very little visual processing that occurs during a saccade so this measure is often excluded in many research applications (Salvucci & Goldberg, 2000). The analysis of fixations and saccades requires fixation identification wherein one tracks normal eye-movements to fixated locations and views the saccades in between (Salvucci & Goldberg, 2000).

During each segment, the average saccade rate (the number of blinks in one minute) in one is calculated and compared across conditions.

Gaze location:

The following objects were identified in the truck before the on-road trials, so that we could easily track the time that drivers look at either of these objects.

The wind screen is divided into two sections, left (WindScreenLH) and right (WindScreenRH), left mirror (LMirror), right mirror (RMirror), Cluster, instrument panel(Meters), and control buttons for platooning system (Platooning)

During each segment, the percentage of the time that the driver spent looking at either of these objects is calculated and compared across conditions.

References

Andrusaitis, S. F., Oliveira, R. P., & Barros Filho, T. E. P. (2006). Study of the prevalence and risk factors for low back pain in truck drivers in the state of São Paulo, Brazil. *Clinics*, *61*(6), 503–510. https://doi.org/10.1590/S1807-59322006000600003

Antin, J., Lee, S., Hankey, J., & Dingus, T. (2011). *Design of the In-Vehicle Driving Behavior and Crash Risk Study*. Transportation Research Board. https://doi.org/10.17226/14494

Arefnezhad, S., Hamet, J., Eichberger, A., Frühwirth, M., Ischebeck, A., Koglbauer, I. V., … Yousefi, A. (2022). Driver drowsiness estimation using EEG signals with a dynamical encoder–decoder modeling framework. *Scientific Reports 2022 12:1*, *12*(1), 1–18. https://doi.org/10.1038/s41598-022-05810-x

Ayoub, J., Du, N., Yang, X. J., & Zhou, F. (2021). Predicting Driver Takeover Time in Conditionally Automated Driving. *IEEE Transactions on Intelligent Transportation Systems*, *XX*. https://doi.org/10.48550/arxiv.2107.09545

Bagherzadeh, Y., Baldauf, D., Pantazis, D., & Desimone, R. (2020). Alpha Synchrony and the Neurofeedback Control of Spatial Attention. *Neuron*, *105*(3), 577-587.e5. https://doi.org/10.1016/J.NEURON.2019.11.001/ATTACHMENT/09206838-29C3-41DC-B377-19A206042713/MMC2.MP4

Bakhtiari, R., Tomczak, M. V, Langor, S., Scanlon, J. E. M., Granley, A., & Singhal, A. (2020). Application of tablet-based cognitive tasks to predict unsafe drivers in older adults. *Transportation Research Interdisciplinary Perspectives*, *4*, 100105. https://doi.org/10.1016/j.trip.2020.100105

Brell, T., Philipsen, R., & Ziefle, M. (2019). sCARy! Risk Perceptions in Autonomous Driving: The Influence of Experience on Perceived Benefits and Barriers. *Risk Analysis*, *39*(2), 342–357. https://doi.org/10.1111/risa.13190

Buld, S., Tietze, H., & Krüger, H.-P. (2005). Auswirkungen von Teilautomation auf das Fahren. *Fahrerassistenzsysteme Mit Maschineller Wahrnehmung*, 161–187. https://doi.org/10.1007/3-540-27137-6_8

Catherine, G., Bruno, B., Christophe, J., Marion, B., & Alexandra, F. (2012). Cognitive load measurement while driving. In : Human Factors: a view from an integrative perspective - Archive ouverte HAL. *Human Factors and Ergonomics Society*, 67–80. Retrieved from https://hal.archives-ouvertes.fr/hal-01027475

Chan, M., Nyazika, S., & Singhal, A. (2016). Effects of a front-seat passenger on driver attention: An electrophysiological approach. *Transportation Research Part F: Traffic Psychology and Behaviour*, *43*, 67–79. https://doi.org/10.1016/J.TRF.2016.09.016

Chang, T. H., & Chen, Y. R. (2014). Driver fatigue surveillance via eye detection. *2014 17th IEEE International Conference on Intelligent Transportation Systems, ITSC 2014*, 366–371. https://doi.org/10.1109/ITSC.2014.6957718

Choi, S. Y., Yoo, D. H., & Lee, J. S. (2015). Usefulness of the driveABLE cognitive assessment in predicting the drivingrisk factor of stroke patients. *Journal of Physical Therapy Science*, *27*(10), 3133. https://doi.org/10.1589/JPTS.27.3133

Craig, A., Tran, Y., Wijesuriya, N., & Nguyen, H. (2012). Regional brain wave activity changes associated with fatigue. *Psychophysiology*, *49*(4), 574–582. https://doi.org/10.1111/J.1469-8986.2011.01329.X

Cruikshank, L. C., Singhal, A., Hueppelsheuser, M., & Caplan, J. B. (2012). Theta oscillations reflect a putative neural mechanism for human sensorimotor integration. *Journal of Neurophysiology*, *107*(1), 65–77. https://doi.org/10.1152/JN.00893.2010/ASSET/IMAGES/LARGE/Z9K0121110990006.JPEG

Cummings, M. L., Mastracchio, C., Thornburg, K., & Mkrtchyan Armen Ashot. (2013). Boredom and

Distraction in Multiple Unmanned Vehicle Supervisory Control. *Interacting with Computers*, *25*(1), 34–47. Retrieved from https://dspace.mit.edu/handle/1721.1/110429

Dingus, T. A., Hankey, J. M., Antin, J. F., Lee, S. E., Eichelberger, L., Stulce, K. E., … Stowe, L. (2015). Naturalistic Driving Study: Technical Coordination and Quality Control. *Naturalistic Driving Study: Technical Coordination and Quality Control*, 370. Retrieved from http://onlinepubs.trb.org/onlinepubs/shrp2/SHRP2_S06Report.pdf

Dissanayake, U. C., Steuber, V., & Amirabdollahian, F. (2022). EEG Spectral Feature Modulations Associated With Fatigue in Robot-Mediated Upper Limb Gross and Fine Motor Interactions. *Frontiers in Neurorobotics*, *15*, 192. https://doi.org/10.3389/FNBOT.2021.788494/BIBTEX

Dobbs, A. R., & Résumé, P. (2013). Accuracy of the DriveABLE cognitive assessment to determine cognitive fitness to drive. *Canadian Family Physician*, *59*(3), e156. Retrieved from /pmc/articles/PMC3596225/

Ellinghaus, D., & Schlag, B. (1984). *Alter und Autofahren: eine zukunftsorientierte Studie über ältere Kraftfahrer*. *Uniroyal-Verkehrsuntersuchung, , Uniroyal-AG* (Vol. 11). IFAPLAN.

Funke, G., Greenlee, E., Carter, M., Dukes, A., Brown, R., & Menke, L. (2016). Which Eye Tracker Is Right for Your Research? Performance Evaluation of Several Cost Variant Eye Trackers. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *60*(1), 1240–1244. https://doi.org/10.1177/1541931213601289

Galéra, C., Orriols, L., M'Bailara, K., Laborey, M., Contrand, B., Ribéreau-Gayon, R., … Lagarde, E. (2012). Mind wandering and driving: responsibility case-control study. *BMJ*, *345*(7888). https://doi.org/10.1136/BMJ.E8105

Halgren, M., Ulbert, I., Bastuji, H., Fabó, D., Eross, L., Rey, M., … Cash, S. S. (2019). The generation and propagation of the human alpha rhythm. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(47), 23772–23782. https://doi.org/10.1073/PNAS.1913092116/SUPPL_FILE/PNAS.1913092116.SAPP.PDF

Hanowski, R. J., Bowman, D. S., Alden, A., Wierwille, W. W., & Carroll, R. (2008). *PERCLOS+: Development of a Robust Field Measure of Driver Drowsiness Evaluation of Light-Vehicle Driver Education Programs Targeting Sharing the Road with Heavy Vehicles: Case Study Analysis View project Federal Motor Carrier Safety Administration's Advan*. Retrieved from https://www.researchgate.net/publication/268504751

Hayashi, H., Oka, N., Kamezaki, M., & Sugano, S. (2021). Development of a Situational Awareness Estimation Model Considering Traffic Environment for Unscheduled Takeover Situations. *International Journal of Intelligent Transportation Systems Research*, *19*(1), 167–181. https://doi.org/10.1007/S13177-020-00231-4/FIGURES/17

Hernández, L. G., Mozos, O. M., Ferrández, J. M., & Antelis, J. M. (2018). EEG-based detection of braking intention under different car driving conditions. *Frontiers in Neuroinformatics*, *12*, 29. https://doi.org/10.3389/FNINF.2018.00029/BIBTEX

Hjälmdahl, M., Krupenia, S. & Thorslund, B. (2017). Driver behaviour and driver experience of partial and fully automated truck platooning – a simulator study. *Eur. Transp. Res. Rev*, *9*, 8. https://doi.org/10.1007/s12544-017-0222-3

Jagannath, M., & Balasubramanian, V. (2014). Assessment of early onset of driver fatigue using multimodal fatigue measures in a static simulator. *Applied Ergonomics*, *45*(4), 1140–1147. https://doi.org/10.1016/J.APERGO.2014.02.001

Jonah, B. A. (1986). Accident risk and risk-taking behaviour among young drivers. *Accident Analysis & Prevention*, *18*(4), 255–271. https://doi.org/10.1016/0001-4575(86)90041-2

Ju, J., Feleke, A. G., Luo, L., & Fan, X. (2022). Recognition of Drivers' Hard and Soft Braking Intentions Based on Hybrid Brain-Computer Interfaces. *Cyborg and Bionic Systems*, *2022*, 1–13.

https://doi.org/10.34133/2022/9847652

Kapitaniak, B., Walczak, M., Kosobudzki, M., Jóźwiak, Z., & Bortkiewicz, A. (2015, December 1). Application of eye-tracking in drivers testing: A review of research. *International Journal of Occupational Medicine and Environmental Health*. Walter de Gruyter GmbH. https://doi.org/10.13075/ijomeh.1896.00317

Khan, M. Q., & Lee, S. (2019). Gaze and Eye Tracking: Techniques and Applications in ADAS. *Sensors (Basel, Switzerland)*, *19*(24). https://doi.org/10.3390/S19245540

Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J., & Ramsey D.J. (2006). The Impact of Driver Inattentionnn On Near-Crash/Crash Risk::: An Analysis Using the 100-Car Naturalistic Driving Study Data.

Knoflacher, H. (1979). Altersspezifische Unfallumstände. *Zeitschrift Für Verkehrssicherheit*, *25*, 131–134.

Kontotasios, A. (2020). *Fatigue in pilots who commute by driving : identification of the problem and investigation with driving and flying simulation*. University of Leeds. Retrieved from https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.826761

Körber, M. (2019). Theoretical considerations and development of a questionnaire to measure trust in automation. *Advances in Intelligent Systems and Computing*, *823*, 13–30. https://doi.org/10.1007/978-3-319-96074-6_2/COVER

Körber, M., Baseler, E., & Bengler, K. (2018). Introduction matters: Manipulating trust in automation and reliance in automated driving. *Applied Ergonomics*, *66*, 18–31. https://doi.org/10.1016/J.APERGO.2017.07.006

Körber, M., Cingel, A., Zimmermann, M., & Bengler, K. (2015). Vigilance Decrement and Passive Fatigue Caused by Monotony in Automated Driving. *Procedia Manufacturing*, *3*(December), 2403–2409. https://doi.org/10.1016/j.promfg.2015.07.499

Korner-Bitensky, N., & Sofer, S. (2009). The DriveABLE Competence Screen as a predictor of on-road driving in a clinical sample. *Australian Occupational Therapy Journal*, *56*(3), 200–205. https://doi.org/10.1111/J.1440-1630.2008.00749.X

Krigolson, O. E., Hammerstrom, M. R., Abimbola, W., Trska, R., Wright, B. W., Hecker, K. G., & Binsted, G. (2021). Using Muse: Rapid Mobile Assessment of Brain Performance. *Frontiers in Neuroscience*, *15*, 56. https://doi.org/10.3389/FNINS.2021.634147/BIBTEX

Krigolson, O. E., Williams, C. C., Norton, A., Hassall, C. D., & Colino, F. L. (2017). Choosing MUSE: Validation of a low-cost, portable EEG system for ERP research. *Frontiers in Neuroscience*, *11*(MAR), 109. https://doi.org/10.3389/fnins.2017.00109

Lal, S. K. L., & Craig, A. (2001). A critical review of the psychophysiology of driver fatigue. *Biological Psychology*, *55*(3), 173–194. https://doi.org/10.1016/S0301-0511(00)00085-5

Larsen, R. S., & Waters, J. (2018). Neuromodulatory correlates of pupil dilation. *Frontiers in Neural Circuits*, *12*, 21. https://doi.org/10.3389/FNCIR.2018.00021/BIBTEX

Larue, G. S., Rakotonirainy, A., & Pettitt, A. N. (2011). Driving performance impairments due to hypovigilance on monotonous roads. *Accident Analysis & Prevention*, *43*(6), 2037–2046. https://doi.org/10.1016/j.aap.2011.05.023

Le, A. S., Suzuki, T., & Aoki, H. (2020). Evaluating driver cognitive distraction by eye tracking: From simulator to driving. *Transportation Research Interdisciplinary Perspectives*. https://doi.org/10.1016/j.trip.2019.100087

Lee, S., Kim, M., Jung, H., Kwon, D., Choi, S., & You, H. (2020). Effects of a Motion Seat System on Driver's Passive Task-Related Fatigue: An On-Road Driving Study. *Sensors*, *9*, 1–15. https://doi.org/10.3390/s20092688

Li, G., Huang, S., Xu, W., Jiao, W., Jiang, Y., Gao, Z., & Zhang, J. (2020). The impact of mental fatigue on brain activity: A comparative study both in resting state and task state using EEG. *BMC*

Neuroscience, *21*(1), 1–9. https://doi.org/10.1186/S12868-020-00569-1/FIGURES/6

Liu, D., Sun, P., Xiao, Y. Q., & Yin, Y. (2010). Drowsiness detection based on eyelid movement. *2nd International Workshop on Education Technology and Computer Science, ETCS 2010*, *2*, 49–52. https://doi.org/10.1109/ETCS.2010.292

Liu, J., Wang, C., Liu, Z., Feng, Z., & Sze, N. N. (2021). Drivers' Risk Perception and Risky Driving Behavior under Low Illumination Conditions: Modified Driver Behavior Questionnaire (DBQ) and Driver Skill Inventory (DSI). *Journal of Advanced Transportation*, *2021*. https://doi.org/10.1155/2021/5568240

Lutz, A., Slagter, H. A., Dunne, J. D., & Davidson, R. J. (2008). Attention regulation and monitoring in meditation. *Trends in Cognitive Sciences*, *12*(4), 163. https://doi.org/10.1016/J.TICS.2008.01.005

Matthews, G. (1993). Cognitive processes in driver stress. In *Proceedings of the 1993 International Congress of Health Psychology* (pp. 90–93). Tokyo: ICHP. Retrieved from https://www.google.com/search?q=Matthews%2C+G.+1993%2C+Cognitive+processes+in+driver+stress%2C+in+Proceedings+of+the+1993+International+Congress+of+Health+Psychology+(Tokyo%3A+ICHP)%2C+90±93.&rlz=1C1CHBF_enCA959CA959&oq=Matthews%2C+G.+1993%2C+Cognitive+pr

Matthews, G., Desmond, P. A., Joyner, L., Carcary, B., & Gilliland, K. (1997). A comprehensive questionnaire measure of driver stress and affect. *Traffic and Transport Psychology: Theory and Application*. Retrieved from http://trid.trb.org/view.aspx?id=635125

Matthews, G., Dorn, L., Hoyes, T. W., Davies, D. R., Glendon, A. I., & Taylor, R. G. (1998). Driver stress and performance on a driving simulator. *Human Factors*, *40*(1), 136–149. https://doi.org/10.1518/001872098779480569

Matthews, G., Dorn, L., & Ian Glendon, A. (1991). Personality correlates of driver stress. *Personality and Individual Differences*, *12*(6), 535–549. https://doi.org/10.1016/0191-8869(91)90248-A

May, J. F., & Baldwin, C. L. (2009). Driver fatigue: The importance of identifying causal factors of fatigue when considering detection and countermeasure technologies. *Transportation Research Part F: Traffic Psychology and Behaviour*, *12*(3), 218–224. https://doi.org/10.1016/j.trf.2008.11.005

Michiels, W., & Schneider, P. A. (1984). Traffic offences: Another approach to description and prediction. *Accident Analysis & Prevention*, *16*(3), 223–238. https://doi.org/10.1016/0001-4575(84)90016-2

Miyamoto, M., Shirai, Y., Nakayama, Y., Gembun, Y., & Kaneda, K. (2000). An epidemiologic study of occupational low back pain in truck drivers. *Journal of Nippon Medical School = Nihon Ika Daigaku Zasshi*, *67*(3), 186–190. https://doi.org/10.1272/jnms.67.186

Neubauer, C., Matthews, G., Saxby, D., & Langheim, L. (2010). Simulator methodologies for investigating fatigue and stress in the automated vehicle. *Advances in Transportation Studies*, (SPEC. ISSUE), 7–18. https://doi.org/10.4399/97888548416592

Noguchi, Y., Nopsuwanchai, R., Ohsuga, M., & Kamakura, Y. (2007). *LNAI 4562 - Classification of Blink Waveforms Towards the Assessment of Driver's Arousal Level - An Approach for HMM Based Classification from Blinking Video Sequence*. *LNAI* (Vol. 4562).

Nuñez, A., & Buño, W. (2021). The Theta Rhythm of the Hippocampus: From Neuronal and Circuit Mechanisms to Behavior. *Frontiers in Cellular Neuroscience*, *15*, 31. https://doi.org/10.3389/FNCEL.2021.649262/BIBTEX

Öz, B., Özkan, T., & Lajunen, T. (2010). Professional and non-professional drivers' stress reactions and risky driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *13*(1), 32–40. https://doi.org/10.1016/J.TRF.2009.10.001

Park, C., Shahrdar, S., & Nojoumian, M. (2018). EEG-Based classification of emotional state using an autonomous vehicle simulator. *Proceedings of the IEEE Sensor Array and Multichannel Signal Processing Workshop*, *2018-July*, 297–300. https://doi.org/10.1109/SAM.2018.8448945

Pauzié, A. (2008). A method to assess the driver mental workload: The driving activity load index (DALI). In *IET Intelligent Transport Systems* (Vol. 2, pp. 315–322). https://doi.org/10.1049/iet-its:20080023

Pauzié, A., & Manzano, J. (2007). Evaluation of driver mental workload facing new in-vehicle information and communication technology. In *Enhanced Safety in Vehicle Conference*. Retrieved from www.noehumanist.org

Qu, W., Zhang, Q., Zhao, W., Zhang, K., & Ge, Y. (2016). Validation of the Driver Stress Inventory in China: Relationship with dangerous driving behaviors. *Accident Analysis and Prevention*, *87*, 50–58. https://doi.org/10.1016/j.aap.2015.11.019

Randall, J. G., Oswald, F. L., & Beier, M. E. (2014). Mind-wandering, cognition, and performance: a theory-driven meta-analysis of attention regulation. *Psychological bulletin,* 140(6), 1411–1431. https://doi.org/10.1037/a0037428

Robb, M. J. M., & Mansfield, N. J. (2007). Self-reported musculoskeletal problems amongst professional truck drivers. *Ergonomics*, *50*(6), 814–827. https://doi.org/10.1080/00140130701220341

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the Eye Tracking Research and Applications Symposium 2000*, 71–78. https://doi.org/10.1145/355017.355028

Saxby, D. J., Matthews, G., Warm, J. S., Hitchcock, E. M., & Neubauer, C. (2013). Active and passive fatigue in simulated driving: discriminating styles of workload regulation and their safety impacts. *Journal of experimental psychology. Applied,* 19(4), 287–300. https://doi.org/10.1037/a0034386

Schlag, B. (1987). " In meinem Alter fährt man an sich recht zügig": Untersuchungen zu Risikoverhalten und Risikobereitschaft junger Fahrer. .

Sherman, M. A., Lee, S., Law, R., Haegens, S., Thorn, C. A., Hämäläinen, M. S., ... Jones, S. R. (2016). Neural mechanisms of transient neocortical beta rhythms: Converging evidence from humans, computational modeling, monkeys, and mice. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(33), E4885–E4894. https://doi.org/10.1073/PNAS.1604135113/SUPPL_FILE/PNAS.1604135113.SAPP.PDF

Soukupová, T. (2016). Real-Time Eye Blink Detection using Facial Landmarks. In *21st Computer Vision Winter Workshop*. Rimske Toplice, Slovenia,.

Spivey, M. J., Richardson, D. C., & Fitneva, S. A. (2004). Thinking outside the Brain: Spatial Indices to Visual and Linguistic Information. *The Interface of Language, Vision, and Action: Eye Movements and the Visual World* , 161–189.

Statistisches Bundesamt. (1987) (Wiesbaden). Kohlhammer, Stuttgart.

Tavakoli, A., Lai, N., Balali, V., & Heydarian, A. (2022). How are Drivers' Stress Levels and Emotions Associated with the Driving Context? A Naturalistic Study. https://doi.org/10.48550/arxiv.2205.06116

Thiffault, P., & Bergeron, J. (2003). Fatigue and individual differences in monotonous simulated driving. *Personality and Individual Differences*, *34*(1), 159–176. https://doi.org/10.1016/s0191-8869(02)00119-8

Thompson, W. T., Lopez, N., Hickey, P., DaLuz, C., Caldwell, J. L., & Tvaryanas, A. P. (2006). *Effects of Shift Work and Sustained Operations: Operator Performance in Remotely Piloted Aircraft (OP-REPAIR). Human Systems Wing (311th) Brooks AFB TX.* Retrieved from https://apps.dtic.mil/sti/citations/ADA443145

Ting, P. H., Hwang, J. R., Doong, J. L., & Jeng, M. C. (2008). Driver fatigue and highway driving: A simulator study. *Physiology & Behavior*, *94*(3), 448–453. https://doi.org/10.1016/J.PHYSBEH.2008.02.015

Tran, Y., Craig, A., Craig, R., Chai, R., & Nguyen, H. (2020). The influence of mental fatigue on brain activity: Evidence from a systematic review with meta-analyses. *Psychophysiology*, *57*(5).

https://doi.org/10.1111/PSYP.13554

Van der Laan, J., Heino, & Waard. (1996). A Simple Procedure for the Acceptance of Advanced Transport Telematics. Retrieved March 30, 2021, from http://www.hfes-europe.org/accept/accepttrc97OK.pdf

Wang, Z., Zheng, R., Kaizuka, T., & Nakano, K. (2019). Relationship between Gaze Behavior and Steering Performance for Driver-Automation Shared Control: A Driving Simulator Study. *IEEE Transactions on Intelligent Vehicles*, *4*(1), 154–166. https://doi.org/10.1109/TIV.2018.2886654

Weber, E. U., Blais, A. R., & Betz, N. E. (2002). A domain-specific risk-attitude scale: measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making*, *15*(4), 263–290. https://doi.org/10.1002/BDM.414

Wolkin, J. R. (2015). Cultivating multiple aspects of attention through mindfulness meditation accounts for psychological well-being through decreased rumination. *Psychology Research and Behavior Management*, *8*, 171. https://doi.org/10.2147/PRBM.S31458

Yeon Kim, J., & Gu Ji, Y. (2013). A Comparison of Subjective Mental Workload Measures in Driving Contexts. *Journal of the Ergonomics Society of Korea*, *32*(2), 167–177. https://doi.org/10.5143/JESK.2013.32.2.167

Yuen, N. H., Tam, F., Churchill, N. W., Schweizer, T. A., & Graham, S. J. (2021). Driving With Distraction: Measuring Brain Activity and Oculomotor Behavior Using fMRI and Eye-Tracking. *Frontiers in Human Neuroscience*, *15*, 413. https://doi.org/10.3389/FNHUM.2021.659040/BIBTEX

Zakerian, S. A., Zia, G., Nasl Seraji, G., Azam, K., & Mortezapour, A. (2018). Reliability and Validity of the Driver Activity Load Index for Assessing Mental Workload among Drivers in Production Companies. *Journal of Occupational Hygiene Engineering*, *5*(2), 65–71. https://doi.org/10.21859/JOHE.5.2.65

Zhang, Y., Ma, J., Zhang, C., & Chang, R. (2021). Electrophysiological frequency domain analysis of driver passive fatigue under automated driving conditions. *Scientific Reports 2021 11:1*, *11*(1), 1–9. https://doi.org/10.1038/s41598-021-99680-4

Zhou, F., Alsaid, A., Blommer, M., Curry, R., Swaminathan, R., Kochhar, D., … Tijerina, L. (2021). Predicting Driver Fatigue in Automated Driving with Explainability, 1–30. Retrieved from http://arxiv.org/abs/2103.02162

Zoellick, J. C., Kuhlmey, A., Schenk, L., & Blüher, S. (2021). Method-oriented systematic review on the simple scale for acceptance measurement in advanced transport telematics. *PLOS ONE*, *16*(3), e0248107. https://doi.org/10.1371/JOURNAL.PONE.0248107

Zuckerman, M. (1979). Sensation Seeking and Risk Taking. *Emotions in Personality and Psychopathology*, 161–197. https://doi.org/10.1007/978-1-4613-2892-6_7