FEASIBILITY STUDY FOR THE ON-TRACK TESTING
OF USER EXPERIENCE UNDER TRUCK PLATOONING
CONDITIONS

Contract# T8080-190419

Prepared by:
Jeffrey S. Hickman, Ph.D.

Virginia Tech Transportation Institute
Virginia Polytechnic Institute and State University
3500 Transportation Research Plaza (0536)
Blacksburg, VA 24061
FEASIBILITY STUDY FOR THE ON-TRACK TESTING OF USER EXPERIENCE UNDER TRUCK PLATOONING CONDITIONS

INTRODUCTION

Surface transportation, which relies heavily on large trucks, has become the primary means of transporting goods. In contrast to the transportation system’s gradual evolution, vehicle technology is undergoing rapid changes that could affect all types of road transportation, and its effects on trucking could have a particularly important effect on society. Increasing demand for consumer goods and just-in-time inventory strategies (i.e., receiving goods only as they are needed) place a significant demand on truck drivers as well as the Canadian highway system as more and more goods are delivered by trucks.

Traffic congestion is one of the most critical challenges compromising the efficiency of the transportation system. Each year, delays keep travelers stuck in their vehicles for hours, wasting fuel. Traffic congestion leads to higher crash rates and negative environmental impacts resulting from increased CO\textsuperscript{2} emissions and noise. All of these congestion effects degrade the public’s quality of life.

Platooning is one connected automated vehicle (CAV) feature that can potentially curb energy consumption and greenhouse gas emissions in the transportation sector. Platooning is a demonstrated method of two or more trucks traveling close together with the following vehicle(s) actively coordinated in formation at high speed by the lead vehicle. This has the potential to reduce energy consumption and greenhouse gas emissions resulting from aerodynamic drag. Trucks are ideal applications for platooning due to their technical characteristics and mode of operation. And, from an economic perspective, increased fuel economy has a significant benefit to fleets, as ~25\% of truck costs are associated with fuel.\textsuperscript{1}

Fuel Economy

Lammert et al.\(^2\) compared and contrasted six truck platooning studies, including works from the North American Council for Freight Efficiency (2013), National Renewable Energy Laboratory (NREL, 2014), Auburn University (2015), and National Research Council (NRC) Canada/Lawrence Berkeley National Laboratory (2016), along with wind tunnel test results from Lawrence Livermore National Laboratory (LLNL), and computational fluid dynamics simulations from Denso. Figure 1 shows the comparison of these studies, with the percent of fuel savings on the y-axis and the following distance on the x-axis. Results indicated that the following platooning vehicle experiences a benefit at longer than anticipated distances, but still experiences a reduction in savings at distances closer than 50 ft. Lammert et al.\(^3\) indicated the best team savings for a 65,000 pound truck was a speed of 55 mph and a following distance of 30 ft.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Platooning Evaluations Comparisons in Lammert et al.\(^2\)}
\end{figure}


Potential for Fatigue in Platooning

The terms “drowsiness” and “fatigue” are used interchangeably as drowsy driver and driver fatigue; however, they are defined very differently. Fatigue is a state of reduced physical or mental alertness that impairs performance and is often the result of physical or mental exertion, whereas drowsiness is the inclination to sleep resulting from lack of sleep, boredom, hunger, or other outside factors. Here we are most concerned with the former, which, during platooning, may result in active fatigue in the lead vehicle driver or passive fatigue in the following driver, based on workload regulation. Active fatigue (or overload conditions) may be experienced in driving during high workload conditions, such as driving in inclement weather, traffic, and, possibly, as the lead driver in a platoon. Passive fatigue (underload conditions or hypovigilance) is most likely induced in low workload conditions when the driver assumes more of a monitoring role and less control of the vehicle, such as when the following driver(s) is in a truck platoon.

The most robust impact of fatigue is called the vigilance decrement. Via its effect on alertness, fatigue decreases operators’ abilities to maintain their attention on a source of signal for a prolonged period in order to detect critical signals that are necessary to perform the task at hand. The vigilance decrement, or hypovigilance, happens early in the fatigue process and is associated with various performance failures that can lead to crashes, including signal omissions, loss of situational awareness, and slow reaction times.

Per the Statement of Work, the rule of thumb is that the level of vigilance is always the result of a systemic interaction between the basic endogenous state of the organism, the tasks that a person is performing, and under what conditions those tasks are being performed. The main influences are endogenous, as they determine the base level of alertness, but a low-demanding monotonous task has an exogenous impact on alertness that is enough to create hypovigilance even in well-rested individuals. Science is clear to the effect that task monotony has a negative impact on alertness and vigilance and that it is a fatigue facilitator. Specifically, monotony has the potential to exacerbate or to counteract, to some extent, the impact of endogenous factors by pulling alertness towards the sleep end of the sleep-wake continuum. The exogenous task-related effect of monotony on alertness and vigilance is called passive fatigue.

There are theoretical arguments and growing empirical evidence\textsuperscript{6} that it is counter-productive to partially remove drivers from the driving task while relying on them to monitor the traffic environment, as the following driver(s) in a truck platooning scenario are often expected to do. If drivers are expected to remain vigilant, science indicates they should be actively involved in the control of the task. Tracking and speed management are analogous tasks that necessitate continuous real-time adaptation to an ever-changing environment, providing a means for the driver to maintain alertness and remain in tune with the driving situation. Nevertheless, even under those normal driving conditions, a significant number of fatigue-related crashes take place. Currently, there are a growing number of reasons to believe these numbers could rise if the task monotony increases by removing lateral and/or longitudinal control while still relying on the driver to monitor the traffic environment. As such, it appears that task-induced fatigue is a problem that needs to be mitigated in automated driving, or at least partial automation, in order for platooning to be successful from a safety perspective.

The extant literature indicates that vehicle automation technologies are associated with passive fatigue and loss of situational awareness.\textsuperscript{6,7,8,9,10,11,12} The relationship is mainly explained by the fact that automation increases task monotony, which decreases alertness, creates a state of hypovigilance, and facilitates the transition towards the sleep end of the sleep-wake continuum. A critical issue is driver vigilance in intermediate levels of automation (SAE International Level 1 to level 3 [L1 to L3]), where the driver is required to monitor the driving environment for hazards (potentially in truck platooning). Various dependent measures, including electroencephalogram

\textsuperscript{7} Hjalmdahl, M., Krupenia, S., & Thorslund, B. (2017). Driver behavior and driver experience of partial and fully automated truck platooning – A simulator study. European Transportation Research Review, 9(8), 1-11.
subjective ratings of alertness and workload, ocular measures, cardiac measures, reaction time to a stimulus, and steering measures have been shown to be effective in identifying operator passive fatigue under vehicle automation. The latter, though effective, is not an appropriate dependent measure for vehicle automation that is SAE L2 or greater, given that the vehicle in SAE L2 maintains lateral and longitudinal control.

Although commercial platooning developers have privately performed extensive testing with fleet customers, in the public sphere, commercial vehicle platooning has been demonstrated in small-scale testing only, mostly with respect to fuel economy. A simulator experiment on truck platooning confirmed the relationship between automation, hypovigilance, and fatigue; however, the apparatus (i.e., a simulator) is more likely to induce fatigue due to drivers’ lack of motivation in a simulated driving environment. The only published assessment of driver drowsiness in truck platooning in the real-world comes from Daimler Trucks, where the attentiveness of 16 truck drivers on a test track was assessed while using the Highway Pilot system for 4 hours without a break. Fatigue was measured with an EEG and electrocardiogram (ECG). Results indicated a 25% reduction in fatigue and subjective reports of better attentiveness when in autonomous mode; however, the drivers in this study could perform interesting secondary tasks (e.g., use a tablet computer) while piloting the vehicles.

A comprehensive deployment plan needs to be developed in order to facilitate larger-scale testing and deployment to address questions important to all highway users. One issue is driver vigilance in automated driving. Little real-world public data are available on potential effects on drivers in platooning operations where the following driver (in a platoon with two or more trucks) is required to monitor the driving environment for potential hazards (i.e., as in SAE L1 to L3 vehicles). And, those that have been published evaluated platooning over a short period of time (less than 60 min). In these situations, engagement in secondary tasks, especially visual non-driving tasks, which is illegal to perform while driving, may limit the ability of drivers to monitor the traffic situation and

respond to a safety-critical issue such as a take-over-request or silent failure. The Daimler Trucks platooning study\textsuperscript{13} is not an appropriate evaluation of driver vigilance in SAE L1 to L3 applications, as drivers could perform an illegal non-driving task during the study. Secondary tasks may result in increased driver vigilance, but also likely result in decreased situational awareness, especially in SAE L1 to L3 applications where the driver is required to monitor the driving environment for hazards. The proposed study will address this gap by evaluating driver vigilance in an SAE L1 and/or L2 truck platooning scenario where the lead and following drivers are required to monitor the driving environment.

METHODS OVERVIEW

A high-level overview of the proposed methods is presented in this section, including the design, power analysis, participants, setting, technologies, dependent variables, analysis approach, timeline, and cost estimates. This overview does not provide a description of specific research protocols or instructions to participants.

Design

Two different designs are presented, which reflect differing costs and periods of performance. The scenarios are meant to capture, at a minimum, the required elements provided in the Statement of Work. This assumes a platoon with two trucks; the addition of a third truck to the platoon will reduce the number of data collection days. Blanco et al. assessed the impact of driving hours, work hours, and breaks on driving performance in a naturalistic truck driving study in a sample of 97 U.S. drivers. Using a hybrid log consisting of video and drivers’ self-reports, they found that, on average, drivers drove for approximately 8.5 hours per duty period, with the fourth driving hour being the most common hour for a break.\textsuperscript{19} Note that this is from a sample of U.S. drivers and may not reflect average Canadian driver schedules, where the hours-of-service allow greater driving and on-duty periods compared to the U.S.

It is assumed the maximum driving hours for participants will be 8 hours, and 8 hours is the assumed maximum amount of daily data collection at the test track. As the Statement of Work indicated, one of the independent variables is the driver (lead platoon truck X with following platoon truck), and this was included as an independent variable. This independent variable can be eliminated to reduce the number of required participants if a membe(s) of the research team is the

lead platoon truck driver. However, lead vehicle driver vs. following vehicle driver is viewed as the primary comparison in the proposed study with respect to driver vigilance under truck platooning conditions, and this comparison should therefore be kept if possible.

**Scenario 1: Between Subjects Design**

Scenario 1 assumes a between-subjects 2 Driver (lead platoon truck, following platoon truck) X 2 SAE Level (SAE L1, SAE L2) X 2 Following Distance (Distance 1, Distance 2) design. SAE Level 1 is when the headway of the following vehicle is controlled by the CAV (i.e., the following platoon driver is required to steer the truck). SAE Level 2 is when the headway and lateral control of the following platoon truck is controlled by the CAV (i.e., the following platoon driver is required to monitor the roadway). Scenario 1A assumes 8 hours of driving in data collection running one group per day (see Figure 2), Scenario 1B assumes 4 hours of driving in data collection running two groups per day (see Figure 3).

![Figure 2](image-url)  
**Figure 2. Scenario 1A with 8 Hours of Driving in Data Collection (One Group per Day).**

![Figure 3](image-url)  
**Figure 3. Scenario 1B with 4 Hours of Driving Data Collection (Two Groups per Day).**

The advantages of this design include minimizing potential learning effects across conditions, shorter time investment for each participant (i.e., less likely to lose participants as they only need to attend one experimental condition), and easier to setup (i.e., no need to counterbalance conditions). The disadvantages of this design include the inability to control for potential individual differences that may affect the dependent variables, such as susceptibility to fatigue, personality, driving style, etc., and more participants to recruit, which could increase the time necessary to recruit a sufficient number of qualifying participants.
Scenario 2: Within Subjects Design

Scenario 2 assumes a within-subjects 2 Driver (lead platoon truck, following platoon truck) X 2 SAE Level (SAE L1, SAE L2) X 2 Following Distance (Distance 1, Distance 2) design. Scenario 2A assumes 8 hours of driving in data collection, Scenario 2B assumes 4 hours of driving in data collection. Scenario 2 is identical to Scenario 1; however, assignment to the independent variables is counterbalanced using a Latin Square. As there are an even number of conditions (six), the first row of the Latin Square will follow the formula 1, 2, n, 3, n-1, 4, n-2, where n is the number of conditions. For subsequent rows, one is added to the previous, returning to 1 after n. Table 1 and Table 2 illustrate the presentation of the six conditions (columns) for each participant (rows) and four conditions (columns) for each participant (rows), respectively. The presentation of the conditions can be counterbalanced if only two conditions are present.

Table 1. Latin Square Design Using Six Experimental Conditions.

<table>
<thead>
<tr>
<th>Participant</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
<th>5th</th>
<th>6th</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>B</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>C</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>D</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>6</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>E</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>F</td>
<td>6</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 2. Latin Square Design Using Four Experimental Conditions.

<table>
<thead>
<tr>
<th>Participant</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>B</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>C</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>D</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>E</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>F</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>
The advantages of this design include the ability to control for potential individual differences that may affect the dependent variables (i.e., each participant is in each experimental condition) and the fact that there will be fewer participants to recruit. The disadvantages of this design include potential learning effects across conditions; longer time investment for each participant, leading to increased likelihood of participant attrition, as they need to attend each experimental condition; and more difficult setup (i.e., counterbalancing conditions).

**Power Analysis**

The power analysis was conducted with respect to the primary dependent variable of driver fatigue. It is assumed the candidate technology will measure PERCLOS, which is one of the most robust measures of driver fatigue. The power analysis will require re-calculation if the candidate technology measures a different variable. There were no studies that evaluated driver fatigue as a function of following distance in a platooning scenario, lead driver vs. following driver, or SAE Level 1 vs. Level 2. However, Jarosch et al.\textsuperscript{16} assessed driver fatigue in a highly automated vehicle. The drive was in a simulator and only lasted 50 minutes; thus, it can be interpreted in terms of the power analysis with the knowledge the drive was relatively short compared to the proposed study and it was conducted in a driving simulator, where fatigue is more likely to be induced. One condition involved a monitoring task, similar to what is proposed under SAE Level 2, while the other condition involved drivers performing a non-driving related task under highly automated driving. The power analysis considers the former condition to reflect the vigilance of drivers under normal circumstances, where they are actively engaged, controlling the vehicle’s headway and lateral movement.

The power analysis used the mean PERCLOS in the monitoring task group at (M = 8.92), the mean PERCLOS in the non-driving related tasks group (M = 3.39), and an overall standard deviation of 5.82.\textsuperscript{17} There are six conditions (or “k”) in the 2 Driver (lead platoon truck, following platoon truck) X 2 SAE Level (SAE 1, SAE 2) X 2 Following Distance (Distance 1, Distance 2) design, which reflect 15 pair-wise comparisons [k(k-1)/2]). Using a power of 0.80 and an alpha of 0.05, the required sample size is 32 participants in Scenario 2 and 192 total participants in Scenario 1 (32 participants in each condition). If two of the conditions are eliminated (e.g., Following Distance), this reduces the number of pair-wise comparisons to six, which would reduce the required sample size to 27 participants in Scenario 2 and 108 total participants in Scenario 1 (27 participants in each condition). If four of the experimental conditions are eliminated (e.g., Following Distance and SAE Level), this reduces the number of pair-wise comparisons to one,
which would reduce the required sample size to 18 participants in Scenario 2 and 36 total participants in Scenario 1 (18 participants in each condition).20

This last power analysis could also serve to inform the number of participants (18) in an exploratory analysis on driver vigilance in the following driver; the lead driver would be a member of the research team in this case. This would reduce cost and the overall period of performance; however, the disadvantage of this exploratory analysis is the lack of a lead driver vs. following driver comparison. Thus, this type of exploratory analysis would not be able to determine if any of the dependent variables differ as a function of lead driver vs. following driver, which is viewed as an important comparison with respect to driver vigilance.

Use of steering measures and lateral position have been found to be similar in accuracy as PERCLOS in detecting fatigue.21,22 Thus, the outputs in the power analysis should be acceptable. However, Jarosch et al.16 used an SAE L2 scenario; thus, more participants should be added in an SAE L1 scenario irrespective of the experimental setting (as task-induced fatigue is more likely in an L2 scenario). Moreover, Jarosch et al.16 used a driving simulator where it is easier to induce fatigue/drowsiness due to lower driver motivation; thus, this should be considered when interpreting the power analysis for use in a test track or field operational test.

Caution: consider increasing the sample size if using an SAE L1 scenario or data collection takes place in a naturalistic setting (i.e., a field operational test).

Participants

Eligible participants will be active Canadian truck drivers with a valid Commercial Driver’s License and medical report. Active means the driver is currently employed as a commercial motor vehicle driver (i.e., transporting goods) for a Canadian-domiciled company, or an owner operator, that operates a Class 8 (> 33,000 lbs.) articulated truck. These drivers should be employed at a company in line-haul or long-haul operations (short-haul operations should be excluded). Drivers

with untreated sleep disorders should be excluded as should drivers who operate primarily at night (i.e., their duty period starts at 4 p.m. or later). If possible, the study should exclude drivers with experience driving a truck with automatic cruise control and/or automatic emergency braking, unless this significantly reduces the potential participant pool during recruitment.

Although there should be no general restrictions on age or gender, a range of ages should be included (e.g., 23 to 65 to exclude newly licensed drivers and older drivers). Age, which is a proxy measure of experience, can be used as a covariate in the analysis. However, drivers with 2 years or less experience in driving a Class 8 articulated truck should be excluded. A driver’s working hours since their last restart, working hours the day prior to data collection (if they drove), and sleep in the last 24 hours before data collection should be reported and used as covariates in the analysis. Drivers should be instructed to avoid caffeine in the 24 hours prior to data collection and there should be no non-driving secondary tasks performed during data collection. Although these procedures limit generalization to real-world driving, where caffeine and some non-driving secondary task are allowed, they are meant to induce fatigue under shorter driving conditions. Participants should be reminded of participation and given the study location 48 hours prior to their scheduled date and time to limit no-shows.

Setting

The proposed study will be performed on a test track. Start times for each scenario should be identical to control for circadian effects.

Dependent Variables

Below is a comprehensive list of the suggested dependent variables in the proposed study.

Technologies

Real-time Fatigue Detection: Appendix A lists real-time fatigue detection technologies in order of validity and reliability. Smart Eye was listed as the best candidate technology given that it uses an unobtrusive, one camera system to detect over 100 raw ocular metrics of fatigue, resulting in a more robust and reliable system. Measured parameters include raw ocular measures, such as pupil diameter, blink frequency, blink durations, and PERCLOS (e.g., can be set at percentage of time the eyes are closed more than 75% in 10,000 frames). The system can detect head and ocular measures if participants use their prescription glasses and/or sunglasses; thus, there are no restrictions in excluding these potential participants during recruitment. Eye tracking can be added with the purchase of additional cameras. The Dikablis glasses were rated as the second candidate
technology, and provides similar outputs, but requires participants to wear the technology as frames during the experimental drive; these can also be worn over regular eyeglasses and sunglasses. Note the SOW requested real-time fatigue detection technologies; however, as the system in the proposed study will not warn the driver regarding his/her fatigue/drowsiness, any technology that reliably measures and records ocular measures will be sufficient.

The suggested real-time fatigue detection technologies assume SAE L2 platooning. However, in an SAE L1 platooning scenario, steering and lane position measures could be used to reliably detect drivers’ fatigue and drowsiness. Signs of driver drowsiness and fatigue include a reduction in the number of micro-corrections to the steering wheel, rough steering adjustments, zigzags and slow oscillations, larger erratic steering movements and corrections,\textsuperscript{23,24} standard deviation of lateral position, and unintentional lane deviations\textsuperscript{25}. Verster & Roth\textsuperscript{23} and SAE\textsuperscript{26} provide guidance on procedures for using driving performance measures. High-precision devices (along with a data logger) that measure steering moment, steering angle, and steering speed are relatively low cost (~$1,000). A video camera-, laser-, or infrared-based system (along with a data logger) that detects and records lane position relative to the left and right lane markers can be used to measure the standard deviation of lateral position, and unintentional lane deviations. However, use of these dependent measures precludes comparisons with SAE L2 systems.

**Psychomotor Vigilance Test (PVT):** The PVT was invented by Dr. David F. Dinges, through support from the U.S. Office of Naval Research, and has been validated to detect slowing of psychomotor speed and lapses of attention\textsuperscript{27} as well as vigilance decrements and instability in behavioral alertness, which are common adverse effects on performance of fatigue due to inadequate sleep, wakefulness at night, and prolonged time-on-task.\textsuperscript{28} The PVT also outperformed other widely used brief performance measures of fatigue. In a comparison of various cognitive

performance tests known to be sensitive to fatigue induced by sleep loss, investigators at the Walter Reed Army Institute of Research concluded that “the [PVT] was among the most sensitive to sleep restriction, was among the most reliable with no evidence of learning over repeated administrations, and possesses characteristics that make it among the most practical for use in the operational environment.” Dr. Dinges and colleagues empirically developed an algorithm for PVT stimulus delivery rate and response quantification that resulted in a shorter, 3-min PVT-B. These experiments demonstrated that performance on the 3-min PVT-B tracked with performance on the 10-min PVT throughout total and partial sleep loss. The PVT-B will be used to provide data on drivers’ behavioral alertness.

**Situation Awareness:** The Detection Response Task (DRT) can be used to assess situation awareness (i.e., measuring a response, such as a button press to a visual or auditory cue). However, this task is likely to increase drivers’ vigilance and minimize the fatigue manipulation. Another variable used to assess situation awareness is the driver’s reaction time to a takeover request; however, a takeover request will not be manipulated in the proposed study. Lastly, eye tracking could be used as a proxy for situation awareness, with gaze behavior toward the forward roadway and mirror checks compared to glances away from these locations. However, gaze behavior is limited as a proxy measure for situation awareness as it cannot determine “looked, but didn’t see,” and the technology is not viewed as reliable when used in real-world driving; gross measures of gaze behavior, such as on/off road, are reliable, but specific locations are not. Thus, there is not likely to be a solution for this dependent variable unless the DRT is used (with the understanding that it will increase drivers’ vigilance). Or, a compromise might be to use a DRT, such as a takeover request, once per drive, likely at the end of the experimental drive.

**Questionnaires**

**Demographic Questionnaire:** The research team will create a demographic questionnaire. Items on the questionnaire are expected to include the following: age, gender, weight, height, driving experience, prior crash rate, prior moving violation rate, experience with automated driving system (e.g., automated braking systems, automatic cruise control), caffeine use in last 48 hours, sleep in the last 24 hours, and typical driving schedule over the last month (e.g., daily driving hours, weekly driving hours, start/stop times).


Karolinska Sleepiness Scale (KSS): This scale measures the subjective level of sleepiness. On this scale, subjects indicate which level best reflects the psycho-physical state experienced in the last 10 min. The KSS is a measure of situational sleepiness and is sensitive to fluctuations. The KSS is a 9-point scale, where 1 = extremely alert, 3 = alert, 5 = neither alert nor sleepy, 7 = sleepy – but no difficulty remaining awake, and 9 = extremely sleepy – fighting sleep.31

Short Stress State Questionnaire (SSSQ): Matthews et al.32 presented psychometric and experimental evidence based on studies with the Dundee Stress State Questionnaire (DSSQ). The authors, in their developed SSSQ, identified three broad higher-order state factors: task engagement, distress, and worry. While the DSSQ has been found to be a reliable and useful measure of the three higher-order dimensions of subjective stress state, it contains 90 items and is therefore quite lengthy and time consuming to complete. The SSSQ is based on the DSSQ, but includes only 24-items, which makes it useful in applied settings.33

NASA Task Load Index (NASA-TLX): The NASA-TLX is a standard measure of workload based on ratings of task demands and subjective reactions to the task. Total workload is divided into six subjective subscales that are represented on a single page, serving as one part of the questionnaire: mental demand, physical demand, temporal demand, performance, effort, and frustration. Each subscale is rated for each task within a 100-points range with 5-point steps.34

Human Trust in Automation Systems Scale (HTASS): Jian et al. constructed a multi-item measure of operator trust in automated systems. The scale contains 12 items; seven assess trust and five assess distrust in automated systems. Means for trust and distrust can be computed.35

Van der Laan Acceptance Scale: This scale assesses “system” acceptance on two dimensions: a usefulness scale and an affective satisfying scale. The scale has a total of nine 5-point rating scale items.36

Analysis Approach

Most of the dependent variables will be analyzed with mixed-design analysis of variances (ANOVAs). The addition of covariates, such as sleep in the prior 24 hours, would necessitate a mixed-design analysis of co-variance (ANCOVA). Parametric assumptions (normal distribution, homogeneity of variances, and sphericity, where applicable) should be examined for each dependent variable. If the sphericity assumption is violated, a Greenhouse-Geisser can be used to correct the F-value.

Strawman Procedures

A high-level, step-by-step procedure for the 8-hour drive is described below. The 4-hour drive will be similar, but steps 7 to 9 would be eliminated.

1. Greet participant and review informed consent procedures.
2. Complete Demographic Questionnaire.
3. Experience with the truck, technology (if in an SAE L1 or L2 condition), and test track.
4. Instructions on test procedures and Smart Eye equipment.
5. Complete KSS and SSSQ.
6. Drive first 4-hour block.
7. Complete KSS, SSSQ, and NASA TXL followed by 30-min rest.
8. Complete KSS and SSSQ.
9. Drive second 4-hour block.
11. Thank participants for their time and arrange participant compensation.

Timeline

Table 3 displays the estimated timeline for each task by the test block and required sample size combination. The timeline assumes a two-vehicle platoon; addition of another truck will increase throughput and reduce the data collection timeline. Preparatory work (i.e., finalizing study procedures, institutional review board approval, test track preparation, etc.) is expected to take 6 months, recruitment is likely to take 2–3 months, dependent on the existing database of potential participants that meet inclusion/exclusion criteria and participant pay. Data collection is likely to be limited to weekends (Saturday and Sunday), as most drivers have these days off. There may be a few weekdays where a driver has an off-duty day; however, these were excluded in the calculation to account for potential missed test track opportunities due to poor weather, equipment malfunctions, and participant no-shows. Thus, these should be accurate estimates of the number of test track days.

Eight (8) hours of data collection limits the study to running two participants per day (lead vehicle driver and following vehicle driver) for a total of four participants each weekend (16 each month). This can be accomplished over

- 12 months for 32 participants per condition,
- 6.75 months for 27 participants per condition, and
- 2.5 months for 18 participants per condition.

Four hours of data collection reflects half the time, or

- 6 months for 32 participants per condition,
- 3.5 months for 27 participants per condition, and
- 1.25 months for 18 participants per condition.

Adding a third 4-hour block each day reduces the time to

- 4 months for 32 participants per condition,
- 2 months for 27 participants per condition, and
- 0.75 months for 18 participants per condition.

Data analysis is expected to take 6 months and completion of the final report is expected to take 3 months. The estimated timeline is presented in Table 3.
<table>
<thead>
<tr>
<th>Conditions</th>
<th>Tasks</th>
<th>Prep Work</th>
<th>Recruitment</th>
<th>Data Collection</th>
<th>Data Analysis</th>
<th>Final Report</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>8-hour Driving Block Each Day</td>
<td>32 per Condition</td>
<td>6 Months</td>
<td>3 Months</td>
<td>12 Months</td>
<td>6 Months</td>
<td>3 Months</td>
<td>30 Months</td>
</tr>
<tr>
<td></td>
<td>27 per Condition</td>
<td>6 Months</td>
<td>3 Months</td>
<td>6.75 Months</td>
<td>6 Months</td>
<td>3 Months</td>
<td>24.75 Months</td>
</tr>
<tr>
<td></td>
<td>18 per Condition</td>
<td>6 Months</td>
<td>2 Months</td>
<td>2.5 Months</td>
<td>6 Months</td>
<td>3 Months</td>
<td>19.5 Months</td>
</tr>
<tr>
<td></td>
<td>Exploratory (n = 18)</td>
<td>6 Months</td>
<td>2 Months</td>
<td>1.25 Months</td>
<td>6 Months</td>
<td>3 Months</td>
<td>18.25 Months</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two 4-hour Driving Blocks Each Day</td>
<td>32 per Condition</td>
<td>6 Months</td>
<td>3 Months</td>
<td>6 Months</td>
<td>6 Months</td>
<td>3 Months</td>
<td>24 Months</td>
</tr>
<tr>
<td></td>
<td>27 per Condition</td>
<td>6 Months</td>
<td>3 Months</td>
<td>3.5 Months</td>
<td>6 Months</td>
<td>3 Months</td>
<td>21.5 Months</td>
</tr>
<tr>
<td></td>
<td>18 per Condition</td>
<td>6 Months</td>
<td>2 Months</td>
<td>1.25 Months</td>
<td>6 Months</td>
<td>3 Months</td>
<td>16.25 Months</td>
</tr>
<tr>
<td></td>
<td>Exploratory (n = 18)</td>
<td>6 Months</td>
<td>2 Months</td>
<td>0.75 Months</td>
<td>6 Months</td>
<td>3 Months</td>
<td>15.75 Months</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Three 4-hour Driving Blocks Each Day</td>
<td>32 per Condition</td>
<td>6 Months</td>
<td>3 Months</td>
<td>4 Months</td>
<td>6 Months</td>
<td>3 Months</td>
<td>22 Months</td>
</tr>
<tr>
<td></td>
<td>27 per Condition</td>
<td>6 Months</td>
<td>3 Months</td>
<td>2 Months</td>
<td>6 Months</td>
<td>3 Months</td>
<td>20 Months</td>
</tr>
<tr>
<td></td>
<td>18 per Condition</td>
<td>6 Months</td>
<td>2 Months</td>
<td>0.75 Months</td>
<td>6 Months</td>
<td>3 Months</td>
<td>15.75 Months</td>
</tr>
<tr>
<td></td>
<td>Exploratory (n = 18)</td>
<td>6 Months</td>
<td>2 Months</td>
<td>0.5 Months</td>
<td>6 Months</td>
<td>3 Months</td>
<td>15.25 Months</td>
</tr>
</tbody>
</table>
As shown in Table 3, the Exploratory condition (18 participants) would be completed in 18.25 months, 15.75 months, and 15.25 months in the 8-hour driving block, two 4-hour blocks, and three 4-hour blocks, respectively.
APPENDIX A: VALIDATED REAL-TIME FATIGUE DETECTION TECHNOLOGIES

This inventory of existing (i.e., commercially available) fatigue detection technologies details key features, functions, and applications of systems used by, or that have applications in, the commercial motor vehicle industry. Here we limit the review to those technologies that have been shown effective in identifying fatigue. The technologies are rank ordered based on their functionality in the proposed study, with access to raw data outputs and sensor fusion given a higher priority.

1. **SmartEye** ([https://smarteye.se/](https://smarteye.se/))

Smart Eye’s AntiSleep technology uses one camera to track multiple fatigue factors, focusing on gaze direction, eyelid closure, and head position and orientation. AntiSleep uses 3-D mapping and an algorithm to derive real-time data output. The system uses a single standard VGA camera together with infrared flash illuminators tuned to frequencies that receive minimum interference from outdoor light, making the system robust to all-natural illumination conditions in automotive applications. It is currently an extensible system; although it does not give any specific alarms or feedback to the driver following the detection of fatigue, such features can be integrated by manufacturers as they wish. One-camera systems, such as the Smart Eye AntiSleep, are cheaper, easier to operate, and easier to install in a vehicle compared to multi-camera systems, which are more accurate and widely available. A one-camera system is most suitable for in-vehicle applications, such as systems that warn drivers of drowsiness or internal distractions.39

**Effectiveness**

No studies provided, used in several vehicle OEMs.

**Cost**

Cost is ~$20,000 for one camera system that detects ocular measures. Additional cameras can be purchased to provide eye tracking (~$45,000 for three camera system).

Pro and Cons in Proposed Research

Allows collection of raw outputs to be processed post-hoc. Measured parameters include raw ocular measures, such as pupil diameter, blink frequency, blink durations, and PERCLOS\textsuperscript{40} (e.g., can be set at percentage of time the eyes are closed more than 75\% in 10,000 frames). The system can detect head and ocular measures if participants use their prescription glasses and/or sunglasses; thus, there are no restrictions in excluding these potential participants during recruitment. Eye tracking can be added with the purchase of additional cameras.

2. Dikablis Eye Tracking

Ergoneers’ Dikablis Glasses are available in three versions—cable, mobile and wireless—and capture the eye-movement binocular at 60 Hz. They have a Full HD scene camera whereby the markers of the patented marker technology for automated gaze data analysis can be very small and unobtrusive. The Dikablis Professional Glasses are designed in such a way that there’s nearly no restriction in sight, which leads to a totally natural gaze behavior. This is combined with a comfortable and non-slip fit. The Dikablis Professional Glasses are wearable over normal eyeglasses and allow perfect adjustment to any given head-geometry or application area because of the adjustable scene- and eye-cameras.

Ergoneers D-Lab3 data acquisition and analysis software offers a measurement and analysis platform for user and behavior studies which allows flexible and synchronized recording of multiple data channels at different (native) frequencies and their common analysis. The following different data streams can be freely combined with each other: eye-tracking (head-mounted, remote and in head-mounted displays), videos, audio, physiology, motion data and “data streams“ via TCP/IP and CAN bus. The measurement and analysis modules can be combined flexibly, enabling users to tailor D-Lab to their exact requirements. D-Lab3 supports the whole testing process—from test planning and synchronous data capture through to results analysis and

visualization. The system has the ability to measure PERCLOS\textsuperscript{31} with eye tracking. Heart rate monitoring is also available and can be synchronized using the D-Lab3 software.

**Effectiveness**

No studies cited, potential to record PERCLOS\textsuperscript{31} and other ocular measures as well as eye tracking.

**Cost**

Glasses and software cost $36,000.

**Pro and Cons in Proposed Research**

Allows collection of raw ocular outputs to be processed post-hoc. Includes multiple channels of data that can be combined for a more robust measure of fatigue (if purchased). Also, includes eye tracking, which can be used to measure situational awareness. The device allows drivers to use their own prescription glasses and sunglasses so data collection can continue in all light conditions. The drivers must wear the glasses for the system to operate, and some individuals might find the glasses uncomfortable or feel they obstruct their view of the road. No effectiveness data are provided; however, given the outputs are raw data, PERCLOS\textsuperscript{34} and heart rate (if the heart rate monitor is purchased) can be combined into a multi-channel fatigue algorithm.

**3. Smartcap** (http://www.smartcaptech.com)

Life by SmartCap is a band of sensors that can be worn in a hat or by itself around the head. The Lifeband measures the ability to resist sleep by monitoring brainwaves. The Life system provides real-time feedback to the wearer and audibly and visually alerts the wearer before microsleeps occur through a display or an app on a smartphone or tablet. Though this technology relies on self-monitoring, it can alert management and/or family of fatigue levels.

**Effectiveness**

Evaluation of the SmartCap was conducted by Monash University. Researchers used the Osler (Oxford Sleep Resistance Test) task to evaluate the SmartCap. The Osler is a behavioral measure of sleep latency in which four consecutive misses are indicative of having brief periods of EEG-
defined sleep. Researchers found that an average score of 4 with the SmartCap (very drowsy) provided a high sensitivity (94.75%), correctly identifying most of the 1-minute periods when severe sleepiness was present. An average score of 4 with the SmartCap had a specificity of 81.9% with an area under the receiver operating characteristic curve of 0.89. Thus, it had a small to moderate false positive rate.\textsuperscript{41} In addition, the University of Chile determined the SmartCap uses signals that reliably represent EEG and reflect expected circadian patterns.\textsuperscript{42}

**Cost**

Must inquire for exact pricing. Costs include a one-time purchase of the required hardware and the potential purchase of a LifeDisplay; however, transport customers can integrate Life into their telematics system, so there may be no need for an additional screen. The charge for all software licensing, maintenance, and support is on a per user per month basis. A special version of the SmartCap, including raw EEG signals, was used by Bongers\textsuperscript{33} in his evaluation of the technology; thus, this might be an option in the proposed study.

**Pros and Cons in Proposed Research**

Life works in most operational environments. There are no known environmental limitations, such as day versus night or rainy versus dry weather. However, drivers are always required to wear the cap (or at a minimum the band) while driving and keep the battery charged for it to work. The Life battery needs to be charged while the driver sleeps. Alerts are given directly to the driver; however, fleet managers have the option of being made aware of the fatigue levels of their drivers. If fleet managers choose to be alerted to a fatigued driver, they must filter through and monitor all the fatigue alerts they receive and get in touch with fatigued drivers to address concerns. Drivers must feel comfortable wearing the Life sensors in order to remain compliant. Drivers must also clean the hat/band approximately every 3 months for optimal performance. Device is suitable if the alerts can be silenced and the raw data can be provided.


The Drowsimeter R100 uses images of the eye acquired at 120 Hz by a high-speed camera integrated into glasses to provide automatic, objective, and real-time measurements of several drowsiness and eye metrics in most lighting conditions. The level of drowsiness is calculated from several ocular parameters related to the eyelids’ movements (i.e., blinks) and eyeballs’ movements (i.e., saccades). The ergonomics and the high-frame rate of Phasya Glasses ensure accurate and continuous measurements without disturbing the user. The easy-to-use software (setup and calibration in less than 1 minute)—Drowsilogic—allows users to visualize and export all the data. The Drowsimeter R100 can be adapted with specific ocular parameters related to eyelid and eyeball movements, and pupil diameter can be provided upon request.

The standard version of Drowsimeter R100 outputs the following data:

- **Drowsiness metrics:** Level of Drowsiness (from 0 = fully awake to 10 = fully drowsy), PERCLOS 70, mean blink duration, blink frequency, percentage of LEYECLOS (= long eyelids closure), mean LEYECLOS duration;
- **Eye metrics:** Eyelids gap, pupil position, pupil diameter; and
- **Eye images.**

Prototype software allows data fusion, combining cardiac and ocular data. The combination of ocular and cardiac data enables a better accuracy of the drowsiness measurement. Phasya also develops software for monitoring stress, cognitive load, and mind wandering, etc.; however, they provide no details on how this was validated.

**Effectiveness**

Wertz et al. used a driving simulator with 14 participants to validate a fully automatic drowsiness monitoring system (software/algorithms) based on the subject’s physiological state. This system uses ocular parameters extracted from images of the eye (i.e., photooculography) to determine a level of drowsiness on a continuous numerical scale from “0 to 10,” with “0” corresponding to

---

"very awake" (or "very vigilant") and "10" to "very drowsy." The results show standard deviation of lane position increases when the (computed) level of drowsiness increases, and (2) that the level of drowsiness increases when the level of sleep deprivation increases. Francois et al.\textsuperscript{44} compared ocular parameters to the PVT and EEG measures in 24 participants in various sleep conditions. The Pearson’s correlation coefficient between photooculography and EEG-based levels of drowsiness was significant ($R = 0.54$, $p < 0.01$). The same finding was obtained for the correlation coefficient between the photooculography and the percentage of lapses in the PVT ($R = 0.52$, $p < 0.01$).

**Cost**

The cost is ~$25,000 for the Drowsimeter R100, which records ocular measures. Software to add stress and cognitive load measurements costs an additional ~$25,000.

**Pro and Cons in Proposed Research**

Allows collection of raw ocular outputs to be processed post-hoc. Includes multiple channels of data that can be combined for a more robust measure of fatigue. Limited information on monitoring stress, cognitive load, and mind wandering; however, the objective, continuous measures of these dependent variables is desirable to the extent that are valid and reliable. Glasses cannot be used with prescription lenses or sunglasses, thus limiting the potential pool of participants. The drivers must wear the glasses for the system to operate, and some individuals might find the glasses uncomfortable or feel they obstruct their view of the road.

---

5. **Seeing Machines** ([www.seeingmachines.com/guardian](http://www.seeingmachines.com/guardian))

This system is composed of two cameras. One camera captures the forward roadway and the other captures the driver. The forward-facing camera uses a wide-angle lens to capture footage in the event of an incident. The in-vehicle system uses infrared sensors to detect fatigue and distraction through proprietary face- and eye-tracking algorithms that measure eyelid closure, blink rate, and the driver’s head position. When these sensors detect microsleeps or driver inattention, the system alerts the driver with an audio tone and vibrating seat. The 24/7 SafeGuard Center provides around-the-clock fleet protection through live data analysis and human intervention. In case of a verified fatigue event, managers will be notified within 2 minutes.

**Effectiveness**

The system runs optimized real-time proprietary computer vision algorithms that measure the position and orientation of the head in three dimensions, as well as the extent of eyelid opening; however, there is not study that evaluated this algorithm against a validated measure of fatigue or drowsiness. A study of three long-haul trucking companies in South Africa showed a 93.2% reduction in fatigue events per hour in the intervention period (incidence rate ratio [IRR]: 0.068, 95% confidence interval [CI]: 0.059–0.078, \( p \leq 0.001 \)) versus the baseline period during which the trucks were installed with Guardian but no feedback was given to the driver. When fatigue-related events were compared on a per distance travelled basis, there was a 90.9% reduction in fatigue related events (IRR: 0.091, 95% CI: 0.080–0.105, \( p \leq 0.001 \)).

**Cost**

Can rent for $240/month. Unit can be purchased for $2,000 with a $30 monthly subscription fee.

---

45 Lenne, M., & Fitzharris, M. (2016). Real-time feedback reduces the incidence of fatigue events in heavy vehicle fleets. *Proceedings from 23rd ITS World Congress (ITS-AN-SP0293).*
Pros and Cons in Proposed Study

Guardian operates and maintains tracking integrity through a range of lighting from bright sunlight to nighttime. It also functions with most glasses, even sunglasses, as the system will then use head pose to determine fatigue level. Managers are notified and can act should a driver receive a fatigue alert. They can also use data from the system to analyze the entire fleet and find ways to mitigate fatigue. Moreover, should an event occur, the video from the forward-facing camera offers the company additional protection. Drivers may like the privacy feature that the camera only records if it senses a fatigue or distraction event. Raw outputs cannot be accessed; thus, the system relies on a proprietary algorithm based on eye and head movements (will only get information on how many times the criterion measure for fatigue was met or exceeded). Alerts can be silenced, which is needed in the proposed research; however, raw outputs are not available.

6. Optalert (http://www.optalert.com)

The Eagle Industrial and Eagle Portable are glasses that monitor eye and eyelid movements. The system calculates multiple variables from the neuromuscular function of muscles in the eyelids during their reflex-controlled movements with each blink. The glasses have infrared reflectance oculography that provides a continuous measurement of drowsiness using the validated Johns Drowsiness Scale (JDS). The JDS is reported to the wearer continuously in real time every minute as a value between 0.0 and 9.9. The constant feedback to the driver allows the driver to self-manage fatigue. Drivers are warned 15 to 20 minutes before they have microsleeps. An auditory alert sounds when a high level of fatigue is detected. Managers can access the drowsiness levels of all their drivers in real time, as well as opt to receive an email or text alert when a driver receives a “High Risk” warning. Data collected by the system can be analyzed and used to help fleets mitigate fatigue. There are two versions available: a portable system and an industrial system. The portable system is a plug-and-play version that uses a smartphone as the display for the JDS and wireless glasses. Both systems use over-the-air software updates.
**Effectiveness**

Validation of the JDS scores was conducted by assessing homeostatic and circadian change. Fourteen participants completed 30 hours of wakefulness while wearing the Eagle glasses and completing bi-hourly neurobehavioral tests, including the KSS and PVT. Researchers concluded that average JDS scores above the cautionary level were associated with delayed response times and subjective sleepiness when compared to average JDS scores below the cautionary level.\(^{46}\) Stephan et al.\(^{47}\) investigated lane deviations of alert and sleep-deprived drivers up to 30 min. after a driver scored a cautionary and critical JDS. Their study showed a significant increase in the proportion of time a vehicle left its lane during the 30 min. following a cautionary or a critical JDS level for periods that cautionary or critical JDS levels were not reached. In addition, under sleep-deprived conditions, the study showed a sensitivity range of 70.6% to 75.0% for 5 to 30 min. following a cautionary JDS and 45.8% to 56.3% for 5 to 30 min. following a critical JDS value. The study showed a specificity range of 65.4% to 71.4% for 5 to 30 min. following a cautionary JDS and a specificity range of 56.3% to 70.0% for 5 to 30 min. following a critical JDS value.

**Cost**

Cost is $1,000 per unit if 100 or more units are ordered and one’s own Android tablet is used as a display. There is also a monthly subscription fee of $20 to $50 depending on the level of service requested.

**Pros and Cons in Proposed Study**

The glasses have three different interchangeable shades of lenses so data collection can continue in all light conditions (bright sunlight, dim light, and complete darkness). The lenses can be adapted for prescriptions as well. However, if the light changes during a driving shift, drivers may try to change lenses while driving, wear the incorrect shade lenses for the environment, or take off the glasses altogether in order to see properly. Fleet managers have the option to receive alerts regarding drivers if they receive a “high risk” warning. The drivers must wear the glasses for the system to operate, and some individuals might find the glasses uncomfortable or feel they obstruct

---


their view of the road. The glasses come with or without a wire for portability. Software updates are automatically downloaded and installed. The Eagle Portable version is plug-and-play, so there is no downtime for fleets during installation. The system does not indicate if raw outputs can be accessed, and relies on a proprietary algorithm based on JDS scores (not PERCLOS). It is unclear if the alerts can be silenced, a requirement of the proposed research.