DRIVING IN A SIMULATOR VERSUS ON-ROAD: THE EFFECT OF INCREASED MENTAL EFFORT WHILE DRIVING ON REAL ROADS AND A DRIVING SIMULATOR

by

Jessica Anne Mueller

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Engineering

MONTANA STATE UNIVERSITY
Bozeman, Montana

April 2015

DEDICATION

For my family back East, and my Bozeman family.

ACKNOWLEDGEMENTS

In the seven years I've spent at MSU, I have been relentlessly supported by a huge number of people. Dr. Laura Stanley has mentored for me over seven years and two degrees, and I'll always be thankful for taking me on and giving me the support I needed over the years. Big thanks are due to my dissertation committee members Frank Marchak, Maria Velazquez, Robert Marley, and Mark Anderson for their guidance and support. Thanks also to Kezia Manlove for great statistical advice; John McIntosh and Will Hollender for saving the day when I really needed it.

Lenore Page and Kaysha Young constantly kept me motivated and gave me the encouragement I needed to pull this all together. Steve Albert and Western Transportation Institute were critical in supporting my projects over the years. Friends and family kept me focused and were my cheerleaders for this entire process and cannot be thanked enough: Josh Dale, Melissa Dale, Emily Pericich, Katherine McLachlin, the rest of Team Awesome, Jody and Don Mueller, Kathy Lemery-Chalfant, Eric Mueller and Audrey Albright—too many to name, but all have my eternal gratitude.

This project was supported by National Science Foundation Grant No. 1116378, Montana State University's Benjamin Fellowship, Montana Academy of Sciences' Student Research Grant, and the U.S. Department of Transportation and National Highway Transportation Safety Administration's Eisenhower Graduate Research Fellowship Program.

TABLE OF CONTENTS

TABLE OF CONTENTS – CONTINUED

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES – CONTINUED

LIST OF FIGURES – CONTINUED

Figure Page

ABSTRACT

The objective of this thesis is to study human response to increased workload while driving in a driving simulator compared to real world behavior. Driving simulators are a powerful research tool, providing nearly complete control over experimental conditions—an ideal environment to quantify and study human behavior. However, participants are known to behave differently in a driving simulator than in an actual real-world scenario. The same participants completed both on-road and virtual drives of the same degree of roadway complexity, with and without a secondary task conditions. Data were collected to describe how the participants' vehicle-handling, gaze performance and physiological reactions changed relative to increases in mental workload. Relationships between physiology and performance identified physiological, performance, and gaze-related metrics that can show significant effects of driving complexity, environment, and task. Additionally, this thesis explores the inadequacy of multinomial predictive models between the simulator and instrumented vehicle. Relative validity is established in the performance-physiology relationship for on- and off-road fixation frequencies, but few correlations between the simulator and instrumented vehicle are apparent as mental workload increases. These findings can be applied to the real world by providing specific variables that are adequate proxies to detect changes in driver mental workload in on-road driving situations; valuable for in-vehicle driver assistance system research. Overall, the simulator was a suitable proxy to detect differences in mental workload in driving task; and initial steps have been taken to establish validity, and to supplement on-road driving research in these high-demand driving scenarios.

OBJECTIVE

The objective of this set of studies is to validate human response to increased workload while driving in a driving simulator to real world behavior, to illustrate the capability of simulators as a suitable proxy for real world driving. Driving simulators are a powerful research tool, providing nearly complete control over experimental conditions—an ideal environment to quantify and study human behavior. However, participants are known to behave differently in a driving simulator than in an actual real-world scenario. By comparing driver response in an on-road real-world setting with driver response data measures obtained in a driving simulator, we will be able to isolate driver compensatory strategies associated with increased mental workload in both a simulator and real vehicle, to compare how drivers react in both environments. This research is unique by looking specifically at the physiological and performance relationships while mental workload is applied—a condition commonly found in driving simulator research, but not yet studied for validity. If this study can show a reliable relationship between real and simulated driving, it will serve as a major step toward the validation of driving simulators as a suitable proxy for real-world driving situations.

BACKGROUND

Driving Simulators

Driving is a common and universal task performed by humans daily; and also a task subject to a high degree of risk. Motor vehicle crashes are the leading cause of death for individuals aged 11 to 27; there were over five million police-reported crashes in 2011 alone, leading to the deaths of 32,367 people (NHTSA, 2013). While these numbers are alarming, advances in transportation safety have resulted in 2011 being an impressive year—2011 had the lowest fatality rate, at 1.10 fatalities per vehicle mile traveled (VMT), compared to a fatality rate of 1.50 in 2002 (NHTSA, 2013). As a highly complex task, driving involves a wide range of sensory, perceptual, cognitive, and motor functions (Allen, Rosenthal, & Cook, 2011). One of the popular technologies researchers have been using to investigate transportation safety research revolves around the use of driving simulators to study driver behavior.

Early driving simulators in the late 1960's used different methods of presenting environmental imagery and cues to the drivers. These older computing and graphics capabilities were not sufficient to support refresh rates or computational delays that would allow researchers to accurately measure human performance. Some of the original simulators consisted of computationally generated imagery despite these shortcomings. One simulator consisted of showing the driver film of a moving belt model whose belt speed was representative of the vehicle velocity, and camera angle and position were used to represent the driver's heading and lane position (Weir &Wojcik, 1971). Another type of simulator was film-based projection (35mm), which while it provided excellent resolution was associated with high production costs. In this approach, the speed of the driver was represented by the speed of the projector, and the driver heading was represented by the angle of the projector pan angle (Hutchinson, 1958).

With the evolution of digital processing capabilities and the wider availability of graphics components capable of supporting driving simulation, the development of driving simulators has largely moved to PC-based systems. The difficulties that were present in early simulators (low refresh rates, and computational delays) can still be found, and are generally associated with highly complex visual scenes that require additional computer resources to support. Information about specific simulator visual or motion delay was unable to obtain via published simulator research; but it is a factor that has been identified in aviation simulator that impairs performance and increase physiological workload (Flad, Nieuwenhuizen, Bulthoff & Chuang, 2014). To counter this issue, compensation techniques have been developed to try to predict the changing visuals to decrease computation load (Ricard, Cyrus, Cox, Templeton & Thompson, 1978; Hogema, 1997).

Advantages

Simulators provide a safe environment to observe human responses in unsafe experimental conditions. It would be unethical and potentially harmful to the driving participant as well as the general public to study a person's driving behaviors or patterns while excessively fatigued, experiencing effects of drugs or alcohol, executing difficult or dangerous evasive driving techniques, or while using new and untested technologies within a driving context (texting, GPS, other in-vehicle accessory systems). Driving simulators provide an environment where these types of studies can be conducted without putting the driver or the general public at risk.

Real on-road driving environments can be difficult to fully control, but conditions for both internal and external variables can be controlled in a driving simulator, providing complete experimental control to a driving simulation study. Aspects of environments which can be controlled include weather, road conditions (dry, slick, icy; potholes, hazards), pedestrian location or pace, ambient traffic density, wind, and time of day. By using a simulator, researchers can ensure identical experimental conditions for all participants; compared to the highly variable and often unpredictable presence of these elements encountered while driving in the real world.

In addition to complete experimental control, simulators also provide an opportunity for increased exposure to events which may not typically be encountered while driving on real roads. Scenarios can be modeled that include exposure to events including animal/vehicle encounters, erratic traffic behavior, sudden braking by forward vehicles, or vehicle crashes. By providing an avenue where driver response to these events can be measured during an experimental session (as opposed to waiting for these rare events to occur naturally, on road), simulators can reduce the time and cost required to conduct a study.

While very high-fidelity simulators may have a prohibitive initial financial cost for many research institutions, there are a great many different configurations of simulators which can be used at more affordable costs. Currently, it remains to be seen if published research studies can be reproduced across the wide variety of existing simulator configurations, but low-fidelity simulators are a widely available option within the budgets for smaller research institutions.

Disadvantages

Individuals not familiar with driving simulators have been known to experience some discomfort, which can impact the behaviors they exhibit in the vehicle, occasionally showing one or many of a group of symptoms known as “simulator sickness (SS) (Kennedy, Lane, Berbaum & Lilienthal, 1993).” While tools exist to help prescreen study participants for sensitivity to simulator sickness, it occurs in a higher proportion as drivers age, and women display an increased sensitivity to the symptoms (Classen, Bewernitz & Shechtman, 2011). Beyond the physical discomfort, the disadvantage to simulator sickness is reduced representation of several demographics (older drivers and women) due to onset of simulator sickness symptoms. While participant selection can be impacted here, it is not largely known the characteristics of the general public that do not experience simulator sickness symptoms, or if the behavior of individuals insensitive to simulator sickness can be reliably used to make inference to the general population. Simulator characteristics can also contribute to higher simulator sickness rates, specifically refresh rates, scenario design, duration of experimental session, simulator configuration, and simulator calibration (Classen, Bewernitz & Shechtman, 2011). A survey was developed intended to be used to quantify the effect and focus of simulator sickness in participants (Kennedy, Lane, Berbaum & Lilienthal, 1993), which is currently in use as a screening tool to pull subjects from active participation if the effects of simulator sickness become too great.

The current level of technology used in high-fidelity simulator facilities can be so expensive that the experimental conditions cannot be easily reproduced by alternate facilities (Hancock & Sheridan, 2011). Simpler systems may consist of a desktop computer, with a monitor display and video game-style pedals and steering wheel; compared to a state-of-the-art facility like Iowa's National Advanced Driving Simulator (NADS); see Figure 1 below.

Figure 1. Desktop Simulator, and National Advanced Driving Simulator Facility

Currently, no standard system exists to classify simulator facilities based on different variable characterizing their complexity of construction, motion, and visualization.

At present levels of technology capability, it is impossible to completely and perfectly replicate conditions found in the real world. The natural limitations in computational speed and graphics capability currently result in a tradeoff between performance and complex visual scenes. A human vision system's ability to perceive images is limited by the resolution capabilities of the eye, while the resolution detail in a driving simulator is typically limited by the hardware it uses (projector or television screen; typically). Resolution is generally described as the degrees visual angle for a single pixel—or unit of differentiation. The resolution of the human eye has approximately a 0.016 degrees visual angle; but a projector screen has a much lower resolution depending on the size of the display and built-in resolution. If a driving simulator display was projected to a visual angle of 60 degrees by 40 degrees, using a common resolution of 1024 by 768, a single pixel would be 0.058 by 0.051 degrees visual angle (Andersen, 2011). As a wide variety of driving simulators are in use in research contexts, the degrees visual angle for each individual simulator is an important characteristic that can be compared between different facilities.

By conducting driving simulator experiments in a lab setting, drivers may display non-natural behaviors or responses due to their knowledge that they are being observed. For example, crashes in a driving simulator have few consequences when compared to crashes in a real driving environment, which may lead to a driver being less careful since he knows his actions have few real-world consequences.

Experiments using a driving simulator often use a methodology that requires the drivers to participate in multiple trials, perhaps with different experimental treatment combinations for each replication. Repeated exposure in a simulator can have an effect on the driver's response to subsequent exposures. For example, say a driver in his first treatment had to react to a scenario where he had to maneuver to avoid a suddenly braking forward vehicle—he would then be ‘primed’ for the future treatments to anticipate that same event; which could affect his response by braking faster or driving more carefully. One study found that subjecting the driver to surprise events beyond the initial event reduced the drivers’ reaction speed for the non-initial event by about a half-second (Olson & Sivak, 1986), so learning effects are present in simulator environments.

Transportation Applications

One application for simulator research is for driver training. Simulators provide a highly-controlled, safe, and relatively cost-effective method to train new drivers.
One advantage over the real world is with simulator’s ability to trigger events which are relatively rare in natural driving conditions, such as crashes or animal strikes. As a training tool, simulators are relatively new, and the relationships between training received in a simulated environment and its impact on driving behavior or performance is still being investigated.

Simulators have been used as part of rehabilitation therapy, for individuals suffering from some form of physical impairment. Devos, et al. (2010) used driving simulators as part of training for people who had experienced a stroke in the past six to nine weeks. Participants were administered 15 hours of driving therapy, and were evaluated before training, after training, six months after training, and 60 months after training. The simulator training was found to speed up the restoration of driving skills after a stroke, with effects observed at 6-months post-stroke (no effect at the 60-month evaluation). [Also see Carroz, Comte, Nicolo, Deriaz & Vuadens, 2008]While not an exhaustive list, driving simulators have been used to successfully rehabilitate individuals with traumatic brain injuries (Cox, Singh & Cox, 2010; Mazer et al., 2005), post-traumatic stress disorder due to motor vehicle crashes (Beck, Palyo, Winer, Schwagler & Ang, 2007), post-stroke patients (Akinwuntan et al;, 2005; Devos et al., 2009), and driving-related phobias (Tomasevic, Regan, Duncan & Desland, 2000).

Simulators can be used to evaluate driving performance. Several studies used simulators to differentiate between safe or unsafe drivers (Lee, Lee, Cameron, & Li-Tsang, 2003; Lew et al., 2005; Patomella, Tham & Kottorp, 2006), and predicting likelihood of future crash involvement (Lee & Lee, 2005). One study noted the advantage of simulators in driving assessments that with the correct data collection, driver assessments can be conducted while the participant is alone in the vehicle, potentially improving the ecological validity of the assessment (Bedard, Dahlquist, Pakkari, Riendeau & Weaver, 2010).

Driving simulators can also play a role in the evaluation and testing of novel or untested roadway geometry designs (Granda, Davis, Inman, & Molino, 2011). Roadway geometry generally refers to different physical characteristics of a roadway: lane design, horizontal and vertical curvature, and superelevation are all generally considered. The Federal Highway Safety Administration (FHWA) simulator has conducted several studies that were exploring how people behave in new intersection designs. In one study, the proposed roadway (a “diverging diamond interchange;” see Figure 2) would temporarily shift drivers lane position to the left side of the road—the designers were concerned that the drivers might inadvertently bear to the right lane out of habit and cause crashes. After evaluating driver performance in the simulated interchange, researchers found that there drivers did not perform the predicted wrong-way errors (FHWA, 2007). Another study simulated a real-world intersection that had high crash-rates, to try to identify features that could predict crashes. The researchers found increased crash rates were associated with corners with stop lines closer to intersections (increased deceleration rates), and right-turns on red where the driver did not come to a complete stop (Yan, Abdel-Aty, Radwan, Wang & Chilakapati, 2008).

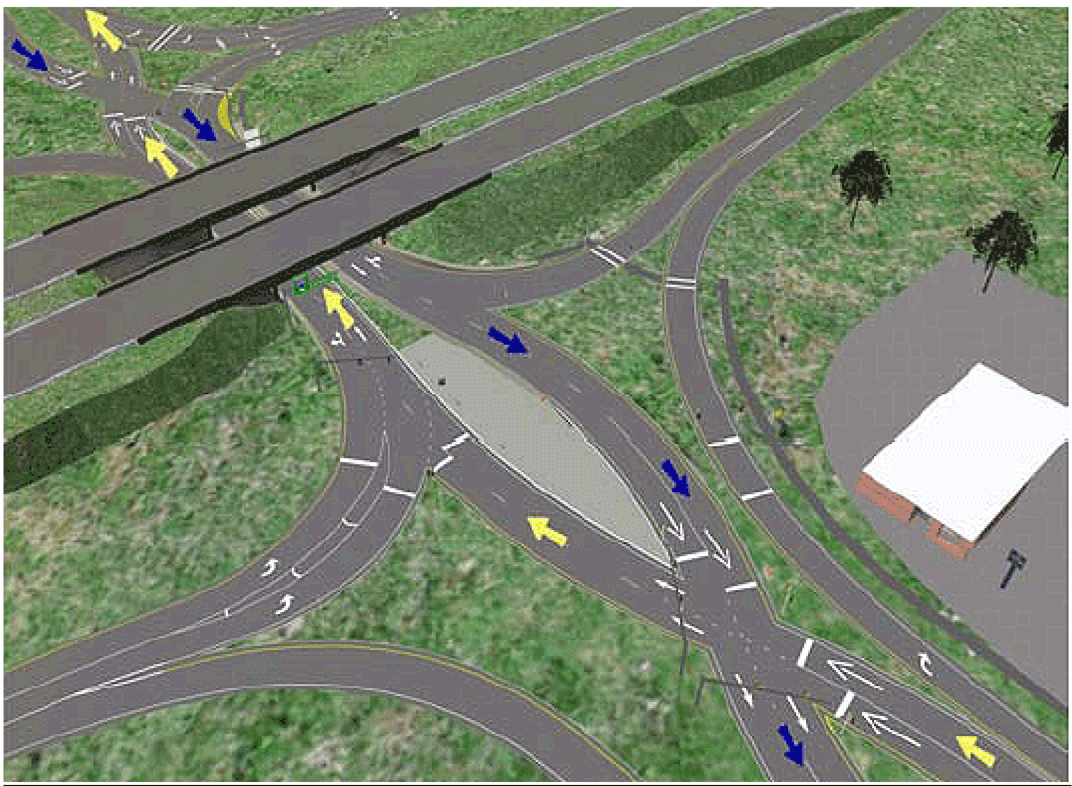

Figure 2. Overhead view of a diverging diamond interchange, (FHWA, 2007)

One useful application of simulators involves studying human-computer interface (HCI) research, specifically looking at how drivers interact with different computer devices while performing driving tasks. As in-vehicle electronic devices and interfaces are becoming more common in vehicles, the science behind how drivers multi-task between those devices and the main task of driving is coming under more scrutiny. In-vehicle electronic systems typically fall under three different types of systems: information-based (navigation, vision enhancement, hazard warnings), control-based systems (adaptive cruise control, lane keeping, self-parking, and collision avoidance), and those that do not support the main driving task (cell phones, music, and entertainment) (Burnett, 2011). These types of systems can lead to two main scenarios affecting the driver: overload can be caused by excessive information affecting the driver's workload leading to distraction, or underload which is where control-based systems automate enough of the driving task that the driver can become uninvolved and eventually lead to reduced awareness, negative driving behaviors, and loss of skill (Burnett, 2010). Wang et al. (2010) studied driver interaction with a surrogate system where the driver needed to input an address in both a mid-fidelity driving simulator and in an instrumented vehicle on roads in the real world; they found near-absolute validity in many different performance measures collected during task performance.

Other Applications (Non-Transportation)

Simulation has been adopted as a training tool in many domains outside of transportation. In contrast to the transportation-based methods, most of these applications of simulation have not been thoroughly validated. Many focus on transfer of training, or on using the simulations as a rehabilitative tool.

Virtual representations have been used as a cognitive behavioral therapy tool (Poeschl and Doering, 2012; Slater, Pertaub & Steed, 1999; Pertaub, Slater & Barker,

2002) to expose patients with social phobias to a virtual audience. These audiences were enabled with different attributes so that the researchers could manipulate characteristics such as negative or positive response to the speaker, audience members chatting with one another instead of listening, or even an audience member that could get up and cross the room mid-speech. An evaluation of this type of audience showed that speakers appropriately characterized the audience reaction and level of interest (Slater et al., 1999; Pertaub et al., 2002).

Virtual training has also played a role in the rehabilitation world, and with coping with different disabilities. The transfer of spatial knowledge has been confirmed in disabled children—the children interacted with a virtual multi-story building that they could not have otherwise interacted with. After interaction with the virtual world, the children were able to identify spatial information better than unexposed control group of children (Wilson, Foreman & Tlauka, 1996). Another study tried to use virtual environments to train inexperienced power wheelchair users, who ended up reporting that the physical tasks were much more difficult in virtual reality than in real life (Harrison, Derwent, Enticknap, Rose & Attree, 2002), indicating poor physical validity at the time the experiment was conducted. Virtual environments have also been explored to try to assist autistic children in collaborative virtual environments where multiple users can interact with one another with virtual representations of themselves (Moore, Cheng, McGrath & Powell, 2005). Moore et al. (2005) used these avatars to help train the children to recognize basic emotional expressions.

Different aspects of combat training have used simulations to train and evaluate humans. One study looked at situational awareness in an urban terrain simulation, using infantry squads to evaluate the virtual world (Kaber et al., 2013). The simulation led to positive skill development, and helped to assess squad leaders as they lead their teams. Other types of combat virtual training include a program developed for combatants to use that is aimed at training a range of emotional coping strategies that are intended to benefit stress resilience (Rizzo et al., 2012), and also aimed at exposure therapy for treating military personnel with posttraumatic stress disorder (PTSD) symptoms (Rizzo et al., 2010), finding that 80% of participants (n=20) improved PTSD symptoms following the virtual treatment. As this is a more recent area of simulation training, there are no long-term efficacy or validation studies yet concerning the effects of treatment or training over time.

The medical domain is increasing their support of virtual simulations in order to train staff in different procedures. Research has looked at a host of different surgery procedures, including post-traumatic craniomaxillofacial reconstruction (Tepper et al., 2011), coupled tissue bleeding (Yang et al., 2013), laproscopic surgery (Wilson et al., 2010)—the list is extensive. Virtual surgery training is used to train surgeons, as well as to identify the behavioral patterns that can separate more skilled practitioners from novices. Some advantages of medical simulations are similar to those found in transportation: this is a safe environment, where a patient cannot accidently be harmed, and also some of these procedures are not particularly common. Simulation provides the chance for medical workers to interact with uncommon events at an artificially inflated rate—by studying these types of events under high mental workload, driving simulation study findings can be applied toward the highly-demanding training scenarios commonly found in medical simulation.

Typical Simulator Configurations

Three main classifications exist for driving simulator configurations: low-, mid-, and high-fidelity facilities. Low-fidelity simulators are typically cheaper to build, consisting of a desktop computer monitor with a steering wheel and pedal configuration typically used in some video game setups. These simulators have little driver feedback. Mid-fidelity simulators generally consist of a projected or wide-angle display composed of many monitors that blend together a composite image representing the driver's view of the simulated world. The driver seat is typically surrounded by some sort of a vehicle shell, lending a sense of physical validity of the vehicle interior modeled in the simulator. Typically there is some sort of feedback on the steering wheel and pedal arrangement to provide the sense of pedal or wheel pushback encountered during actual driving. High-fidelity simulators provide a near-360 degree field of view, consisting of a blended projected image of the vehicle's simulated surroundings. In addition to steering wheel and pedal kinesthetic cues, high-fidelity simulators are mounted on extensive motion bases that provide the driver with motion cues representative of those felt while driving on real roads (Jamson, 2011).

With increasing fidelity, the cost to build the different types of simulators can vary wildly. The prevailing theory is that research involving higher-fidelity simulators yields driver behavior that more closely represents those found in real world driving applications. Santos, Merat, Mouta, Brookhuis, and de Waard (2005) conducted a study comparing how drivers performed when engaged in the same tasks (cognitive, visual, and a dual-task) in low-, mid-, and high-fidelity simulators. The authors found that the same relative results were found—a reduction in speed— when engaged in the tasks. However, as simulator fidelity increased, the effect size findings became more sensitive. A different study by Jamson and Mouta (2004) compared driver performance in a low- and a mid-fidelity simulator while engaging in the same primary and secondary tasks. The mid-level simulator showed more sensitivity to the driver speed reduction. In the mid-fidelity system, it was easier to detect differences in driver behavior than in the low-fidelity system. Allen and Tarr (2005) compared four increasing levels of driving simulators (low-, mid-, and high-fidelity) and found that as simulator fidelity increased, so did driver performance.

Typical Simulator Driving Scenarios – Literature

The goal of developing a specific driving simulator scenario is to generate a driving situation that enables the researchers to test specific hypotheses of interest. Driving simulator study objectives are quite varied, and the scenario designs reflect those specific objectives. For example, a study looking at driver behavior in a construction zone will by its nature need to incorporate construction zone elements in order to present a realistic scenario environment. Any dynamic events (e.g., a lead vehicle, cars or pedestrians turning in front of the participant, or pedestrians walking alongside the road) can be added as necessary for specific testing purposes.

One thing that should be considered when designing driving scenarios is the likelihood of a scenario to cause simulator sickness in drivers. The addition of extra roadside features such as trees or buildings can help deliver cues to the driver about motion and increase the realism of the scenario; however, these visual objects also help trigger cue conflict for the drivers leading to simulator sickness (Stoner, Fisher & Mollenhauer, 2011). A completely featureless scenario has no optic flow and will not lead to simulator sickness, but will consequently deliver no cues to the driver about motion, location, roadway, or speed—a combination that is not useful to researchers. The general recommendations for scenario design to minimize simulator sickness include minimizing rapid changes in direction, and the number of sharp decelerations.

Human Performance and Stress

Hockey (1997) expanded Kahneman's (1973) model of what happens when human task performance is subjected to high workload or stress. Hockey's model incorporates a dual-level system, shown in Figure 3. From this figure, Loop A represents automatic processes that occur with minimal thought and effort. In this loop, task performance is monitored. If performance is found to be off-target, then resources are diverted toward accomplishing the given task. When additional workload or stress is introduced into the process, the effort monitor moves the process from Loop A up to Loop B. This can be thought of as a dual-level system. Loop A represents how human task performance is monitored when low demand is applied; Loop B represents how the process works during periods of high workload or stress.

Throughout Hockey's model, the level of task performance is compared to target performance goals—and when additional cognitive resources are needed in order to bring task performance back to target, different compensatory strategies are adopted. For Loop A, this is referred to as "Active Coping" and is characterized by high behavioral stability, low effort, and increased mental energy. Using Frankenhauser's (1986) definitions of challenging situations, this type of coping can be called "effort without distress."

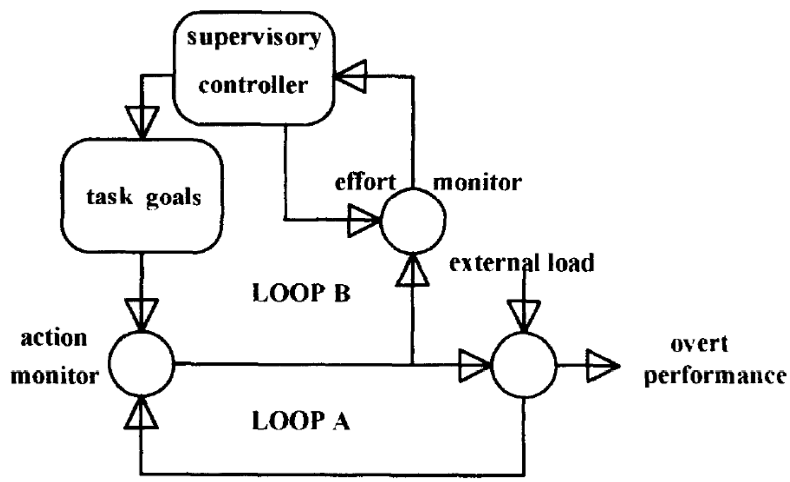

Figure 3. Hockey's (1997) model of performance and effort

When excessive stress or workloads are applied to the model, there are three typical compensatory strategies: strain coping, passive coping, and complete disengagement. In strain coping, the process has increased the maximum available resource budget in order to meet demand. Target performance will still be observed, but there will be an "energetical" cost for increasing the resources. This strategy is associated with increased sympathetic dominance responses, increased catecholamine, and increased cortisol. Frankenhauser would call this strategy "effort with distress." In passive coping, instead of increasing resources to meet target goals, the goal performance target is reduced so that it can be reached without additional resources. This is associated with increased adrenocortical activity, and can be classified as “distress without effort.” Complete disengagement involves the whole abandonment of task goals, and is typically seen in activities that can be naturally started or stopped—Hockey’s example of an activity where complete disengagement is common involves the stop-and-go nature of some academic writing. Research targeting driving and distraction typically focuses on Wickens’ multiple resource theory—the reason for the selection of Hockey’s model over Wickens’ is due to how Hockey proposes a system focusing on how driver physiology and performance are related. Wickens’ theory specifically targets different modalities of attention individually. Because the driving task encompasses many different attentional modes, Hockey’s model was more appropriate to examine the objectives in this dissertation. Hockey’s model could go a long way toward explaining some discrepancies in findings in performance or physiology-based simulator findings, but has not yet been specifically studied in the driving simulator.

Validating Driving Simulators

Because driving is such a familiar task, there is a lower limit on the acceptable level of “fidelity and accuracy (p11-1)” required in a simulator configuration so that the simulator yields useful information (Schwarz, 2011). Blaauw (1982) began the movement studying validation of simulators. He describes the two main definitions of validity as referring to 1) the equivalence of the driver behaviors between the simulated and real environments; and 2) the physical and dynamic characteristics of the simulated and real environments. In short, driving simulator validity is the ability of the driving simulator to elicit the same human behaviors and responses as would be encountered in a real world environment.

Western Transportation Institute's high fidelity simulator has already been validated to an identical vehicle in on-road conditions (Durkee, 2010) in the context of brake force and steering force required for vehicle handling at different speeds; the task of remaining validation lies with studying driver behaviors between a simulated and real environment. Durkee was involved in a tuning effort in 2009 which compared the simulator vehicle dynamics and noise exposure to conditions measured in a representative vehicle under identical conditions. Inputs that were measured and validated during this comparison include brake force required at different speeds, steering column resistance, and acceleration forces; engine and road noises were also compared at different speeds. Following data collection of these parameters on-road, the simulator was tuned so that it achieved the same measures from those variables during use.

Blaauw goes on to describe four main methods used to study behavioral equivalence between a driver in a simulator and on-road: 1) Comparing the two systems given identical experimental conditions by measuring driver behavior or system performance; 2) comparing the two systems using human response to different physiological variables; 3) comparing the two systems through evaluation of self-reported subjective criteria by participants; or 4) establishing the ability of experiences in one environment to "transfer" to the other (Blaauw, 1975).

In order to measure simulator validity, performance differences in identical experimental conditions must be compared in both a real and simulated environment. Those performance differences should be measured in each environment. If the differences have similar direction and magnitude, then they are considered to show "relative" validity; if the differences have equal numerical values, they are considered to show "absolute" validity (Blaauw, 1975). Both types of validity are valuable for experimental inference; while absolute validity marks results much more representative of real behaviors, relative validity in several different variables has already been established in different simulators and studies.

Human Physiological Response

Using human physiological data to evaluate a person's physical state was first done as early as 1939, with a rudimentary polygraph test in order to evaluate "tension" in flight (Williams, Macmillan & Jenkins, 1946). As technology improved, researchers were able to use different methods to evaluate human stress or strain, in studies conducted in the field, simulators, and laboratories. Klimmer and Rutenfranz (1983) further divided this strain into "mental" or "emotional" categories. Boucsein and Backs (2000) provide a summary from past literature of psychophysiological parameters that measure these categories of strain, adding "physical" strain to the list.

The simulator used in this study has previously been physically validated against an identical real-world car (Durkee, 2010), so the physical strain associated with operating the simulator is assumed to not have an impact, and therefore will not be examined. The effects of "Emotional" strain on driving performance or behaviors are not known at this time, and are not included in the scope of this study. Additional preliminary work exploring the physiological validation of a driving simulator was done by Mueller et al. (2013), comparing participant heart and breathing rate at a specific driving scenario. In Mueller's (2013) hazardous scenario, the study finds an initial implied validity between driving on the real roads and a mid-fidelity driving simulator for breathing rate and heart rate. However, the comparison between environments should be extended to examine how drivers behave specifically in high-workload scenarios in both environments; not just in a single instance.

The strain category which is most relevant to driving operations is mental strain. A review of studies conducted since 1985 shows the following responses were strongly associated with increased mental strain (Boucsein & Backs, 2000): decreased EEG alpha activity (8-12 Hz), increased EEG theta activity (4-7 Hz), increased P3 amplitude, decreased 0.1Hz component (cardiorespiratory), decreased respiratory sinus arrhythmia, increased eyeblink rate, and increased epinephrine (Table 1.1, p9).

The physiological parameters that will be used for these experiments were selected in order to measure the elevated sympathetic response associated with strain coping in the higher-order mental processing in Hockey's mental model (1997). Specific justification for selection is expanded in the Methods section.

Human Behavioral Response

Mullen et al. (2011) studied how to validate driver behaviors from a simulator to on-road behaviors, and have compiled a comprehensive list of studies that have established either relative or absolute validity in different driving behaviors. Absolute validity has been established for speed in non-demanding road configurations (Bella, 2008), speed (Blaauw, 1982; Reed & Green, 1999), brake reaction time (McGehe et al., 2000), physiological responses during turns at intersections (Slick et al., 2006), and perceived risk of hazards on road (Yan et al, 2008).

Relative validity has been established in different behaviors as well: Perceived threat from unexpected events (Bedard, 2008), On-road demerit points (Bedart et al., 2010), speed (Bella, 2005; Bella 2008; Bittner et al., 2002; Charlton et al., 2008; Klee et al, 1999; Shinar & Ronen, 2007; Tornros, 1998), lateral displacement (Blaauw, 1982; Phillip et al., 2005; Tornros, 1998; Wade & Hammond, 1998); gaze direction (Charlton et al., 2008; fisher et al., 2007), speed countermeasures (Godley et al., 2002; Reimersma et al., 1990), steering angle (Hakamies-Blomqvist et al., 2001), braking onset (Hoffman et al, 2002), overall performance (Lee et al., 2003; Lee et al., 2007; Lew et al., 2005); reaction time (Toxopeus, 2007) and physiological responses (Slick et al., 2006).

Behavioral performance measures will be used to assess whether or not target goal performance is being met, to establish the difference between strain coping and passive coping with regards to the cost associated with the higher applied mental demand. The specific performance measures selected for this study are expanded on in the Methods section.

Human Subjective Response

There are several different methods currently in place to measure subjective driver workload. The NASA-Task Load Index (NASA-TLX) is a measure of several different subscales of workload (mental demand, physical demand, temporal demand, performance, effort, and frustration). These subscales can be measured on their own (“Raw TLX”) or combined with a pairwise comparison that measures the relative importance of each subscale to the participants (Hart & Staveland, 1988). The Rating Scale Mental Effort survey asks participants to indicate the amount of general effort that a task required; the scale is divided with several different anchors indicating the specific amount of effort (from “Absolutely no effort” to “Extreme Effort”) (Young & Stanton, 2005; Zijlstra, 1993). A third method was initially developed for airplane pilots; the Situation Awareness Global Assessment Technique (SAGAT) has been used in a transportation context to ask drivers about immediate information needs (Endsley, 2000; Jones &Kaber, 2005). The Driving Activity Load Index (DALI) was developed as a revised NASA-TLX method to be used specifically for driver mental workload. Instead of the NASA-TLX subscales, the DALI subscales are effort of attention, visual demand, auditory demand, temporal demand, interference, and situational stress (Pauzie, A., 2008).

Subjective response measures will be used to assess the amount of distress or effort experienced while driving in both environments, related to the coping strategies used when experiencing higher applied mental demand. Specific responses selected for use in this thesis are detailed more thoroughly in the Methods section.

Simulator validation has been a fairly well-explored topic in transportation safety, as simulators are becoming more relied-upon as a research tool. While studying driver workload has been looked at extensively in the literature, the study of driver compensatory behaviors *while subjected to heavy workload* has not been conducted. This research gap will serve an important role, as the ideas behind compensatory behaviors are directly applicable to simulation applications outside of transportation. The basis of Hockey's model of processing under heavy workload will be useful in many domains, and has not yet been studied in any.

Task Complexity

Real-World Complexity

Verwey (2000) conducted several studies wherein participants drove through different scenarios on real roads, in a set route. The scenarios he selected were general road situations found in everyday driving: standing still at a traffic light, straight on inner or outer city roads, driving on curves, roundabouts, or motorway; and turning at uncontrolled intersections. Verwey evaluated the scenarios that the drivers passed through on their route, and the performance of the drivers at both visual and mental cognitive loading tasks in each of these routes. One method to evaluate "real world" driving scenario complexity could be to examine performance measures—road situations that lead to poorer performance on the secondary tasks could be classified as more complex than those where secondary task performance was not impacted.

A German study specifically examined drivers as they interacted with a route-guidance system in a real-world study (Jahn, Oehme, Krems & Gelau, 2005). Jahn et al. selected their on-road segments based on the combinations of high or low levels of demands on information processing and demands on vehicle handling. The specific two classifications that Jahn et al. used were “high demands on information processing and high demands on vehicle handling” or “low demands on information processing and low demands on vehicle handling.”

Fastenmeier and Gstalter (2007) use a complicated system to evaluate specific driving situations for use in driver task analysis. Every potential driving situation was defined by its specific highway type and road design, road layout (horizontal or vertical curvature, type of junction, and junction control), and traffic low information. Fastenmeier and Gstalter then used this information, along with a matrix describing requirements for perception, expectation, judgment, memory, decision, and driver action, combined it with information about typical driver error, and compressed all of this information into a general rating of complexity and risk. Their algorithm was used to arrange many different specific driving situations based on that complexity and risk; an example from their paper is shown in Figure 4 (Fastenmeier & Gstalter, 2007). Specific information about the calculations of these complexity and risk levels was not available.

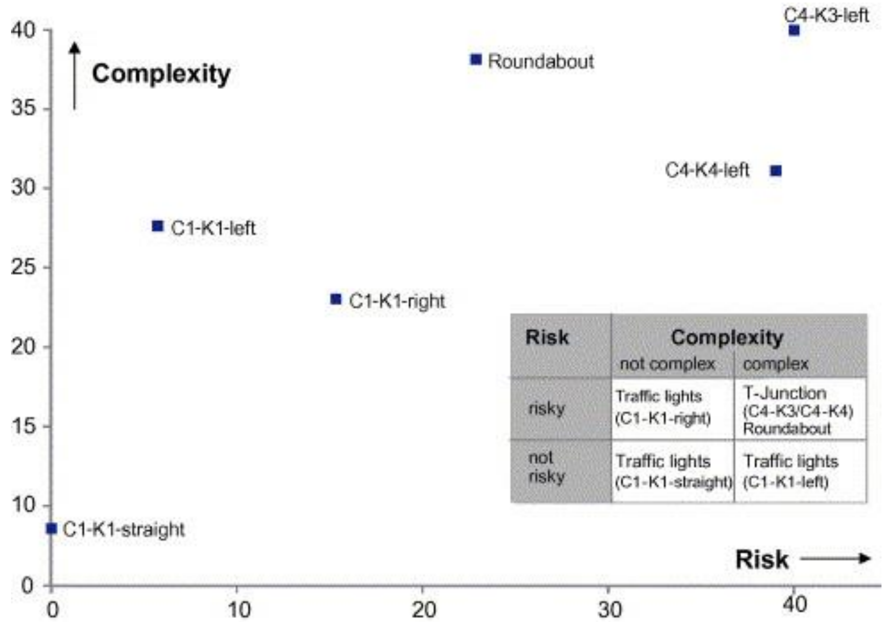

Figure 4. Example of Complexity and Risk from Gastenmeier & Gstalter (2007), p972

Because researchers will only very rarely be able to design real-world driving situations, we will seldom get the chance to capture specific classifications of scenario complexity in the real world. Studies generally select the segment(s) of real road that will be included in studies based on different classification schemes. These classification methods include measuring driver performance in many different driving situations (Verwey, 2000), using demands on information processing and vehicle handling to summarize scenario complexity (Jahn et al., 2005), and compiling a large amount of descriptive data about the roadway and driver task selection characteristics into a complexity value (Fastenmeier & Gstalter, 2007).

Simulated Scenario Complexity

Driving simulator scenarios are generally modeled to represent whatever physical road characteristics are needed for a particular study. Like in the real world, driving simulator scenarios can vary widely as far as the complexity of the situation they are modeled to represent. Studies that have examined cognitive workload in a simulator have used straight roads as a “low” complexity environment, increasing to intersections (Cantin et al., 2009), and can involve elements such as ambient traffic or simulated pedestrians to further increase scenario “complexity”.

There does not seem to be a clear method to define scenario complexity. One method uses “optic flow,” which is a measure of the number of visual elements in the scenario and the rate at which they pass. In this way, a scenario that was populated by 1000 different trees of various sizes or shapes would be considered more complex than a scenario that had fewer or no roadside features. In this way, optic flow is typically represented as ordinal factor levels, not at discrete levels with specific numeric boundaries. Mourant et al. (2007) studied how different scenario configurations affect driver behavior, finding that higher optic flow in scenarios resulted in drivers achieving slightly more accurate target velocities. However, the scenarios were simple, flat, straight roadways and so there was little variation other than the increased optic flow (trees were the only modifier of optic flow).

Cantin et al (2009) studied different driver reaction times as they maneuvered through scenarios of different complexities. Cantin’s study showed that both young and older drivers had increased reaction times as road or task complexity increased. The tasks they used were (from least to most complex) a baseline, straight road, intersection, and while overtaking. The drivers also showed behavioral changes, reducing their speed as they maneuvered through intersections consisting of either “high” or “low” mental workload, and also completed a higher number of braking events at “high” workload intersections when compared to “low.” The results that Cantin et al. found show that there are behavioral effects in simulated driving due to the workload supplied by the external environment. There was no apparent objective method for classifying the complexity of the driving scenarios, only an ordinal increasing between conditions (driving at a constant speed on a straight road, approaching and stopping at an intersection, and overtaking a slower-moving vehicle).

One study had drivers navigate through scenarios of varying complexity while completing different in-vehicle tasks while driving (Horberry et al., 2006). The “complex” scenario had over twelve times as many buildings, oncoming vehicles, and roadside objects compared to the simple conditions. The units they used for complexity were the number of objects (billboards, advertisements, buildings, or objects) per kilometer of road. Because this study examined only driving on straight highway roads, the researchers were not able to make inference on the effects of complexity as applied to the need for divided attention due to engaging in different driving maneuvers (approaching an intersection, interacting with other traffic). No specific justification was provided for the rates of objects (advertisements, roadside effects) between simple (28 objects/km) and complex (484 objects/km).

Bella (2008) conducted a study that examined drives on both real and simulated two-lane rural roads. From the real-world segment of road, the researchers isolated 11 separate points on the road where speed data was collected; these points were also measured in the simulator. The specific complexity was not discussed prior to the analysis segment of Bella's paper; the author post-hoc classified the 11 points as either "demanding" or "non-demanding" based on how different the driving speeds were in the simulator trials compared to the on-road data. In this study, the authors weren't interested in seeing how complexity affected vehicle speed; the authors only present the demanding/non-demanding element of the scenarios as a possible explanation following the analysis.

Another paper where the complexity angle was not directly approached was from Bittner, Simsek, Levinson, and Campbell (2002). In their findings, they classified that their drivers showed higher speed in the simulator for "least difficult curves," and lower speeds in the simulator for "most difficult curves." Here, difficulty may be interpreted as an approximation of scenario complexity—again reported not with objective complexity classifications, but the more ordinal "more difficult" or "least difficult" classification scheme.

Bella (2005) also describes some complex maneuvers based on the estimated difficulty in a paper comparing real world to simulated roads to evaluate work zone design: "from the beginning of the advance warning area at Site 2, the driver always has to perform more difficult maneuvers." In this study, the maneuvers and road sections were not selected due to their complexity levels; instead, scenario complexity was assessed following the scenario selection. Bella found that the fidelity of speeds (simulator compared to the real world) decrease as the maneuver complexity increases, and goes on to suggest that this may be due to the lack of inertial force on the driver since the experiment was conducted using a fixed-base simulator.

Previous work has indicated that there was a relationship between scenario complexity and changes in self-reported workload (Mueller, Martin, Gallagher & Stanley, 2014). Mueller et al. (2013) evaluated drivers on two different simulators while having participants navigate through three driving maneuvers of different levels of complexity: 1) a straight two-lane road without traffic control, 2) a left-hand turn with right-of way, and 3) an unprotected left-hand turn. All levels were repeated with and without an applied secondary task. NASA-TLX results showed a significant increase in mental effort both for increasing scenario complexity, and also when the secondary task was administered. The fidelity of the simulator did not have an effect on self-reported mental workload. This supports Verwey's (2000) study looking at similar behaviors in the real world—indicating that there may be a relative link between the two behaviors.

To summarize, there is no real basis for objectively determining a driving simulator scenario's level of complexity. Different approximations of complexity have been suggested, ranging from an increasing number of roadside objectives (Horberry et al., 2006), a subjective assessment of the difficulty of maneuver as a binary division of "low" or "high" difficulty or complexity (Bittner, Simsek, Levinson & Campbell, 2002; Bella 2005; Bella, 2008), or a general ordinal ranking of difficulty from low to high

(Cantin et al., 2009). Some of these studies are using complexity as an independent variable; others use it as a descriptive method of explaining results.

Applied Workload

Changes in mental workload can be accomplished by applying secondary tasks, on top of a main task. If there is sufficient mental capacity for both tasks, both tasks will be completed as expected. However, as task complexity increases, performance on the secondary task will decline—in this way, secondary task performance can be used to evaluate overall workload (Bridger, 2003). One example of this can be found in Verwey's (2000) study, where drivers on real roads participated in a peripheral detection task while driving. Secondary task performance was then assessed, and the different driving maneuvers that Verwey's participants completed could then be assessed for complexity.

The addition of an n-back secondary task will be used as an independent measure designed to propel drivers from the unconscious effort/performance loop (A) into one requiring additional effort and compensatory strategies to successfully complete the drives. There are a variety of different secondary tasks that are commonly used in different types of research studies; the n-back task was specifically selected because it does not involve a visual or tactile component. Several of the dependent variables in this study rely on driver gaze patterns, so secondary tasks that rely on peripheral vision or object detection would have affected those specific outcomes. The n-back task only uses driver cognitive processing via auditory recitation, which makes it ideal for this study. Mehler, Reimer, Coughlin, and Dusek (2009) used a similar task in their study looking at physiological arousal in young drivers in a simulator. Mehler et al. used multiple conditions of the secondary task, using single digit numbers with an interval between numbers of 2.25 seconds. Through the increasing difficulty, participants showed a “flattening” effect for heart rate, skin conductance, and respiration rate where the physiological variables showed no real increase beyond the 1-back stage, which is why the 1-back was the selected level of n-back for this study.

This dissertation used these independent variables to force participants to adjust their compensatory strategies—the objective of interest here is to find out whether or not the compensatory strategy will be the same in the simulator and the real world, an aspect of driving simulation which has not yet been explored.

Compensatory strategies for heavy mental workload have not been thoroughly studied comparing virtual and real driving. The effects of mental workload while driving are evident in several different aspects of visual gaze behavior, as assessed in past literature. Hockey’s (1997) model focuses on goal performance being the compensatory behavior in the presence of high workload. Typical eye tracking measures that are correlated with increased mental effort include blink amplitude, blink duration, blink rate, gaze dispersion (similarly, “range”), and fixation duration and frequency (Holmqvist et al., 2011). Much research has targeted these as dependent variables in driving studies, with simultaneous applied workload or secondary tasks—the aim of this study is to use these variables as indicators of the drivers’ target goal performance.

The most common variables analyzed while driving involves gaze dispersion (Recarte & Nunes, 2000, 2003; Reimer, 2009; Tsai, Viirre, Strychacz, Chase, & Jung, 2007), along with fixation characteristics toward specific areas in the vehicle and forward scene (Borowsky et al., 2014; Donmez, Boyle, & Lee, 2007; Fu et al., 2011; Harbluk, Noy, Trbovich, & Eizenman, 2007; Nabatilan, Aghazadeh, Nimbarte, Harvey, & Chowdhury, 2012; Recarte & Nunes, 2000, 2003; Sodhi et al., 2002; Tsai et al., 2007; Wang et al., 2010).

It is important to note that studies have not found consistent results when assessing physiological gaze variables alongside several other gaze parameters—if all of these behaviors are in response to increased levels of mental workload, we would expect to see the relevant trends correlated with higher effort. One on-road study that shows this trend found several significant changes in a spatial gaze variable that had no corresponding significant change in pupil diameter during the same tasks (Recarte & Nunes, 2003). This supports the theory of compensatory mechanisms for increased mental workload (Hockey, G. R. J., 1997)—the "passive coping" strategy is shown here wherein subject's performance goal targets are reduced in order to satisfy mental resource demands. In this situation, the driver's visual search patterns are the behavior whose target is reduced, and in that reduction the driver finds himself searching smaller areas to compensate for the increased effort.

This pattern is not limited to on-road driving. In a simulator-based study, drivers engaged in a verbal response task during multiple driving segments while a battery of eye tracking measures were assessed. Drivers showed expected increases in pupil diameter during the dual task, but once the secondary task performance began to decline (indicating the excessive mental workload threshold had been reached), the pupil diameter returned to a baseline level while horizontal convergence simultaneously increased (smaller horizontal visual search area) (Tsai et al., 2007). Similarly, Ting et al. (2010) created an operator interface simulator where operator mental workload was increased by absorbing an increasing amount of tasks from an automated process—both psychophysiological and performance measures were evaluated. In their initial model Ting et al. (2010) found that they were able to create a decision-making process that shared tasks between the automated interface and the operator based on the operator's stress "state," which was proposed as a method of mitigating the operator compensatory behaviors associated with the high levels of mental workload.

What these studies are missing is a definitive comparison between environments to see if the effects of this increased mental workload display absolute or relative validity. Table 1 shows the current type and environments featured in eye tracking mental workload driving studies. One study compared visual measures in a fixed-base simulator with on-road driving while interacting with a mock IVIS display for several typing tasks (Wang et al., 2010). The visual measures that were studied found absolute validity for total glance duration towards a typing task interface during task completion, relative validity with a higher number of glances in the real world compared to the simulator, and absolute validity for percentage of time eyes spent on the road while completing that same typing task. While valuable, the study design used separate participants for the simulator and the field trials and focused mainly on a proxy IVIS task instead of general levels of mental workload, so individual differences that may be present as mental workload compensatory techniques are not explored in these data.

Table 1. Environmental Focus for Eye Tracking Workload Studies

Simulator Only	**On Road Only**	**Both Sim and On Road**
Tsai et al., 2007	Recarte & Nunes, 2003	Wang et al., 2010
Benedetto et al., 2011	Recarte & Nunes, 2000	
Donmez et al., 2007	Sodhi et al., 2002	
Borowsky et al., 2014	Harbluk et al., 2007	
Di Stasi et al., 2010	Reimer, 2009	
(Nabatilan et al., 2012)	(Fu et al., 2011)	

To summarize, literature on the effects of mental workload on driver visual behavior generally consists only of eye gaze measures as standalone metrics, occasionally supplemented by either a single physiological parameter (usually pupil size) or driving performance data, rarely both from the same set of drivers. Further, the studies looking at eye tracking measures related to mental workload typically showcase either simulator eye tracking or real world eye tracking, rarely both. The advantage of the proposed study lies in having all major types of driver responses (physiological, driver performance, subjective, and visual performance) collected from the same participants, across both the simulator and real-world environments.

HYPOTHESES AND OBJECTIVE

The primary objective of this dissertation is to create a comparison between driver behavior on real roads and in a driving simulator, under applied mental workload. The specific hypotheses explored to complete this objective include studying how the independent variables of driver environment, scenario complexity, and whether or not an applied task is present influence dependent variables that describe driver responses. The data collected here were analyzed to identify 1) specific driver response variables that are best suited for on-road comparison to simulators; 2) the relationship between driver performance and physiological responses under heavy mental workload, comparing between the two research environments; and 3) to build a model of driver behavior in the real world, given driver behavioral data collected from the driving simulator. This dissertation is unique from past research in that it focuses on looking at how human behavior changes as it approaches the cognitive limits described in Hockey's (1997) work. By looking specifically at how drivers behave in simulators and on-road while under heavy mental workload, this dissertation provides a window into an aspect of validity that has not yet been explored in the research.

METHODS

Equipment

Instrumented Vehicle

A 2009 Chevy Impala was instrumented with a suite of sensors and data collection equipment, for naturalistic data collection. All instrumentation and support was performed by Digital Artefacts. GPS sensors recorded driver position information in sync with eye tracking and vehicle on board computer data. Seven cameras recorded the events occurring in and around the vehicle, with the video synced to all streaming data collection units.

Figure 5. Instrumented Vehicle Exterior (left) and Interior (right)

Eye Tracking. A 5-camera SmartEye eye tracking system was used to collect all eye tracking data while in the instrumented vehicle. The cameras were placed across the dashboard of the Impala, allowing a 135 degree horizontal field of view to be tracked. The system recorded at 60 Hz, collecting information about where the driver was looking on the road, pupil size and position, and eye closure information. A panoramic camera was mounted on top of the Impala, which collected a wide-angle view of the driver's view throughout all study drives.

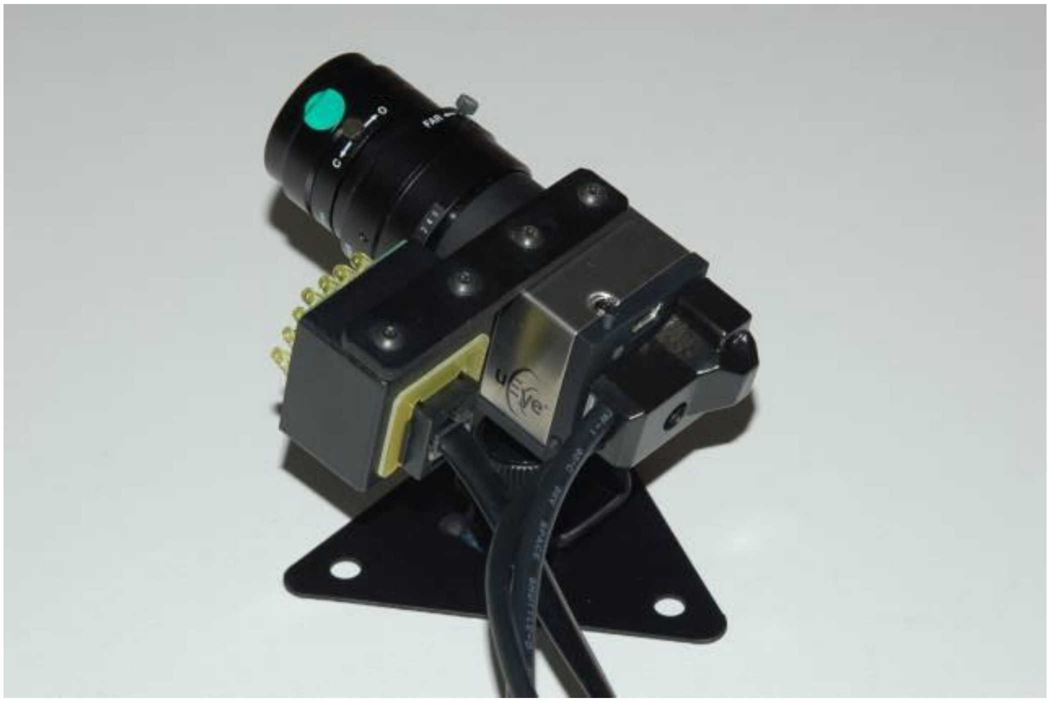

Figure 6. SmartEye Instrumented Vehicle Eye Tracker

<u>Lane Maintenance</u>. Several cameras were placed in specific positions within the interior cab and on both side-view mirrors. These cameras collected video which was later reduced into lane maintenance information, including lane position and steering heading variables.

<u>Vehicle Data</u>. The in-vehicle system tapped into the Impala's on-board computer to collect streaming vehicle data including vehicle speed, brake force, and accelerator force.

Driving Simulator

Western Transportation Institute’s advanced driving simulator consists of a 2009 Chevy Impala chassis, mounted on a Moog motion base (Vendor: Realtime Technologies, Inc). The motion base enables motion cues to be felt while driving, allowing up to 18 inches of movement on six degrees of freedom. The vehicle is surrounded by a wraparound projector screen, providing 240 horizontal degrees field of view. Simulator instrumentation and support was provided by Realtime Technologies, Inc. Cameras within and outside the vehicle cab recorded driver behavior synced to vehicle data.

Figure 7. Advanced Driving Simulator Exterior (left) and Interior (right)

Eye Tracking. A head-mounted Applied Science Labs MobileEye unit was used to collect all eye tracking data while in the simulator. This device consisted of a pair of glasses which contained cameras recording the forward view, as well as a camera recording the driver’s right eye. The system records at 30 Hz, collecting information about where the driver was looking relative to his forward view, and pupil size and position.

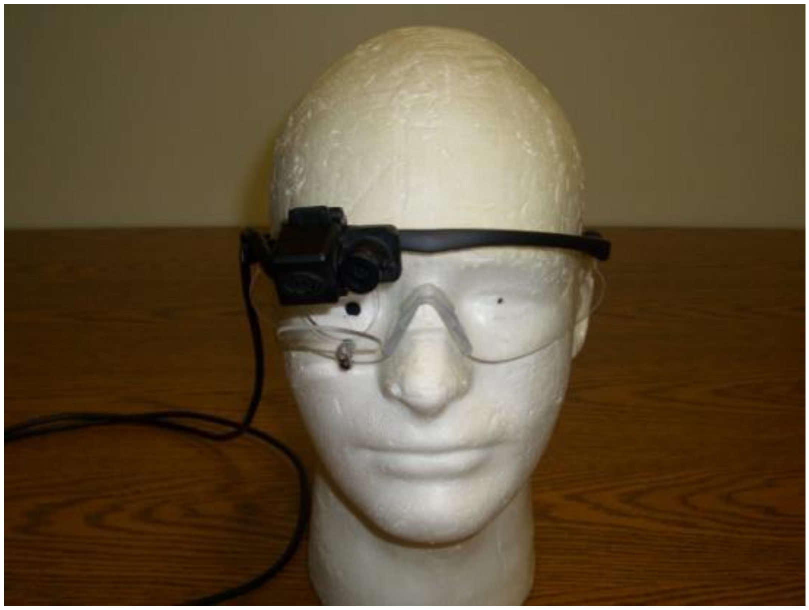

Figure 8. Mobile Eye Head-Mounted Eye Tracker

Lane Maintenance. The driver's physical location relative to the virtual scene was recorded by the simulator, as x-y coordinates (m). The precision of these coordinates allows for simple derivation of standard lane maintenance variables, including standard deviation of lane position (SDLP).

Vehicle Data. The driving simulator collects all streaming vehicle data generated by the driving simulation model, including speed, steering angle, and brake and accelerator force.

BioPac

For all experimental sessions, participants were equipped with adhesive leads that connected to a BioPac MP36 system. Leads were attached to participants' right inner wrist, inner left and right ankles, and the index and middle fingers of the left hand. These leads were used to collect electriocardiogram (ECG) and electrodermal activity (EDA) signals. All signal analysis was performed using BioPac Student Labs (BSL) 4.0. The only features that were used in BSL were the data filtering, data segmenting, and conversion of filtered ECG signal into heart rate; these features are identical to those offered in the research-grade BioPac analysis software.

Figure 9. BioPac MP36 Physiological Data Collection Unit (www.biopac.com, 2015).

To mitigate interaction with leads, the BioPac leads were rested out of the way across the driver's left wrist while the participant was not driving, and taped in position if the participant felt the leads would interfere with driving (Figure 10). There were still some issues with movement artifacts affecting data quality, which are addressed in the Methods' Data Handling section.

Figure 10. Participant wearing ASL eye tracker with EDA leads in place.

Scenario

A high and a low-level complexity scenario were chosen to model low- and high-levels of environmental complexity. Initially, roadways were selected in Bozeman, Montana that reflected the desired roadway characteristics. The low-complexity driving task consisted of a two-lane straight highway, divided by a double-yellow line, with very low levels of ambient traffic and pedestrian involvement. The selected roadway was Bozeman Trail Road. For a high-complexity scenario, a 4-lane straight roadway was selected, with heavy ambient traffic and a heavy pedestrian presence. The most appropriate roadway in Bozeman to satisfy these conditions was East Main Street, in the downtown area (Figure 10).

Figure 11. Scenario Images, by Environment and Complexity

Once the actual roadways were selected, virtual representations of those roadways were built for the simulator using Multigen Creator and Internet Scene Assembler (Vendor: Realtime Technologies, Inc). While not geospecific representations, the modeled roadways accurately reflected the ambient traffic levels, pedestrian presence, road characteristics, and building and vegetation styles of their real-world counterparts.

Secondary Task

A secondary task was selected in order to increase the drivers’ mental workload, without impacting their visual or tactile resources. The task that was used was a 1-back recitation task; in which a stream of 1 digit numbers was played to the drivers at a rate of 0.5 Hz; chosen to be similar to Mehler et al. (2008) who used a 0.44 Hz verbal response task. The drivers were instructed to repeat back the numbers, but had to maintain a delay of one number, instead of an immediate recitation of the number that was just presented. To accomplish this, the driver had to maintain the previous number in his working memory through the duration of the study scenario. The number sequences were generated randomly, and played over audio within the vehicle from a CD controlled by the researcher in the rear of the vehicle.

Independent Variable Selection

Independent variables were selected intentionally to increase task complexity. The variables selected were 1) driving task complexity, 2) secondary task, and 3) driving environment (simulator or real world). Elevated task complexity and multiple tasks are expected to increase applied workload, thereby moving from the automatic processes in Hockey's (1997) "loop A" from Figure 3 up to the elevated workload process in "Loop B." Elevating the workload is expected to elicit Hockey's compensatory strategies, which will be measured using the dependent variables described in the next section.

Dependent Variable Selection

Dependent variables are the same for both experiments, and were selected to reflect the difference outcomes described in Hockey's (1997) performance and effort mental processing model. In this model, the different types of compensatory strategies that people adopt when dealing with higher mental workload can be described using three main types of measures: physiological, performance, and subjective responses.

In Hockey's model, normal and excessive mental demand result in the outcomes displayed below, in Table 2. The dependent variables were selected in order to characterize the outcomes Hockey describes regarding the relationship between physiology and performance at high levels of mental workload.

Table 2. Hockey's (1997) Compensatory Strategies

	Compensatory Strategy	**Physiological Effect**	**Performance Effect**	**Effort Level**
Normal Demand	Active Coping	None observed	Within performance target	"Effort Without Distress"
Excessive Demand	Strain Coping	Increased sympathetic dominance	Within performance target	"Effort with distress"
	Passive Coping	None	Target widened	"Distress without effort"
	Complete Disengagement	Complete abandonment of task goals		

Physiological variables were collected or derived to identify when the sympathetic dominance effect associated with strain coping is evident in the drivers. The sympathetic nervous system is typically associated with a fight-or-flight response, and has several effects associated with different physiological systems. In sympathetic stimulation, heart rate increases with increased force, increased sweating, pupils can become dilated (Bradley & Lang, 2000). To assess heart rate changes, heart rate was collected, and heart rate variability (HRV) was derived from that ECG signal. To measure an increased sweat response, electrodermal activity (EDA) was gathered by recording electrical resistance between electrode leads attached to the driver. Pupil diameter was collected to identify if the sympathetic response toward manipulation of the study independent variables was evident through pupil dilation.

Performance dependent variables were collected to characterize the driver's performance, to see if target goals were being met, or if the driving task performance goals were being widened to compensate for the increased workload. The dependent performance variables measured in this study were selected from variables that have shown some sort of validity between simulators and real driving. The only variable that has repeatedly demonstrated absolute validity is speed (Bella, 2008; Blaauw, 1982, Reed & Green, 1999). Several other variables have shown relative validity between on-road and simulator driving: lateral displacement and lane position (Blaauw, 1982; Hakamies-Blomqvist et al., 2001), steering wheel angle (Hakamies-Blomqvist et al., 2001), mean braking onset (Hoffman et al., 2002), line crossings due to fatigue (Philip et al., 2005), and reaction time to brake (Toxopeus, 2007). These dependent measures will be collected to assess target performance goals, to see if drivers are adjusting their performance as a result of passive coping.

The dependent variables that were used to measure mental effort level include NASA-TLX self-reported measures of mental workload to aid in the characterization of the drivers' effort level.

A summary of all dependent variables can be seen below in Table 3. The selected variables are assumed to be accurate and sensitive enough to detect changes in sympathetic action, as they are known indicators of sympathetic dominance in psychophysiology literature (Bradley & Lang, 2000). The performance variables are expected to be accurate and sensitive enough to use to identify adjusting performance goals, as they have been measured before and found similar results in both simulator and real-world environments. NASA-TLX is a well-documented and validated survey for understanding self-reported mental workload, and so is also considered accurate and sufficient to understand the perceived workload experienced by the drivers.

Participants

Participants were recruited through online job listings, flyers posted on local advertisement boards, and word-of-mouth. Interested participants would contact the researchers through the given phone number, and were given a verbal survey over the phone to ensure that eligibility criteria were met. Eligibility criteria included: 1) a lack of sensitivity to simulator sickness to avoid attrition; 2) drivers aged between 25-35 years to avoid learning effects due to novice drivers, and to avoid age-related simulator sickness sensitivity; 3) a valid driver’s license; 4) no single- or multiple-vehicle crashes within the last 2 years; 5) an average driving exposure of at least 8000 vehicle miles travelled per year; and 6) 20/40 vision without glasses (contacts were allowed) in order to meet Montana’s MVD requirements for licensure as well as to avoid issues with the eye tracking equipment obscured by drivers with glasses. Thirty-four participants satisfying all eligibility criteria were scheduled to come in for multiple sessions to complete the study.

Table 3. Summary of dependent variables

	Dependent Variable	Collection Equipment	Units
Physiological Dependent Variables	Heart Rate	BioPac	Beats per minute
	Heart Rate Variability	BioPac	HF, LF, VLF, ULF [Hz]
	EDA	BioPac	µS
	Pupil Diameter	SmartEye, MobileEye	mm
Performance Dependent Variables	Steering Reversal Frequency	Simulator, IV Video	Reversals/min
	Steering Reversal Magnitude	Simulator, IV Video	°
	Lane Position (SDLP)	Simulator, IV Video	m
	SD of Steering Angle	Simulator, IV Video	°
Gaze-Related Dependent Variables	Horizontal Gaze Dispersion	ASL, SmartEye	No units
	Vertical Gaze Dispersion	ASL, SmartEye	No units
	On-road fixation duration	ASL, SmartEye	sec
	Off-road fixation duration	ASL, SmartEye	sec
	On-road fixation frequency	ASL, SmartEye	fix/minute
	Off-road fixation frequency	ASL, SmartEye	fix/minute
Effort / Workload Dependent Variables	Mental demand	NASA-TLX Survey	Likert (1-20)
	Physical demand	NASA-TLX Survey	Likert (1-20)
	Temporal demand	NASA-TLX Survey	Likert (1-20)
	Performance	NASA-TLX Survey	Likert (1-20)
	Effort	NASA-TLX Survey	Likert (1-20)
	Frustration	NASA-TLX Survey	Likert (1-20)

Over the course of the study, three drivers partially completed the study but had to withdraw due to simulator sickness. Three drivers could not complete the study due to scheduling issues. Drivers were 59.6 percent male, aged 29.86 (3.4) years (minimum age

25, maximum age 37 years), driving on average 15,611 (7,222) vehicle miles travelled per year.

Experimental Sessions

Participants completed two experimental sessions, each lasting approximately 1.5 hours. The sessions were either to complete the instrumented vehicle portion of the study, or the driving simulation portion. The order of the sessions was counterbalanced so that half of the participants drove the simulator first, and half drove the instrumented vehicle initially.

Upon arrival to their first session, participants signed informed consent in accordance with Institutional Review Board policies. Participants completed a vision test on an OPTEC machine, testing near and far visual acuity to ensure a minimum acuity of 20/40 was met in both eyes. The participant’s license was checked for validity, and demographic information was collected. A brief explanation of the session was provided, letting the participant know whether they would be driving on real or virtual roads. The adhesive leads were attached to the appropriate locations on the participants, who were then lead to the vehicle they would be driving. Participants were given instruction about how to complete the recitation task, and a practice one-back task was administered until the participant verbally acknowledged that they were comfortable with the task.

Simulator Session

Drivers were introduced to the driving simulator, where features of the vehicle were explained to participants. The participant was clipped in to the ECG and EDA leads, and the head-mounted eye tracking glasses were placed on the participant. Once comfortable with all of the instrumentation, the participant's gaze was calibrated against the forward image. Participants were instructed to indicate if they were feeling any discomfort, and reminded that they were allowed to withdraw from the study at any point during the session.

Participants first completed a practice drive, to familiarize themselves with the dynamics and visual displays of the driving simulator. The practice drive consisted of a straight two lane road, where the driver was instructed to reach increasing speeds (10, 25, 35, and 40 MPH), maintain that speed for 10 seconds, and come to a smooth controlled stop. After the 5-minute practice session, verbal confirmation was obtained in the participant felt comfortable operating the simulator, or if more practice were required. Participants only moved on to study drives once they confirmed they were comfortable with simulator operation.

Once comfortable, participants completed four separate drives. Each drive took approximately seven minutes to complete, depending on driver speed. Two low-complexity drives were driven back to back, and two high-complexity drives were completed back-to-back. The order of the high- and low-complexity pairs of drives was counterbalanced. During one of each of the pairs high- and low-complexity drives, drivers completed the one-back recitation task (counterbalanced). The eye tracking glasses were checked for calibration following and preceding each drive.

Immediately following each drive, participants completed a NASA-TLX mental workload questionnaire, assessing their self-reported levels of mental workload. Next, a

Kennedy Simulator Sickness Questionnaire was administered to ensure that they were not experiencing simulator sickness. A break was offered, and participants were asked if they would like to continue participation following each drive. If both sessions were completed, participants were debriefed and compensated; if this was the first session, participants were scheduled to come in for their second session.

Instrumented Vehicle Session

All instrumented vehicle sessions were scheduled to maximize ambient traffic levels, so they occurred during the “rush hour” periods of time: 8:30AM, 12:00PM, and 5:30PM. Participants were lead to the instrumented vehicle, which was already running prior to participant arrival. Outside the Impala, vehicle-specific features (automatic gear shifting, location of controls) were identified and explained to the participant, who was then instructed to adjust the rear-view mirror and side-view mirrors. Once the vehicle was adapted for personal use, the driver was hooked into the BioPac leads. Next, the driver completed a calibration procedure for the SmartEye eye tracking system, while the researcher remained in the rear-right seat with all accessory data collection equipment. Once all systems were calibrated and recording data, the driver drove for approximately 7 minutes prior to driving on the experimental roads to ensure comfort driving the new vehicle.

Depending on whether the route began with the low-complexity drives or the high-complexity drives, one of two possible routes was completed. Following each of the four drive segments, participants were directed to a safe parking area. While parked, participants completed a NASA-TLX form regarding the mental workload experienced on the drive segment she just completed. Once finished, the participant re-entered the roadway and drove toward the next segment. All navigation instructions were verbally provided by the researcher.

Upon completion of the session, the driver was instructed to return to the research lab. If all sessions were completed, he would be debriefed and compensated. If this were the first session, the participant was scheduled to come in for his second session.

Data Structure and Handling

Eye tracking data was assessed using MAPPS v3.1 (SmartEye data) and MAPPS v2.0 (ASL data). MAPPS is a utility in which dynamic areas of interest (AOIs) are manually created. These AOIs are a closed polygon surrounding a space of interest to the reviewer—in this study the only AOI that was created was the area surrounding the forward roadway. After creating these AOIs, MAPPS reduces the raw eye tracking data to describe fixations in the forward view that are on- or off-road.

Driver physiological variables were derived from the raw ECG and EDA signals using BioPac Student Lab 4.0. To derive heart rate variability (HRV) and heart rate variables, a FIR band-pass filter between 0.5 and 35 Hz was applied to the raw ECG signal, using 8000 coefficients to remove baseline drift and high frequency signal noise. HR and HRV were derived from the filtered ECG signal. Physiological data plots were visually assessed to identify segments affected by movement artifacts; these segments were manually filtered out during data reduction.

Video Reduction

The on board computer in the Impala was not equipped to automatically collect two groups of data included in this analysis: steering angle and lane position. Cameras in the Impala recorded video of the driver steering wheel throughout the experimental session. The video was processed to derive steering angle from a group of markings affixed to the steering wheel. The output was filtered in MatLab to reduce noise; further operations on the signal to derive variables of interest (steering reversal magnitude and frequency) were performed in SAS. Plots of each road segment's steering angle were visually assessed, any plots with excessive variability were flagged for individual review. Examples of "good" and "bad" plots of steering angle over time are shown below in Figure 12. Video that had an obscured view was not suitable for processing or analysis, and was omitted from the study. If the obscuring feature was temporary (e.g., a hand passing briefly over the steering wheel markings, Figure 12), then the obstructed data was cut from the set and analyses were performed on the remaining unobstructed sections of data.

The steering angle data were processed next in SAS, where the local minima and maxima were identified and flagged to denote individual steering reversals; these were further reduced to include reversals only over a threshold of 0.5 degrees. Information on the magnitude and frequency of these steering reversals was derived from the video output.

Similarly, video data was processed to identify driver's lane position. The video processing depended on the presence of stark boundaries indicating a lane, such as painted lines or a sharp division between asphalt and the shoulder. This data was similarly filtered in Matlab, to derive the distance from the outer edge of the left and right wheels to the lane boundary on either side. One main difference between real and virtual roads is that actual roadway does not have a constant width. To get the most accurate lane

Figure 12. Steering angle for obscured (left) and unobscured (right) video

position, center line boundaries were used as the consistent reference for the roadway lane position.

Separate databases were maintained for demographic information, self-reported survey data, simulator sickness surveys, simulator eye tracking data, instrumented vehicle eye tracking data, simulator vehicle data, instrumented vehicle data, physiological data, and MAPPS fixation output for both the simulator and instrumented vehicle. These databases were reduced to describe variables belonging to four different sets:

physiological, driving performance, gaze-related performance, and self-reported mental workload, all data reduction and data munging were performed using R v 3.1.2, RStudio, SAS v 9.3, and Matlab 2014b.

STATISTICAL ANALYSIS

Driver observations were excluded from study if the corresponding drive ended with a SSQ score greater than 50 (16 drives excluded); this was to exclude driver behavior due to compensating for simulator sickness instead of the mental workload level, which was used as a conservative level of Balk, Bertola, and Inman's (2013) cutoff score of 55 . Two main areas of analysis are explored here: 1) a Multivariate Analysis of Variance approach to look at specific variables that describe driver performance and physiology, and 2) Methods to predict instrumented vehicle performance given simulator driver data as predictors.

Multivariate Analysis of Variance

Each of the four groups of dependent variables were evaluated separately to determine the effects of environment, task, and complexity on the related variables. Each variable was individually assessed for normality using Shaprio-Wilks, and if the variable was found to be non-normal then it was log-transformed for the analyses and back-transformed for reporting results. A Multivariate Analysis of Variance (MANOVA) was conducted on each set, to identify significant factors related to the variable types. MANOVA was selected to test to avoid inflating type I error due to multiple testing, and to take into account potential correlation between the dependent variables. For each separate MANOVA, an initial saturated model was fit consisting of all first-order and second-order interactions; a second reduced model was then created, excluding all non-significant interaction effects. Pillai's criterion was used to assess MANOVA significance, as it is a more robust test statistic than the traditionally used Wilks' lambda when sample sizes are small and there are are unequal cell sizes.

Significant results are reported, along with tables and plots describing the MANOVA effects. Single variable ANOVAs were performed on each significant variable detected in the MANOVA, and post-hoc comparison tests were conducted using Bonferroni correction factors to control for spurious results. All data analyses were performed using SAS 9.3 and R 3.1.2.

Physiological Measures

Using Pillai's trace, there were significant effects of Task (V=0.32, F=15.97, p<0.0001), Complexity (V=0.07, F=2.49, p=0.0461) and Environment (V=0.29, F=14.23, p<0.0001) on physiological dependent variables. The only significant interaction effect was the interaction between Task and Environment (V=0.18, F=7.55, p<0.0001).

Table 4. Physiological MANOVA Results

	DF	Pillai	Approx F	Pr(>F)
Participant	27	2.33880	7.35	<0.0001
Task	1	0.31646	15.97	<0.0001
Complexity	1	0.06731	2.49	0.0461
Environment	1	0.29199	14.23	<0.0001
Task*Environment	1	0.17960	7.55	<0.0001
Residuals	142			

Separate univariate ANOVAs on the physiological outcome variables showed non-significant treatment effects for driver mean EDA (Figure 13). Sympathetic HRV had significant treatment effects from Task (F=11.47, p=0.0009) and Environment (F=8.55, p=0.0039) main effects (Figure 14). Heart rate had significant treatment effects only from Task ($F=46.60$, $p<0.0001$; Figure 15). Pupil diameter relative to the baseline condition had significant treatment effects due to Task ($F=13.01$, $p=0.0004$), Complexity ($F=8.75$, $p=0.0036$), and Environment ($F=36.94$, $p<0.0001$; Figure 17). The interaction effect between Task and Environment was also significant for relative pupil diameter ($F=32.93$, $p<0.0001$).

Table 5. Significant Effects from Physiological Univariate ANOVAs

	P-Values from Follow-up Univariate ANOVAs			
	Task	**Complexity**	**Environment**	**Task*Environment**
Sympathetic HRV	0.0009	NS	0.0039	NS
HR	<0.0001	NS	NS	NS
MeanEDA	NS	NS	NS	NS
Relative Pupil Diameter	0.0004	0.0036	<0.0001	<0.0001

Post hoc least squared mean assessments adjusted for multiple comparisons show that the sympathetic HRV component is significantly higher while the secondary task was being completed ($\mu_{task}=0.625$, $\mu_{NoTask}=0.555$; $p=0.0007$); and was also higher while driving in the simulator compared to the instrumented vehicle ($\mu_{Sim}=0.621$, $\mu_{IV}=0.559$; $p=0.0041$). Driver heart rate was significantly higher while the task was being completed ($\mu_{Task}=78.18$, $\mu_{NoTask}=73.55$ BPM; $p<0.0001$). Relative pupil diameter was higher while a task was being completed ($\mu_{Task}=108.02$, $\mu_{NoTask}=101.81$ percent; $p<0.0001$), higher when driving on a more complex road ($\mu_{Complex}=107.53$, $\mu_{Simple}=102.03$ percent; $p<0.0001$), and higher in the driving simulator compared to real roads ($\mu_{Sim}= 109.08$ percent, $\mu_{IV}=100.75$; $p<0.0001$). While the simulator environment is darker than the real world, these size differences are based on a relative pupil diameter in the low complexity and no-task condition—so the size difference is not due to the simulator’s darker ambient light.

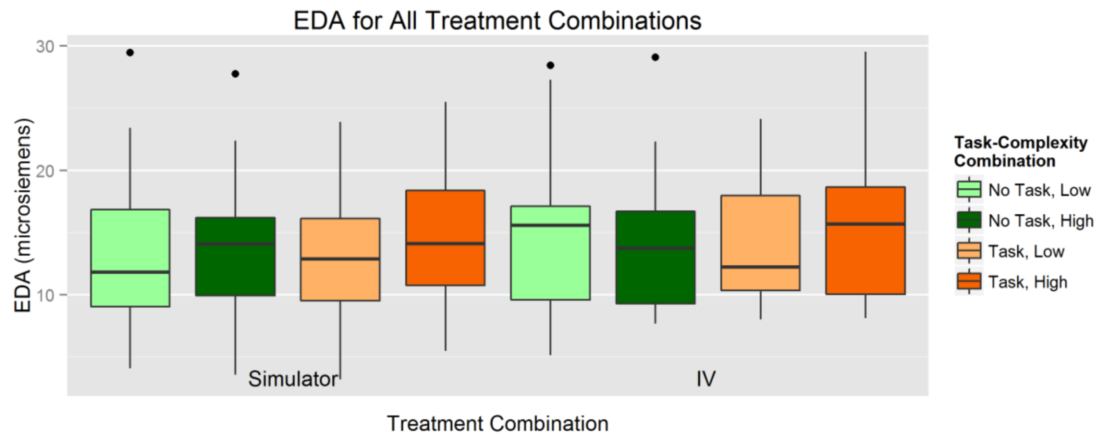

Figure 13. EDA Boxplot for All Treatment Conditions

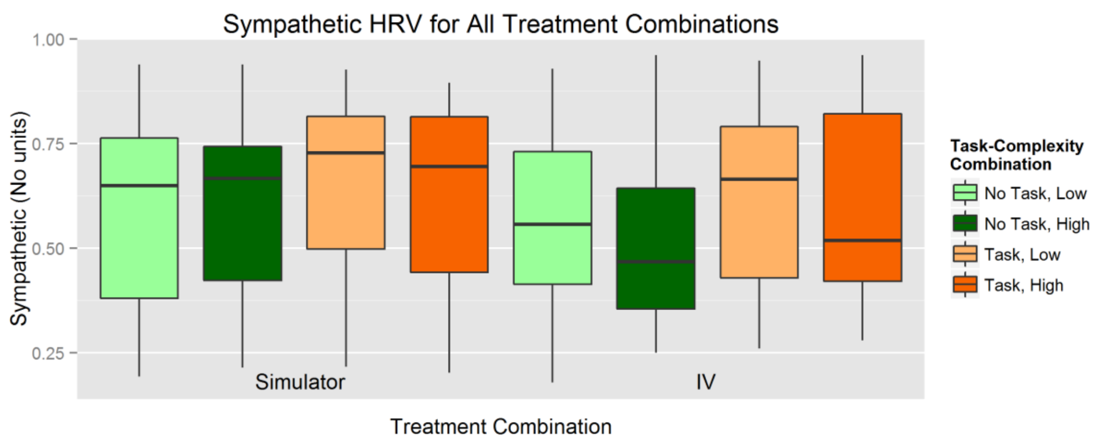

Figure 14. Sympathetic HRV Boxplot for All Treatment Conditions

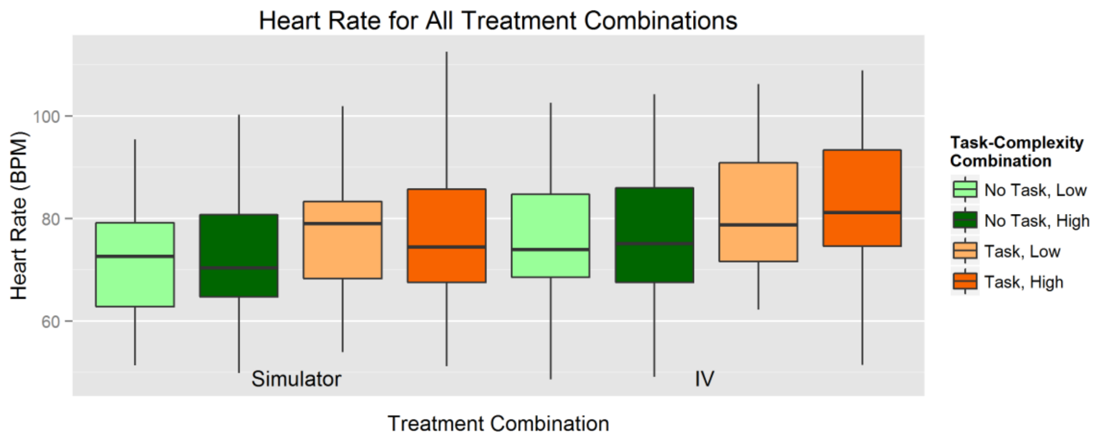

Figure 15. Heart Rate Boxplot for All Treatment Conditions

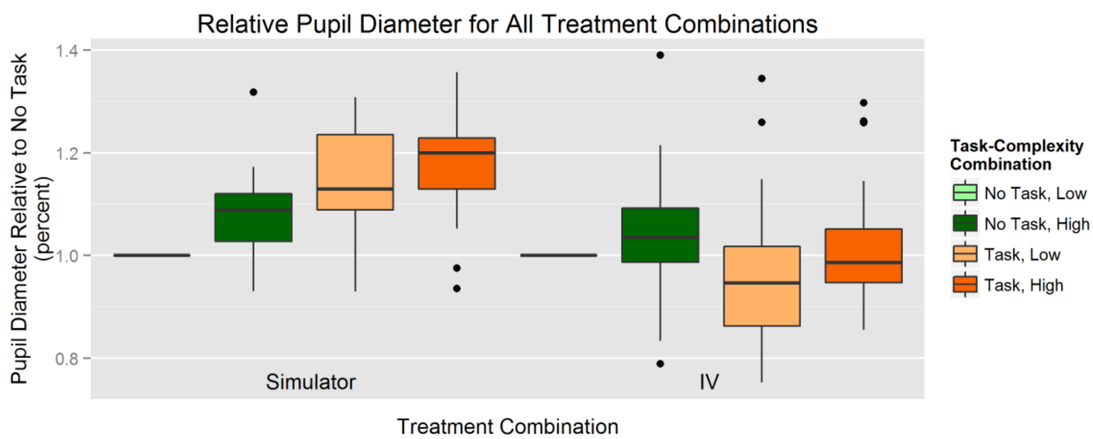

Figure 16. Relative Pupil Diameter Boxplot for All Treatment Conditions

The interaction effect was only significant for the relative pupil diameter dependent variable (Figure 17). For all the combinations of driving environment and whether or not the task is being performed, driving the simulator while performing the secondary task had the highest relative pupil diameter ($\mu_{Sim,Task}$=117.7 percent). The three other conditions were not significantly different from each other ($\mu_{Sim,NoTask}$=104.0 percent, $\mu_{IV,NoTask}$=102.1, μ_{IVTask}=98.7). While it looks similar between the “Task” and

"No Task" conditions in the instrumented vehicle, there is no significant difference between pupil diameter in those conditions.

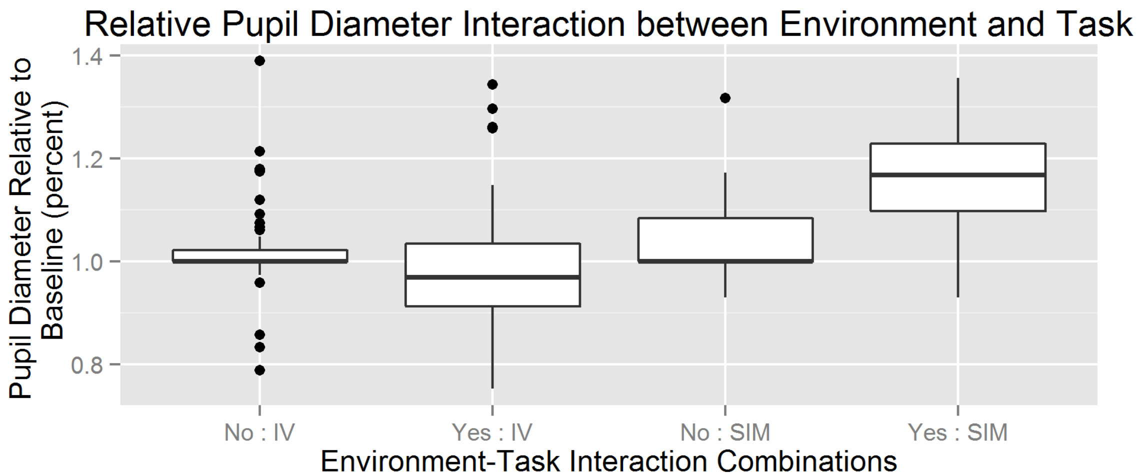

Figure 17. Pupil Diameter Interaction between Environment and Task

<u>Performance Measures</u>

Pillai's trace showed significant effects of Task (V=0.10, F=5.61, p<0.0001), Complexity (V=0.09, F=4.73, p=0.0035) and Environment (V=0.82, F=226.55, p<0.0001) on physiological dependent variables (Table 6). The significant interaction effects were the between Task and Environment (V=0.06, F=3.38, p=0.0200) and Complexity and Environment (V=0.13, F=7.98, p<0.0001).

Table 6. Performance MANOVA Results

	DF	Pillai	Approx F	Pr(>F)
Participant	28	1.41665	4.83	<0.0001
Task	1	0.10147	5.61	<0.0001
Complexity	1	0.08702	4.73	0.0035
Environment	1	0.82019	226.55	<0.0001
Task*Environment	1	0.06371	3.38	0.0200
Complexity*Environment	1	0.13841	7.98	<0.0001
Residuals	149			

Separate univariate ANOVAs on the performance outcome variables (Table 7) showed that only environment was a significant treatment effect for standard deviation of steering angle (F=4.08, p=0.0450). Steering reversal magnitude had scenario complexity as a significant factor (F=4.82, p=0.0297), along the interaction effects between Complexity and Environment (F=5.15, p=0.0247), and a marginally significant interaction between task and environment (F=3.04, p=0.0830).

Table 7. Significant Effects from Performance Univariate ANOVAs

	P-Values from Follow-up Univariate ANOVAs				
	Task	Comp.	Env.	Task*Env	Comp * Env
SteeringSD	NS	NS	0.0450	NS	NS
Reversal Magnitude	NS	0.0297	NS	0.0830	0.0247
Reversal Frequency	0.0049	0.0289	<0.0001	0.0020	0.0044

Post hoc least squared mean assessments adjusted for multiple comparisons show that the steering angle standard deviation was significantly higher while driving in the driving simulator (μ_{Sim}=5.05 degrees; μ_{IV}=2.18 degrees; p=0.0427; Figure 18). Steering reversal magnitudes were larger on more complex roadway ($\mu_{Complex}$=2.94 degrees, μ_{Low}=2.46; p=0.0355; Figure 19). Looking at the interaction between task and environment, reversal magnitudes had an opposite shift. While not significantly different, No Task in the IV had larger reversals than Task drives; but in the simulator the larger reversals were seen while performing the task. Looking closely at the Complexity-Environment interaction for steering reversal magnitude, the combinations of complexity and environment were not very different from each other, except for the low-complexity simulator drives, which had much smaller reversals.

The frequency of steering reversals was significantly higher while a task was being performed (μ_{Task}=28.97 reversals per minute, μ_{NoTask}=25.06; p=0.0024; Figure 20). Drivers had more frequent steering reversals in the high-complexity scenarios (μ_{High}=29.59 reversals per minute, μ_{NoTask}=24.44; p<0.0001). Reversal frequency was also higher in the driving simulator than on real roads (μ_{Sim}=41.71 reversals per minute, μ_{NoTask}=12.32; p<0.0001). The instrumented vehicle reversal frequencies were not different from each other at different task levels, however the simulator had a pronounced difference where there were far more frequent reversals while performing the task than there were without a secondary task. The same trend was evident in the complexity-environment interactions; there was no difference between real world reversal frequencies; but the simulators reversal frequencies were higher than the real car; within the simulator, the higher-complexity scenarios in the simulator had more frequent reversals than the low-complexity drives.

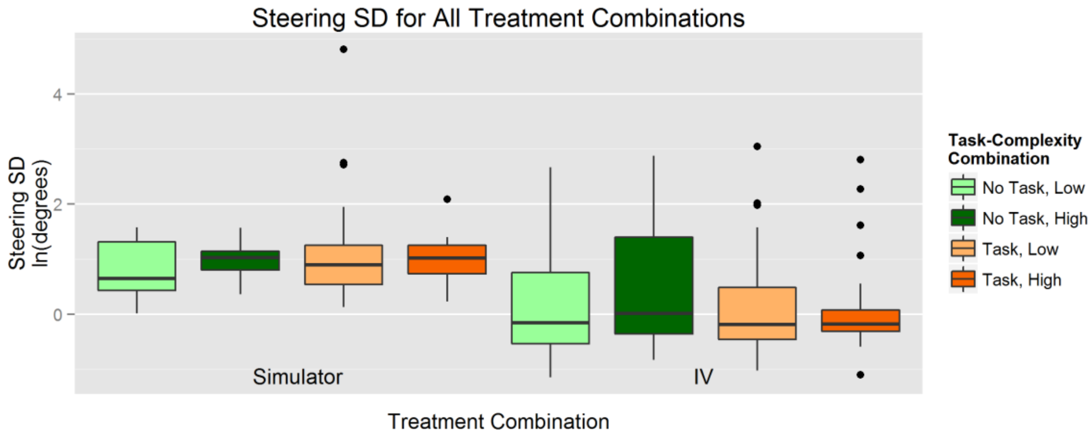

Figure 18. Steering Angle SD Boxplot for All Treatment Combinations

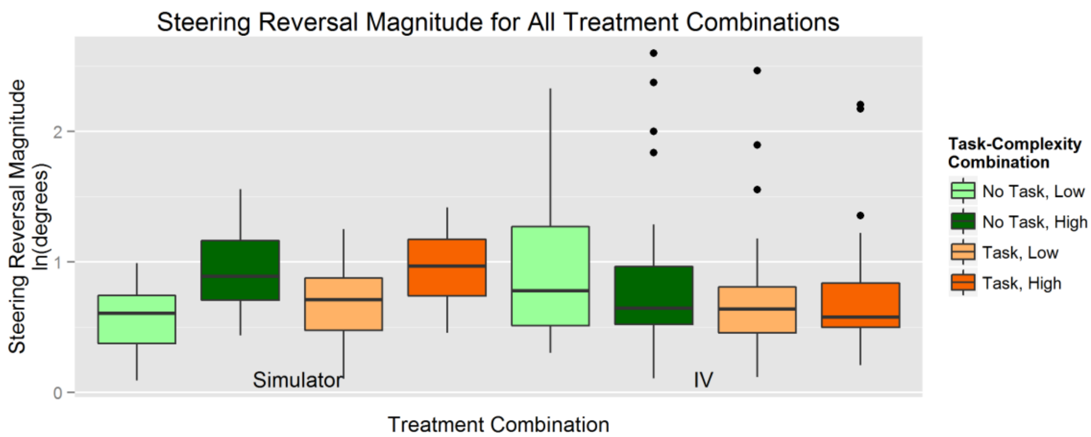

Figure 19. Steering Reversal Magnitude Boxplot for All Treatment Combinations

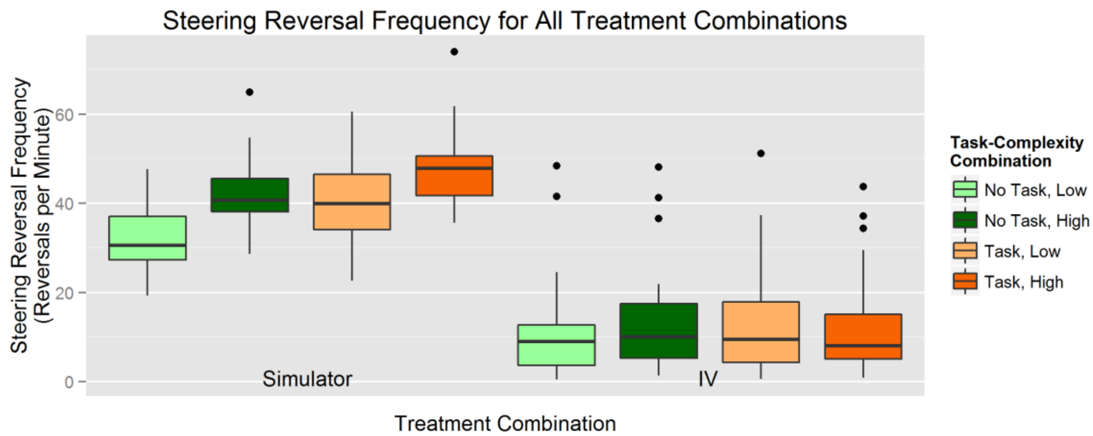

Figure 20. Steering Reversal Frequency Boxplot for All Treatment Combinations

Self-Reported Workload

Pillai's trace showed significant main effects (Table 8) for secondary task (V=0.77, F=89.25, p<0.0001) and complexity (V=0.14, F=4.35, p=0.0004). The only significant interaction effect was between Environment and Complexity (V=0.11, F=3.55, p=0.0025).

Table 8. NASA-TLX MANOVA Results

	DF	Pillai	Approx F	Pr(>F)
Participant	29	2.07100	3.12	<0.0001
Task	1	0.76885	89.25	<0.0001
Complexity	1	0.13950	4.35	0.0004
Environment	1	0.05209	1.47	0.1901
Env*Complexity	1	0.11696	3.55	0.0025
Residuals	163			

Follow-up univariate ANOVAs on gaze outcome variables (Table 9) showed that self-reported mental demand was significantly affected by whether or not a task was being performed (F=472.66, p<0.0001), scenario complexity (F=21.18, p<0.0001), and the interaction between the driving environment and complexity (F=8.00, p=0.0053;

Figure 21). Physical demand was affected by Task ($F=29.43$, $p<0.0001$), Complexity ($F=8.67$, $p=0.0037$, and environment-complexity interaction ($F=9.34$, $p=0.0026$; Figure 22). Temporal demand was affected by Task ($F=142.61$, $p<0001$), Complexity ($F=11.49$, $p=0.0009$), and Complexity-Environment interaction ($F=11.65$, $p=0.0008$; Figure 23). Self-rated driving performance was affected by Task ($F=134.47$, $p<0.0001$), Environment ($F=5.63$, $p=0.0188$), and Environment-Complexity interaction ($F=7.75$, $p=0.0060$; Figure 24). Effort was affected by Task ($F=355.33$, $p<0.0001$), Complexity ($F=12.34$, $p=0.0006$), Environment ($F=4.45$, $p=0.0364$), and Environment-Complexity Interaction ($F=7.97$, $p=0.0053$; Figure 25). Driver frustration was affected by Task ($F=177.07$, $p<0.0001$), Complexity ($F=3.95$, $p=0.0486$), and Environment-Complexity interaction ($F=5.63$, $p=0.0188$; Figure 26).

Table 9. Significant Effects from NASA-TLX Univariate ANOVAs

	P-Values from Follow-up Univariate ANOVAs			
	Task	Complexity	Environment	Env*Complexity
Mental Demand	<0.0001	<0.0001	NS	0.0053
Physical Demand	<0.0001	0.0037	NS	0.0026
Temporal Demand	<0.0001	0.0009	NS	0.0008
Performance	<0.0001	NS	0.0188	0.0060
Effort	<0.0001	0.0006	0.0364	0.0053
Frustration	<0.0001	0.0486	NS	0.0188

Post-hoc least squared means with Bonferroni adjustments for multiple comparisons were done on significant univariate factors. While performing the secondary task, drivers reported higher self-reported mental demand ($\mu_{Task}=13.26$, $\mu_{NoTask}=4.53$; $p<0.0001$), higher physical demand ($\mu_{Task}=5.68$, $\mu_{NoTask}=3.76$; $p<0.0001$), higher temporal demand ($\mu_{Task}=9.28$, $\mu_{NoTask}=4.29$; $p<0.0001$), higher effort ($\mu_{Task}=12.67$,

μ_{NoTask}=4.37; p<0.0001), and higher frustration (μ_{Task}=9.46, μ_{NoTask}=3.55; p<0.0001), with a lower self-rated driving performance (μ_{Task}=12.07, μ_{NoTask}=16.81; p<0.0001).

Drivers in the more complex driving roadway (compared to the simple complexity condition) reported significantly higher mental demand (μ_{High}=9.81, μ_{Low}=7.92; p<0.0001), physical demand (μ_{High}=5.27, μ_{Low}=4.17; p=0.0018), temporal demand (μ_{High}=7.56, μ_{Low}=6.01; p=0.0003), effort (μ_{High}=9.37, μ_{Low}=7.68; p=0.0002), and frustration (μ_{High}=7.02, μ_{Low}=6.00; p=0.0235)—no difference was found regarding self-reported performance and roadway complexity.

While driving in the simulator, drivers reported lower performance (μ_{SIM}=13.88, μ_{IV}=15.00; p=0.0100) and higher effort (μ_{SIM}=9.07, μ_{IV}=7.97; p=0.0192) than while driving in the real world.

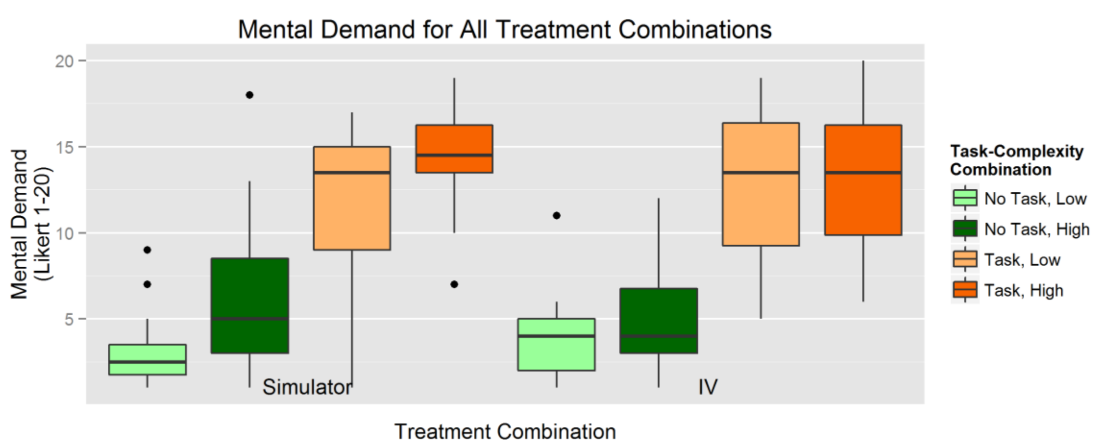

Figure 21. NASA-TLX Mental Demand Boxplot for All Treatment Combinations

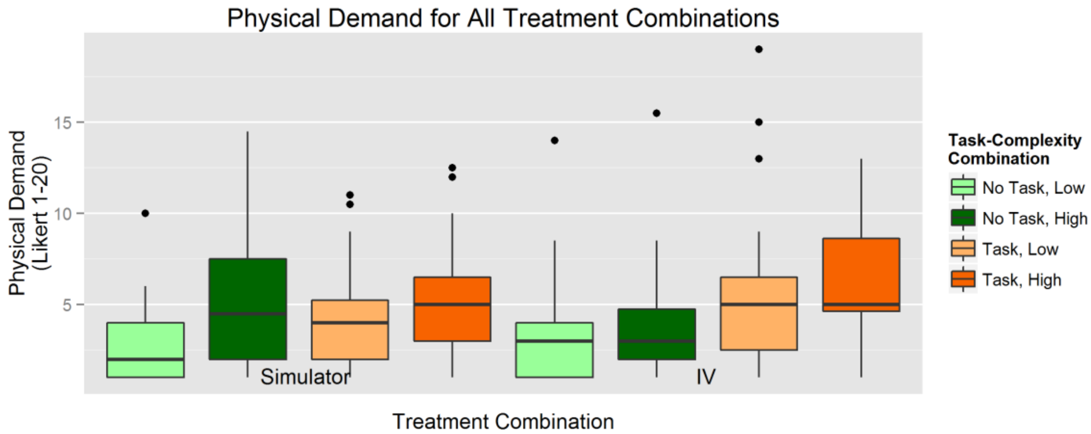

Figure 22. NASA-TLX Physical Demand Boxplot for All Treatment Combinations

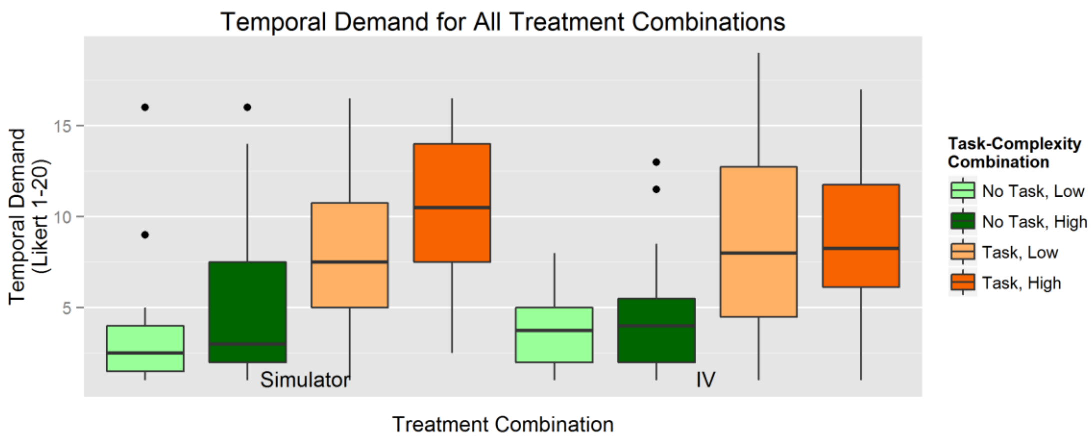

Figure 23. NASA-TLX Temporal Demand Boxplot for All Treatment Combinations

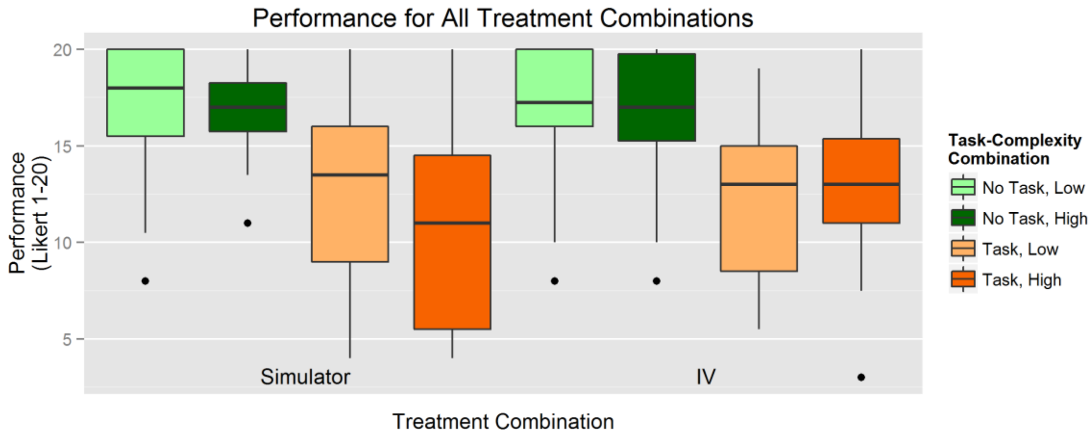

Figure 24. NASA-TLX Performance Boxplot for All Treatment Combinations

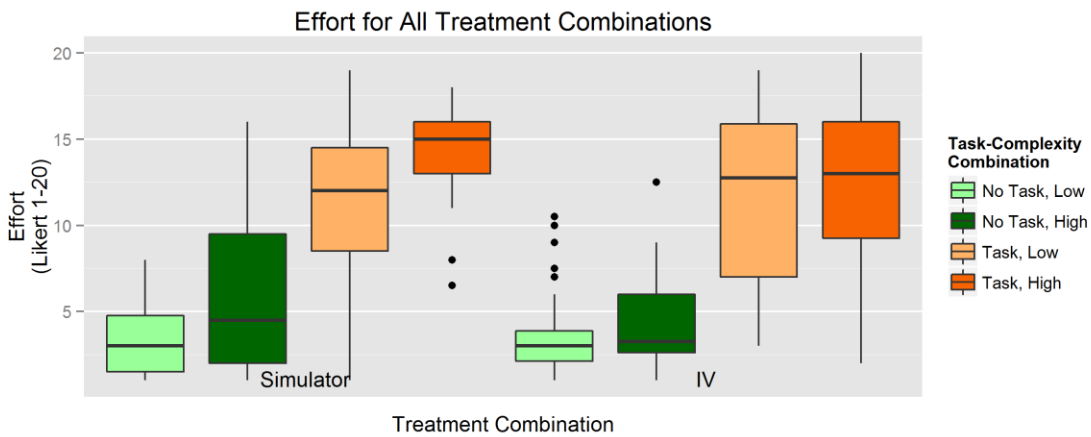

Figure 25. NASA-TLX Effort Boxplot for All Treatment Combinations

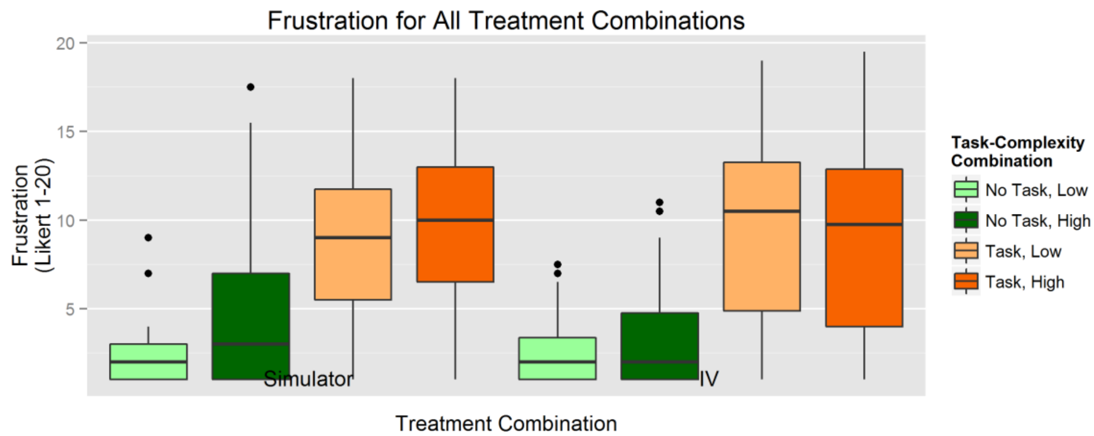

Figure 26. NASA-TLX Frustration Boxplot for All Treatment Combinations

As far as interaction effect, the combination of a high-complexity scenario while driving in the real world had higher reported TLX-metrics than in the other combinations; this is illustrated in the grouped interaction box-plot shown below (Figure 27).

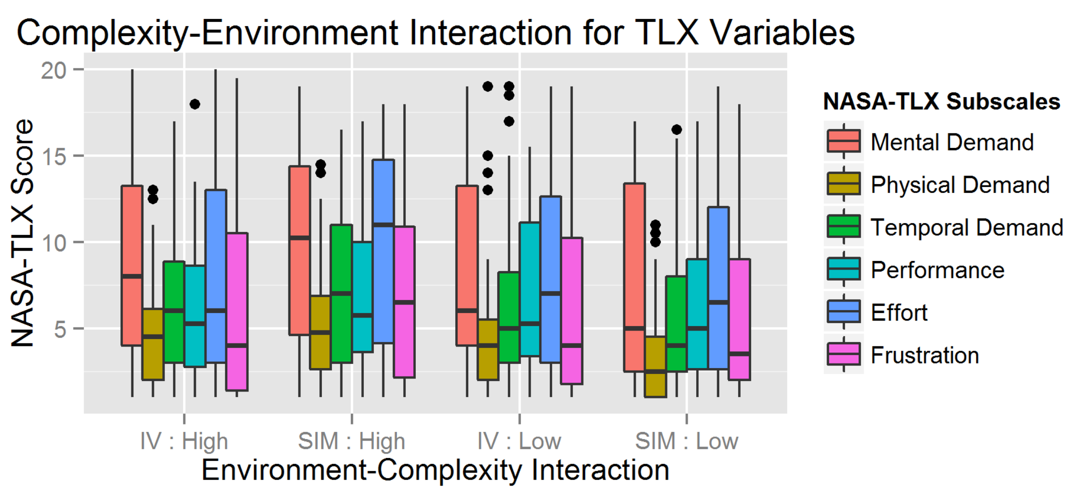

Figure 27. TLX Complexity-Environment Interaction Boxplot

<u>Gaze-Related Performance</u>

Pillai's trace showed significant effects (Table 10) from Task (V=0.40, F=12.73, p<0.0001), Complexity (V=0.22, F=5.40, p<0.0001), and Environment (V=0.77, F=61.57, p<0.0001). The only significant interaction effect was between Task and Environment (V=0.15, F=3, p=0.0042).

Table 10. Gaze-Related MANOVA Results

	DF	Pillai	Approx F	Pr(>F)
Participant	29	2.67871	3.52	<0.0001
Task	1	0.40333	12.73	<0.0001
Complexity	1	0.22269	5.40	<0.0001
Environment	1	0.76575	61.57	<0.0001
Task*Environment	1	0.15233	3.38	0.0042
Residuals	132			

Follow-up univariate ANOVAs on the gaze outcome variables showed no significant treatment effects for off-road fixation durations (Table 11). Horizontal gaze dispersion was significantly affected by the secondary task (F=57.03, p<0.0001), roadway complexity (F=4.48, p=0.0361), driving environment (F=278.92, p<0.0001), and the interaction between the environment and secondary task (F=7.55, p=0.0068). Vertical Gaze Dispersion showed significant effects due to Task (F=18.45, p<.0001), complexity (F=5.04, p=0.0263), Environment (F=245.86, p<0.0001), and the Environment-Task interaction (F=6.49, p=0.0119). Road fixation duration was significantly affected by the driving environment (F=91.49, p<0.0001); the road fixation frequency was affected similarly by driving environment (F=59.42, p<0.0001). There were no significant factors identified for mean off-road fixation duration. Off-Road

Fixations per Minute had significant effects due to the three main effects: Task ($F=17.00$, $p<0.0001$), Complexity ($F=8.17$, $p=0.0050$), and Environment ($F=6.42$, $p=0.0126$).

Table 11. Significant Effects from Gaze Univariate ANOVAs

	P-Values from Follow-up Univariate ANOVAs			
	Task	Complexity	Environment	Task*Environment
GazeDispX	<0.0001	0.0361	<0.0001	0.0068
GazeDispY	0.0001	0.0263	<0.0001	0.0119
RoadFixationDuration	NS	NS	<0.0001	NS
OffRoadFixationDuration	NS	NS	NS	NS
RoadFixationPerMin	NS	NS	<0.0001	NS
OffRoadFixationPerMin	<0.0001	0.0050	0.0126	NS

Horizontal gaze dispersion (Figure 28) increases when no secondary task is being performed ($SD_{Task}=0.0593$, $SD_{NoTask}=0.0770$; $p<0.0001$), increases on the high-complexity roadway ($SD_{Complex}=0.0722$, $SD_{Simple}=0.0642$; $p=0.0023$), and increases in the simulator compared to the instrumented vehicle ($SD_{Sim}=0.0916$, $SD_{IV}=0.0447$; $p<0.0001$). Vertical gaze dispersion (Figure 29) follows the same trends: increases when no secondary task is performed ($SD_{Task}=0.0891$, $SD_{NoTask}=0.1041$; $p=0.0002$), and increases in the simulator ($SD_{Sim}=0.1309$, $SD_{IV}=0.0622$; $p<0.0001$). Average road fixation duration (Figure 30) is higher in the real world than it is in the simulator ($\mu_{IV}=0.72$ seconds, $\mu_{Sim}=0.58$ sec; $p<0.0001$). No post-hoc differences were performed for off-road fixation durations (Figure 31). Drivers in the real world had higher frequency of on-road fixations (Figure 32) per minute than Sim drivers ($\mu_{IV}=37.95$ fixations per minute, $\mu_{Sim}=25.44$; $p<0.0001$). Drivers had fewer frequent fixations per minute off-road (Figure 33) while they were completing the secondary task ($\mu_{NoTask}=3.10$, $\mu_{Task}=1.78$ fixations per minute; $p<0.0001$). Fewer off-road fixations per minute were also seen on the low-complexity roadway (μ_{Simple}=1.92 fixations per minute, $\mu_{Complex}$=2.88; p=0.0038). Drivers in the simulator had more frequent off-road fixations per minute than real-world drivers (μ_{Sim}=2.84 fixations per minute, μ_{IV}=1.95; p=0.0177).

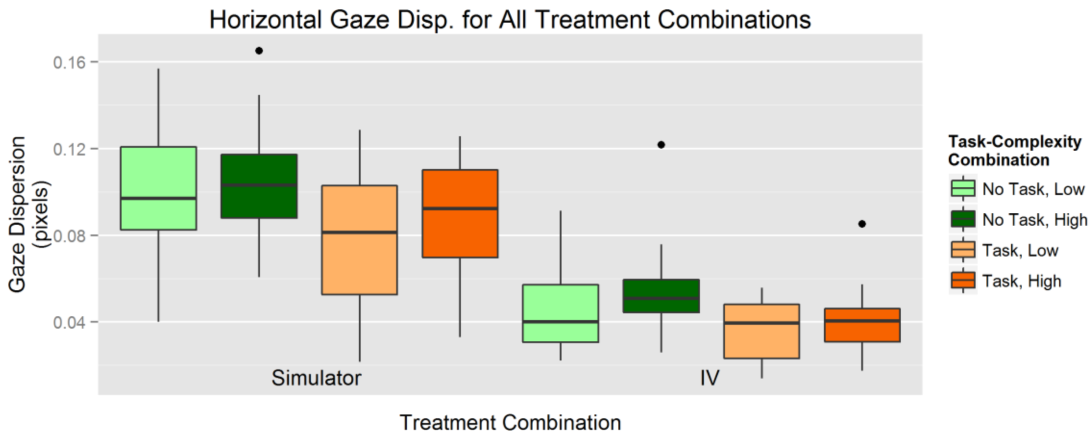

Figure 28. Horizontal Gaze Dispersion Boxplot for All Treatment Combinations

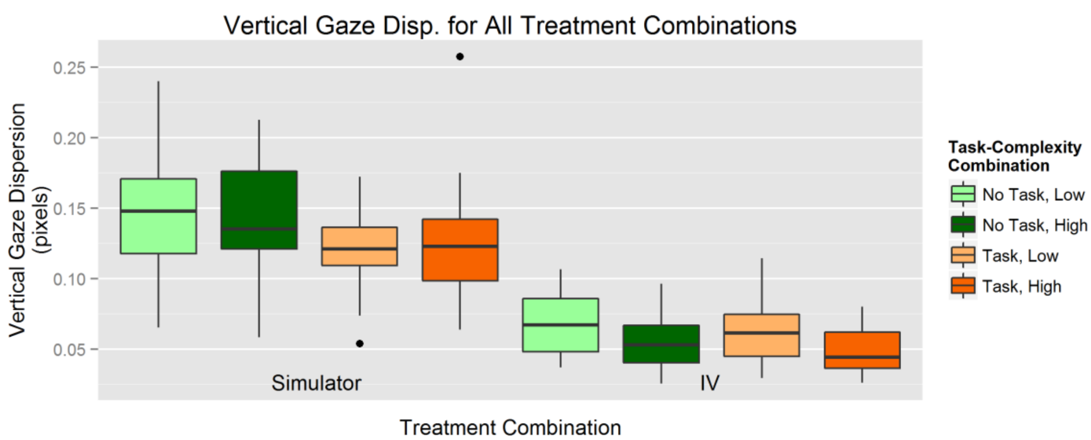

Figure 29. Vertical Gaze Dispersion Boxplot for All Treatment Combinations

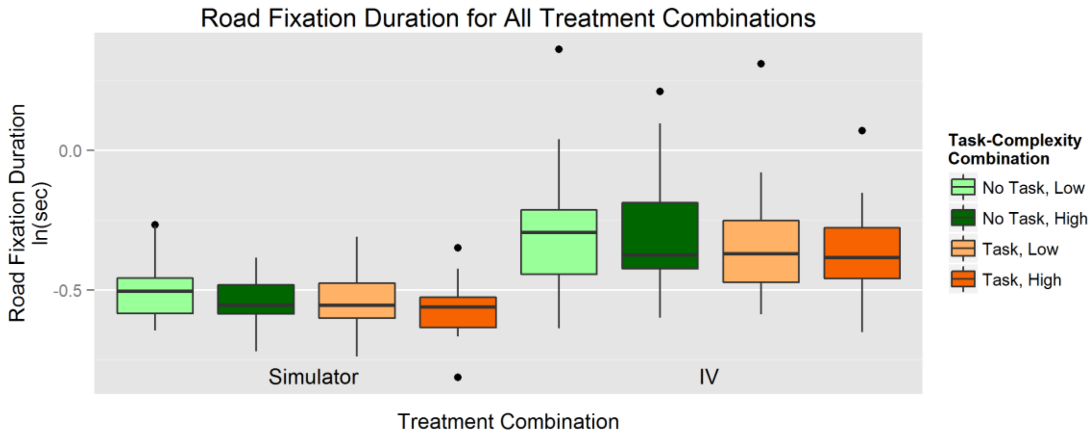

Figure 30. Road Fixation Duration Boxplot for All Treatment Combinations

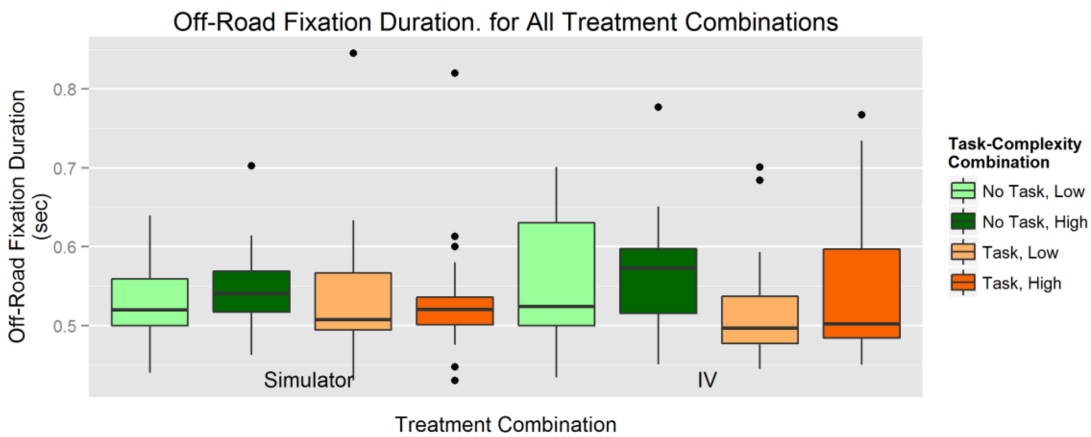

Figure 31. Off-Road Fixation Duration Boxplot for All Treatment Combinations

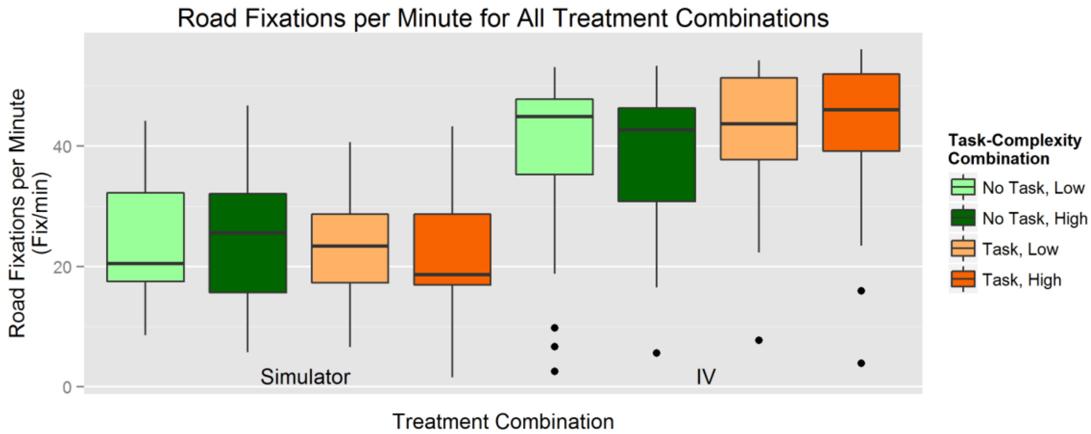

Figure 32. Road Fixation Frequency Boxplot for All Treatment Combinations

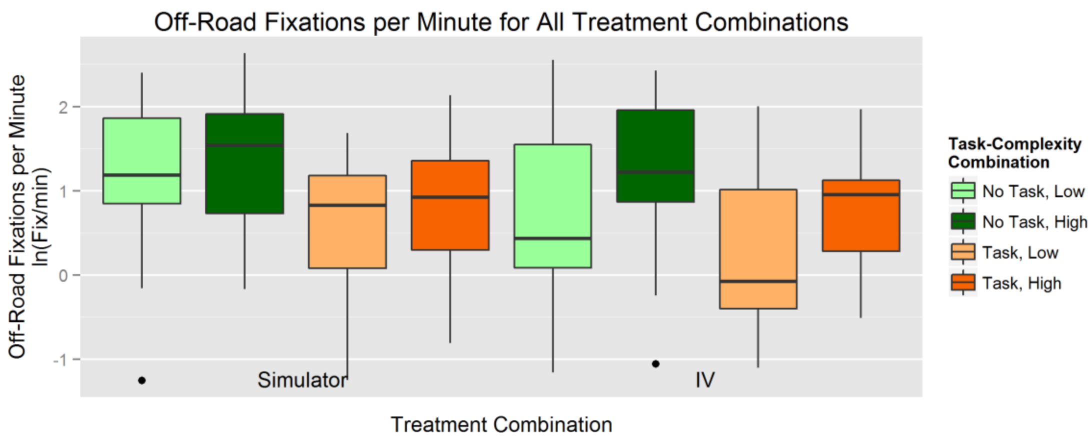

Figure 33. Off-Road Fixation Frequency Boxplot for All Treatment Combinations

To explain the interaction effect, a plot below shows the raw gaze data plotted for both the simulator and instrumented vehicle drivers (Figure 34). In both the simulator and IV drives, dispersion is larger when there is no secondary task happening; however, this effect is larger in the simulator. The spread of the raw gaze data shows this effect below in Figure 34.

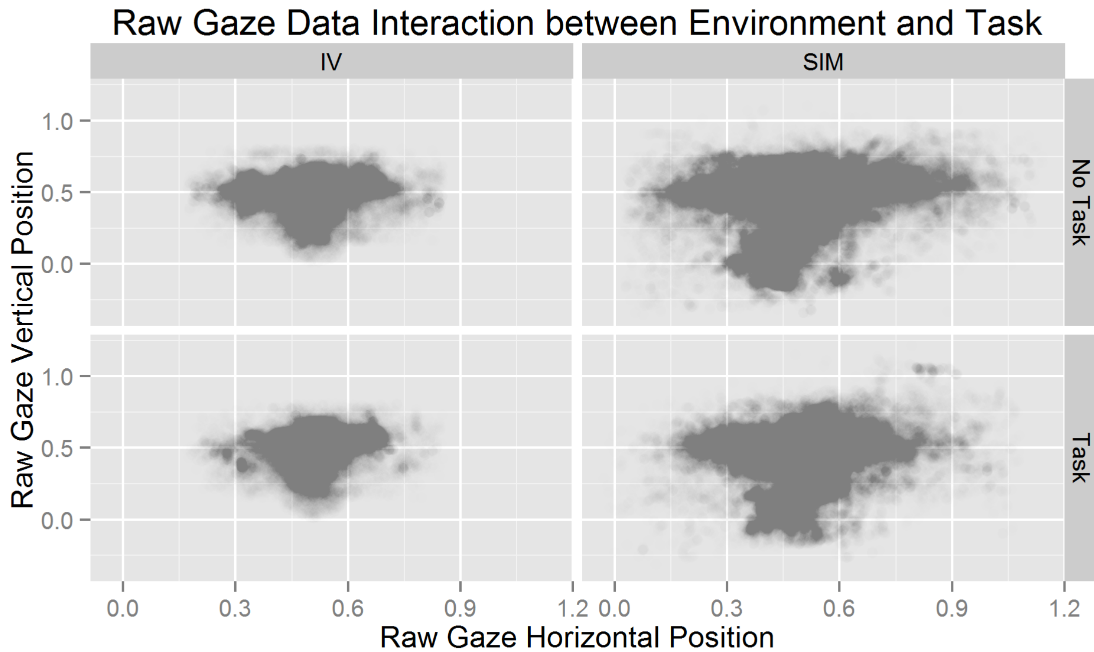

Figure 34. Raw Gaze Data - Interaction Effects between Environment and Task

Predicting Real World Behavior in a Simulator

Overall, the analysis for how each driver behaves in the simulator compared to the instrumented vehicle was not expected to follow the same trends across all individual drivers. One set of drivers may adopt passive coping while a second may lean toward strain coping. In each set, we'd expect to see different trends in the direction of physiological responses. For strain coping, the physiological signals would indicate increased mental workload, while performance metrics from both vehicle handling and eye tracking metrics should remain relatively stable. Alternatively for passive coping, we expect to see minimal change in physiological variables paired with a noticeable decrease in performance goals (e.g., SDLP, gaze dispersion, etc.).

The objective of this dissertation is to study whether an individual exhibits the same trends (in whatever direction) in the driving simulator as they exhibit while driving in the real world.

Nominal Categorization using Principle Components

Because there is no defined level of performance that can be referenced to categorize a driver's compensatory mechanism, a method was developed to facilitate an objective means of classification. Driver data from the instrumented vehicle were assessed using principle components analysis. Data from four groups of variables (physiological, performance, gaze-related, and self-reported measures of workload) were assessed separately. The first principle component (PC1) describing each group of variables was plotted against one another, to look for clusters of behaviorally similar individuals. A cluster analysis program was run on these plots, and these clusters were used to classify each participant's real-world high-mental workload compensation methods.

Physiological Principle Components Analysis. Several observations were omitted from this analysis. From the total of 105 separate drives, observations that had a mean EDA value larger than 27 were omitted (6 observations); from individual inspection, these large values were due to large movement artifacts during data collection. Because direction of travel and time of day have a significant impact on pupil diameter, all observations during which the participant was driving into the sun were omitted (37 omitted). Relative pupil diameters larger than 150% were omitted (3 omitted), as this was found to be due to a relative calculation that was impacted by an initial value where the driver was looking into the sun. Pupil diameters smaller than 0.008m were also omitted (16 omitted), found when participants were facing other light-related artifacts. Figure 35 shows the Scree plot and Pairs plots for the physiological data, to show the variance distribution and various relationships between variables.

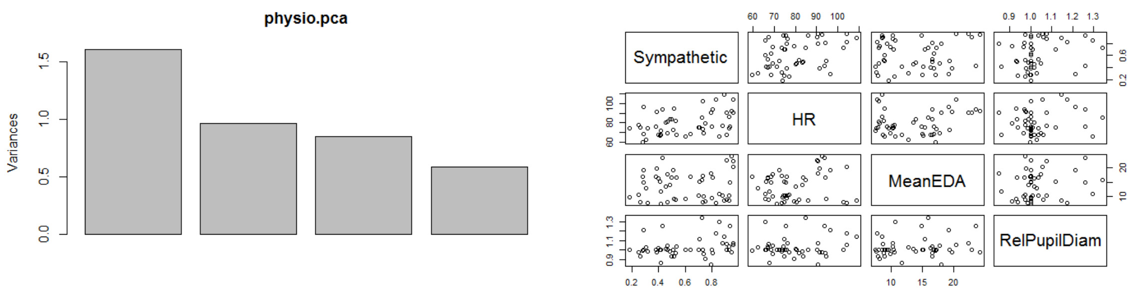

Figure 35. Scree (left) and Pairs (right) plots for Physio Variables in IV.

The rotational loadings are shown below in Table 12. Looking at the first principle component, all of the variables have the same sign—this represents that all variables are increasing (or decreasing) in tandem with each other. This is expected, as the same underlying mechanism controls all of these physiological components during increased mental effort.

Table 12. Rotations for Physio Variable PCA

	PC1	**PC2**	**PC3**	**PC4**
Sympathetic	0.5787739	-0.4338563	-0.06654387	-0.6872856
HR	0.5819383	-0.1689675	-0.46947473	0.6421770
EDA	0.3294353	0.8752638	-0.24986498	-0.2509046
Pupil Diam	0.4667344	0.1308893	0.84423509	0.2286791

Performance Principle Components Analysis. No observations were omitted from this data set. Lane position was not included in this analysis, due to a high level of data loss due to shadows compromising the lane position signal. A log transform was applied to two variables that were found to be non-normal (steering reversal magnitude, and standard deviation of steering angle), and PCA was performed on the transformed variable. Figure 36 shows the Scree plot and Pairs plots for the performance data, to show the variance distribution and relationships between variables.

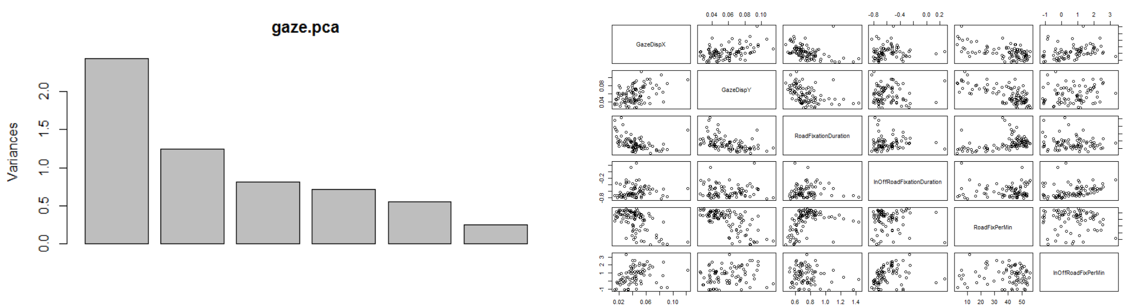

Figure 36. Scree (left) and Pairs (right) plots for Gaze Variables in IV

The rotational loadings are shown below in Table 14. The first PC tells the story we expect to see with increased mental workload. As both vertical and horizontal dispersion decrease, on-road fixations are becoming more frequent and longer in duration, while off-road fixations are becoming less frequent. This behavior is accepted in literature as what traditionally happens as drivers become more mentally loaded: gazes tend to become more concentrated (“tunnel vision”), while attention is directed along the roadway.

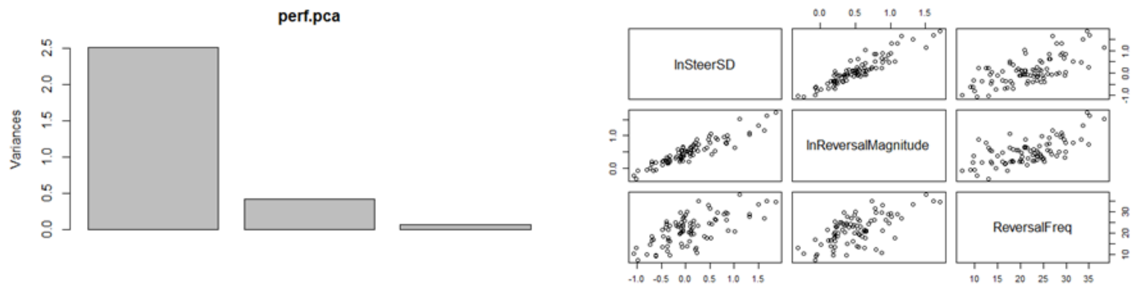

Figure 37. Scree (left) and Pairs (right) plots for Driving Performance Variables in IV.

The rotational loadings are shown below in Table 13. Looking at the first principle component, all of the variables have the same sign—this represents that all variables are increasing (or decreasing) in tandem with each other. Here we see that the standard deviation of steering angle increases, as the frequency of steering reversals and their magnitude increases.

Table 13. Rotations for Performance Variable PCA

	PC1	**PC2**	**PC3**
SteerSD	-0.6027949	0.3489459	0.7175479
Reversal Magnitude	-0.5985719	0.3968735	-0.6958470
Reversal Frequency	-0.5275887	-0.8489571	-0.0303641

Gaze Principle Components Analysis. No observations were omitted from this data set, as the variables that were assessed here were not influenced by any identifiable factors that would affect the various gaze variables. A log transform was applied to two variables that were found to be non-normal, and PCA was performed on the transformed variable. Figure 36 shows the Scree plot and Pairs plots for the NASA-TLX data, to show the variance distribution and various relationships between variables.

Table 14. Rotations for Gaze Variable PCA

	PC1	PC2	PC3	PC4	PC5	PC6
Gaze Dispersion (X)	-0.4969	0.0736	-0.0015	0.2888	0.7413	0.3388
Gaze Dispersion (Y)	-0.5250	-0.1801	-0.1336	-0.0030	-0.5928	0.5681
On-Road Fixation Duration	0.3901	0.0277	-0.6124	0.6657	-0.0721	0.1535
Off-Road Fixation Duration	-0.0388	0.7102	-0.4755	-0.4842	0.0399	0.1120
On-Road Fixation Frequency	0.5192	0.2089	0.4602	-0.0445	0.0316	0.6870
Off-road Fixation Frequency	-0.2330	0.6429	0.3870	0.4868	-0.3022	-0.2332

NASA-TLX Principle Components Analysis. No observations were omitted from this data set, as the variables that were assessed here were not influenced by any identifiable factors that would affect the various gaze variables. Figure 38 shows the Scree plot and Pairs plots for the NASA-TLX data, to show the variance distribution and various relationships between variables.

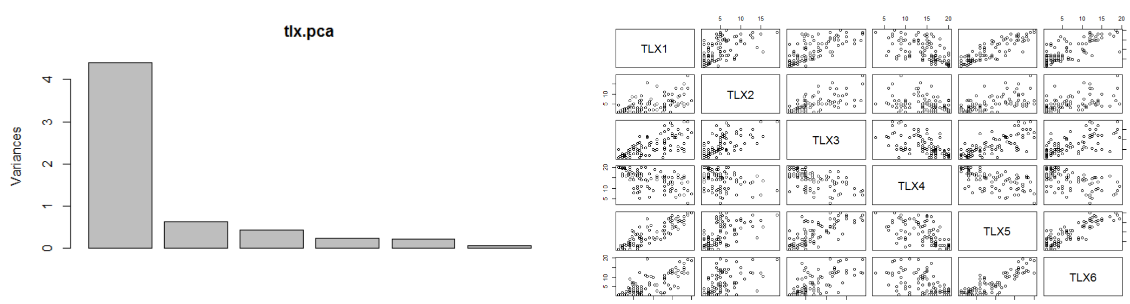

Figure 38. Scree (left) and Pairs (right) plots for Gaze Variables in IV

The rotations for the PCA are shown below in Table 15. The first component fairly overwhelmingly covers the majority of all the variation in the dataset; closer inspection of this component shows expected relationships between the variables as mental demand increases. The variables all increase in the same direction, with the exception of TLX4, which captures how the drivers think they performed. This makes sense thinking about how high performance would correlate with low mental workload, and vice versa.

Table 15. Rotations for NASA-TLX Variable PCA

	PC1	PC2	PC3	PC4	PC5	PC6
Mental Demand	-0.4254	0.0308	-0.4520	0.1325	-0.6585	0.4032
Physical Demand	-0.3562	-0.7186	0.4080	0.4359	-0.0035	-0.0136
Temporal Demand	-0.4206	-0.1444	0.2823	-0.8459	-0.0784	0.0334
Performance	0.3526	-0.6706	-0.5906	-0.2658	0.0608	-0.0532
Effort	-0.4521	0.0928	-0.3271	0.0645	0.0309	-0.8215
Frustration	-0.4318	0.0590	-0.3061	0.0491	0.7454	0.3980

Combining PC1s for Cluster Analysis.

Cluster analysis was performed twice in order to create nominal levels from the clusters for the pending logistic models which were calculated twice: once including the gaze parameters and again excluding gaze parameters to see the effect that gaze variables contribute toward accuracy of the model. Minimum cluster size was manipulated until the output was divided into a minimum of three clusters; three was selected as the most likely to mirror the theorized “types” of individuals suggested by Hockey (1997).

Cluster Analysis Including Gaze Variables

Looking at the dendrogram and cluster pairwise comparisons, the two clusters (minimum cluster size cut off at 7 members), the cluster characteristics can be better categorized with respect to each of the PC1 trends. The dendrogram numbers indicate the drive observation in the instrumented vehicle for all participants. Cluster One tends to have a higher performance PC1 (Low steering standard deviation, fewer steering reversals, and smaller reversals), lower TLX PC1 (higher self-reported mental demand, alongside lower self-reported performance), highly variable Physio PC1, and relatively neutral gaze PC1. The second cluster shows a predominantly lower Physio PC1 (higher heart rate, HRV, EDA, pupil diameter) with higher TLX PC1 (lower self-reported mental demand with higher self-reported performance), neutral performance PC1, and highly variable gaze PC1. In Figure 39 below, the numbers at the bottom of the dendrogram represent unique participant and drive ID information; each number is one participant drive in the driving simulator. A minimum cluster size was set to seven; which dictated the dendrogram cluster height to 7.81 using the Hclust package in R. At that specified cut height, the observations split into three clusters, shown in Figure 40.

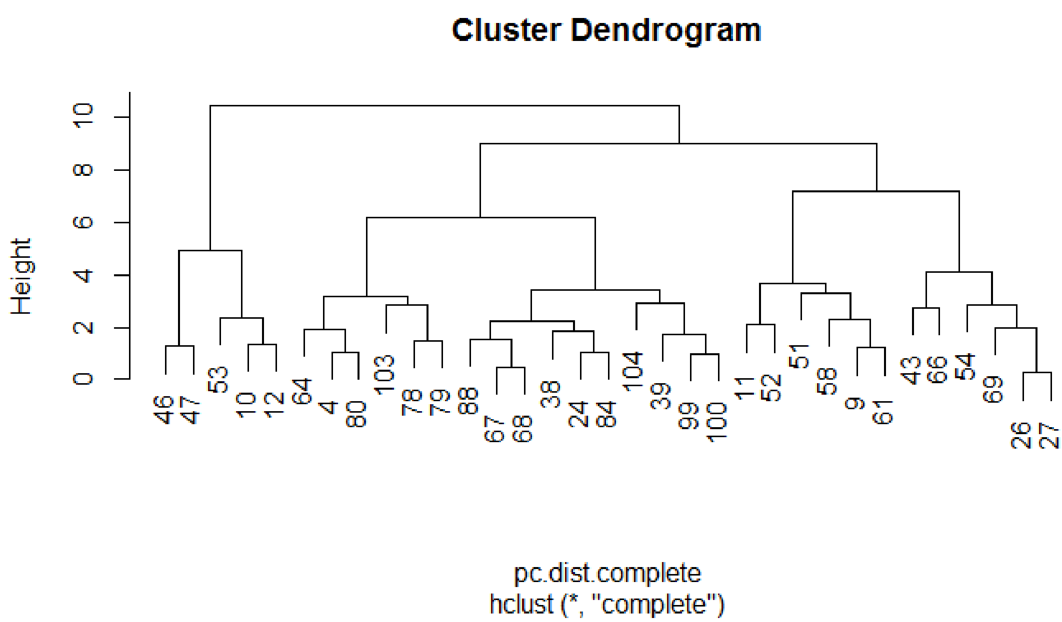

Figure 39. Cluster Dendrogram Output

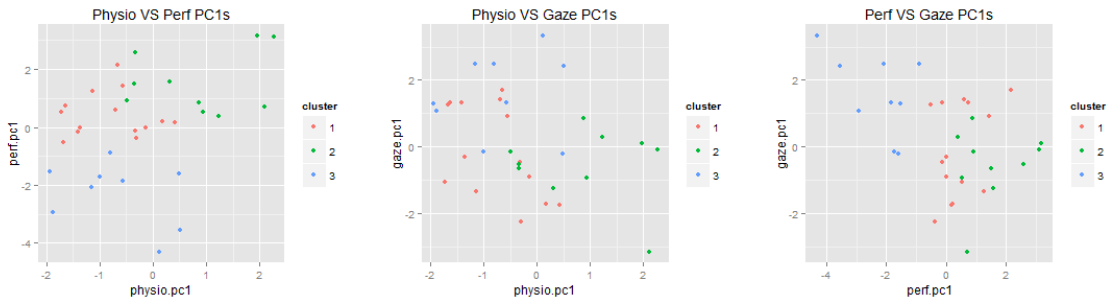

Figure 40. Physio, Performance, and Gaze Cluster Scatter Plots

Table 16. Hclust Descriptions from PC1 Rotations

	Cluster ID		
	1	**2**	**3 “Passive”**
Gaze PC1	No observable trend	Lower gaze PC1	Higher gaze PC1
Perf PC1	Higher and near 0 PC1	Higher PC1	Lower PC1
Physio PC1	Lower PC1	Lower PC1	Lower PC1

If these clusters were based only on the interaction between whether individuals were “active” or “passive” copers when subjected to excessive mental workload, we would expect to see different clusters. If the main divisive factor influencing how the clusters were split gaze behavior, performance, and physiology were based on this division, we would expect to see a cluster of neutral points during low mental workload, and then two other distinct clusters: high physiology PC1 and neutral performance PC1 as the first, with a second cluster near neutral physiology PC1 and low performance PC1. Clearly this is not the case here, so a different classification metric is necessary.

Secondary Approach: Nominal Group Assignment

In order to objectively create nominal categories for the multinomial logistic regression portion of analysis in lieu of PCA, cluster analysis was used to identify groups of driver response patterns relative to increasing mental workload.

First, the data were restructured within each environment so that every difference in driving task and complexity combinations were evaluated with reference to each other. In this way, new metrics were evaluated describing the direction of change in dependent variable categories between the low complexity/no task and the high complexity/high task treatment combinations, for each participant completing the instrumented vehicle portion of the study.

The variables that were examined for this classification schema include one dependent variable from each category studied in the multivariate analyses. The variable with the most noticeable effect when mental workload was applied was selected: heart rate (physiology), steering reversal frequency (performance). The change in heart rate was examined as a percentage, since heart rate is very individual-specific. The low-complexity and no-task treatment combination was used as the baseline level for all relative variables for both environments, separately. The reversal frequency was examined as discrete difference in the number of reversals per minute.

The performance and physiology response variables were rated either “neutral” or “heightened” in response to applied mental workload between the very low and very high conditions. This resulted in four total groups of outcomes: Neutral response, high physiological response, high performance response, and high for both physiological and performance response.

Multinomial Logistic Regression

To study whether these responses are similar, multinomial logistic regression was used to model the real-world driver responses based on simulated driving responses. The outcome categories were based in the different compensatory techniques (“Passive Coping”, “Strain Coping”, or “No Observed Change”) experienced by the drivers under heavy mental stress. The quantitative predictor variables are the percentage of change that the drivers experience in the driving simulator as mental workload is increased between the baseline (low complexity) and highest level of complexity. Each of the four response categories use one representative variable for each category: heart rate predicts physiological response, SDLP predicts driving performance goals, and gaze concentration predicts visual behavior performance. The positive direction for each change is the anticipated direction of change given increasing mental workload (increasing heart rate, increasing SDLP, increasing gaze concentration). Two level interaction effects are also explored.

$$log\left(\frac{\pi_{2i}}{\pi_{1i}}\right) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_{12} x_{i1} x_{i2} + \beta_{13} x_{i1} x_{i3} + \beta_{23} x_{i2} x_{i3} \quad \text{Eq. 1}$$

$$log\left(\frac{\pi_{3i}}{\pi_{1i}}\right) = \alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \alpha_3 x_{i3} + \alpha_{12} x_{i1} x_{i2} + \alpha_{13} x_{i1} x_{i3} + \alpha_{23} x_{i2} x_{i3} \quad \text{Eq. 2}$$

Where:

π_1, π_2, and π_3 represent the probability of the drivers utilizing each specific compensatory mechanism in real world driving (1="No observed change", 2="Passive coping," and 3="Strain Coping", see Table 17 for complete descriptions);

x_{ij} is the j^{th} quantitative explanatory variable measured in the simulator (Physiological, Driving Performance, Visual Performance) for the i^{th} participant;

β_j is a vector of regression coefficients corresponding to the outcome of "Passive Coping" relative to no observed change for simulator predictor *j*;

α_j is a vector of regression coefficients corresponding to the outcome of "Strain Coping" in the instrumented vehicle (relative to no change) for the simulator predictor *j*.

Table 17. Nominal levels in Multinomial Logit Models

Variable	**Outcome Nominal level**	**Response associated with increasing mental workload**
Compensatory Mechanism	"No Observed Change"	No change in physiological AND no change in performance dependent variables
	"Passive Coping"	No change in physiological variables; reduced performance goals observed leading to subsequent increased SDLP, increased steering heading deviation, gaze concentration,
	"Strain Coping"	Increased heart rate, EDA; no observed change in performance goals (SDLP, steering heading deviation, gaze concentration, glances off-road)

The initial goal of this assessment was to observe and model driver behavior, specifically looking at how physiology and performance change as mental workload increases. An initial multinomial model was attempted, but not possible due to the very low number of drivers who met the “Passive Coping” criteria. Closer inspection comparing the changes in driver response between treatment levels shows some interesting patterns that are different between the two environments.

The next few plots are shown to illustrate which variables show recognizable patterns as an explanation for the missing multinomial analysis. Interpretation of the figures is slightly complex. There are four potential treatment combinations: low complexity with no task, low complexity with a task, high complexity with no task, and high complexity with a task. There is a clear hierarchy of “difficulty” to completing these tasks, with one exception: drivers had mixed feelings about the comparisons between driving the low complexity road while completing a task and driving the high-complexity road with no task. All other direct comparisons can easily be ordered so that there is a task requiring additional workload, and one with less. All variables were measured to compare the proportional response for each step in difficulty—the response to the combination requiring more mental capacity was divided by the same response for the easier task to give a proportion of driver response as difficulty increases in the instrumented vehicle. The same operation was applied toward the simulator, and these responses were plotted against each other to show the relationships between them.

An absolutely validated system would appear on this type of plot as a straight line with a slope of one and an intercept at zero; if that were the case, drivers would be experiencing the same relative increase in a variable of interest, and the same absolute measure of that variable. A system with relative validation would appear as a line with a similarly-signed slope (positive or negative). By way of explanation, the physiological response variable plots are below.

Both Heart Rate and the Sympathetic response show a positive trend, which is expected. However, looking at the chart for pupil diameter, there does not appear to be any real observable trend that would suggest validity between the simulator and real world. This is perhaps nearly entirely explainable by the dramatically different settings which have a direct effect on pupil diameter: the driving simulator has a dark environment with no light ever shining toward or near the participant. By necessity, the real world drives were conducted entirely outdoors and were subject to external lighting conditions. Direct sunlight was a factor, along with approaching vehicle headlights, and glare from polished surfaces. Not surprisingly, a broad range of pupil response was observed. Looking only at response changes where the sun was not a factor, no clear pattern emerges. The change in EDA follows a similar lack of pattern; a cluster of data points shows that there were not many detectable EDA changes in either environment.

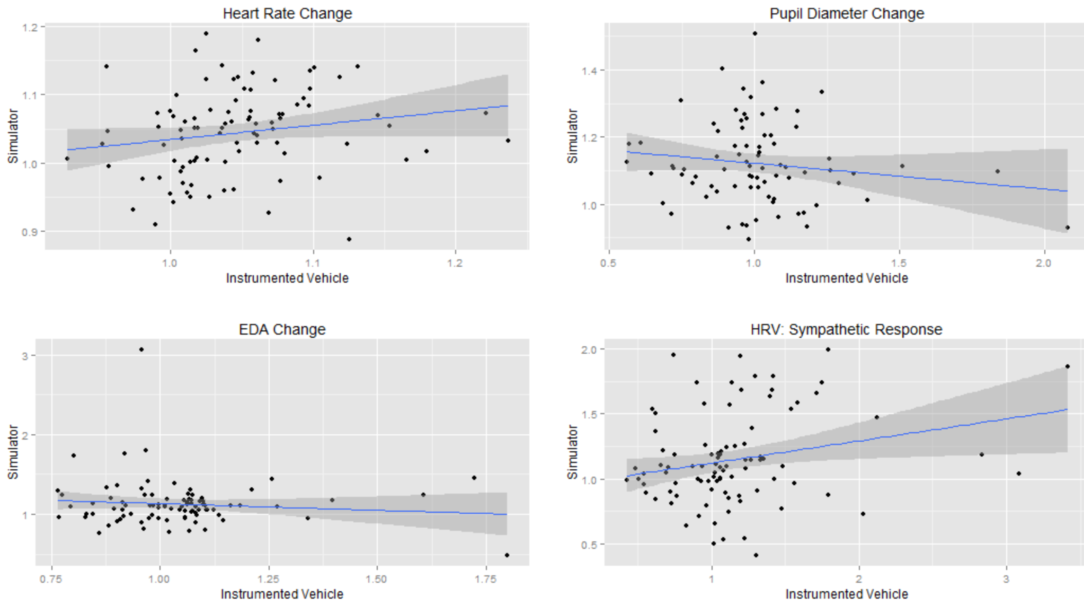

Figure 41. Proportional physiological changes for increases in mental workload between SIM and IV

Similar plots are shown below, for driving performance variables. Steering reversal frequency has the closest relationship between the simulator and the real world, but even that relationship is tenuous at best. This is a reasonable explanation for the failure of logistic regression models to create passable representations of real world driver behavior from driving simulator behavior: that relationship simply does not exist in a capacity that can be modeled directly between the two environments, at least for driver performance variables. This is not to say that the driver performance output from the simulator is bad data, but rather that an absolute, or even relative validity does not exist between the two driving systems as far as solely performance data can show.

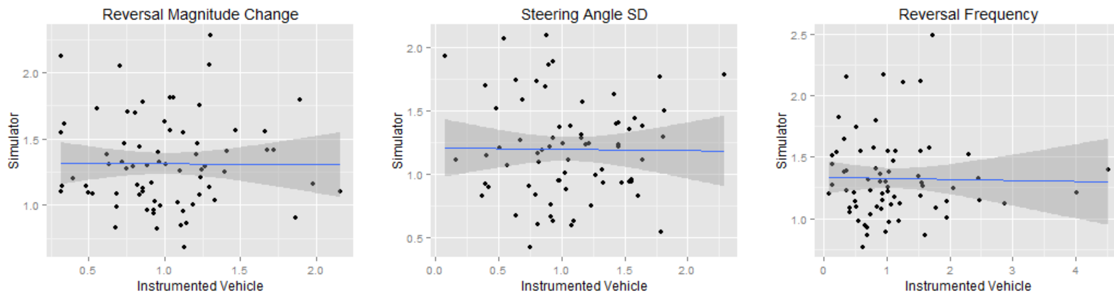

Figure 42. Proportional performance changes for increases in mental workload between SIM and IV

Driver gaze behavior plots show similar patterns. The variable showing the best relationship between the simulator and the instrumented vehicle is horizontal gaze dispersion—While outliers are skewing the scale at the ends, it is important to note that as mental workload increases gaze dispersion *decreases*, so the clustering around the origin is reflective of a fairly representative relationship between the driving simulator and the real world in terms of gaze response during increasing mental workload. No real trend is visible for road fixation duration or frequency changes as mental workload increases. Off-road fixation durations are much longer in the real world than in the simulator for comparable shifts in mental workload.

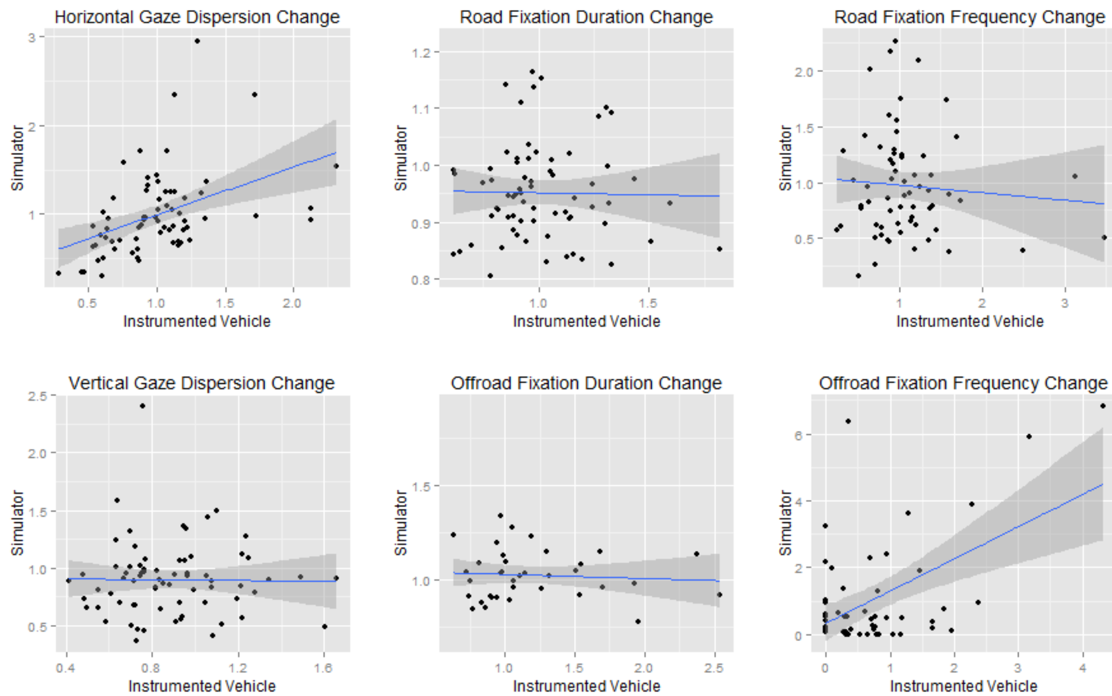

Figure 43. Proportional Gaze changes as mental workload increases between Sim and IV

The lack of relationship may be due in part to driver perception of workload. The next set of plots presents the same proportional response between the simulator and instrumented vehicle, detailing different subscales describing driver self-reported mental workload. These variables have the best validity between the driving simulator and driving in the real world. Mental demand, temporal demand, effort, and performance all show fairly consistent positive linear trends between the two environments, indicating that driver perception of applied workload shows at least relative validity as mental workload increases. Both frustration and physical demand show similar different trends: both have groups of drivers who either a) reported much higher frustration in the simulator than IV, or b) reported much higher frustration or physical demand in the IV than the simulator.

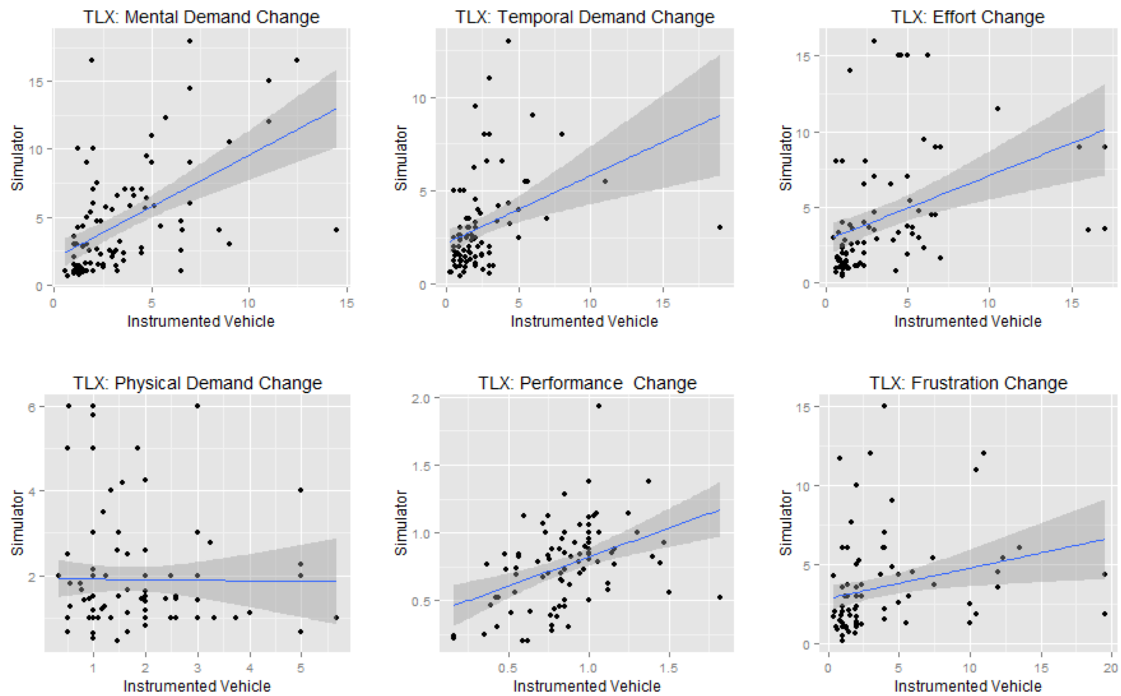

Figure 44. Proportional TLX changes as mental workload increases between Sim and IV

The relationship between performance and driver physiological response within each separate driving environment, in case the environments are eliciting different behavioral patterns. For these plots, steering reversal magnitude was used as the performance response as it appears to maintain the highest degree validity from the variables that were derived. These plots are interpreted differently than those preceding; while the earlier plots were a ratio of IV:Simulator response changes, these plots are a ratio of the performance to the physiology variables to look for relationships expected, given Hockey's theorized relationship between additional stress and coping mechanisms.

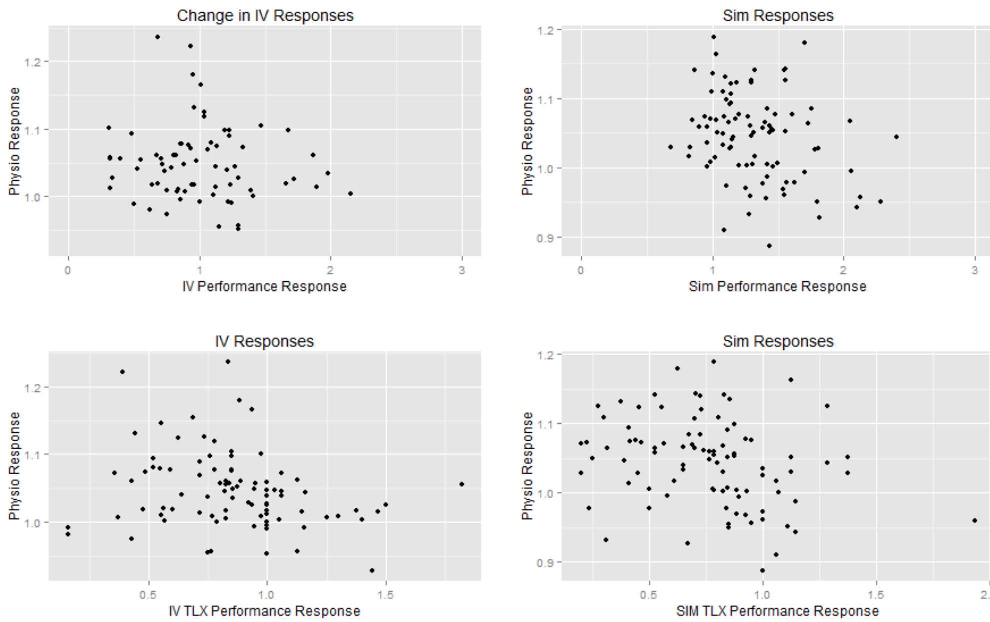

Figure 45. Proportional Changes in Driver Response: Heart Rate and Reversal Magnitude

The first noticeable aspect of these plots are that, while not identical, they do appear to following similar shapes, comparing between the objective steering performance and the TLX performance metric. The differences between the simulator and real vehicle responses are visible in both TLX and steering reversal methods of classifying driver performance. In this plot, we are not looking for positive linear trends—instead, if the relationship suggested by Hockey is visible, we would see clusters of variables following a vertical line at 'Performance = 1' and another at 'Physio = 1' to represent a driver whose physiological response is unchanged ($Performance_{High}$ / $Performance_{Low}$ and at the same time exhibiting a larger change in driving performance; or vice versa.

These clusters are visible in the instrumented vehicle for both the objective and self-reported measures of performance, but these clusters are not highly populated. These are one reason nominal logistic regression was not successful; the other reason can be seen in the simulator Performance vs. Physiology response plots on the right of Figure X. If the underlying relationship between physiology and performance were a direct relationship (i.e., if the a sympathetic physiological response triggered by excessive mental workload at the same time prompted an equivalent decrement in performance due to a lack of mental resources to dedicate to driving performance), we would anticipate a strong downward correlation: as heart rate rises due to mental workload, performance should drop. This is actually closer to what we are seeing in the driving simulator—while not a definitive trend by any means, the expected clusters along ‘physio = 1’ and ‘performance = 1’ are not present.

Specific observations that elicit what is suspected to be a coping strategy in the instrumented vehicle have been flagged in Figure . Looking to the right sections of the figure, there is no obvious method or metric that can be used to identify the same trend occurring in the simulator. Specifically, the points that were flagged from the IV data are so randomly scattered throughout the ‘neutral’ data, there does not appear to be a simple metric that can be used to predict “coping mechanism” style in the real world, given purely simulator data using objective or survey-based performance data collection methods.

Figure 46. Performance vs Physio Plots, with Copers Flagged in IV and SIM

One hypothesis of this thesis is that gaze performance may be an important aspect of driver performance, respective of Hockey's theorized compensatory mechanisms. Similar plots were created, looking at different gaze-related performance variables. Figure 47 shows the relationship between gaze dispersion and heart rate, for the IV (left) and simulator (right). For both horizontal and vertical gaze dispersion, increased mental workload is expected to concentrate the gaze (think "tunnel vision" under stress), so the clustering of data points slightly to the left of the "no change in gaze dispersion" line at 1.0 is expected, as it is a natural decrease as mental load increases. That being said, there are similarities between the shapes of these plots and the earlier IV-performance and TLX performance-physiology plots. However, the simulator plots resemble the same shapes.

Figure 47. Change in Gaze Dispersion and HR with increased workload: SIM and IV

Further examination of the plots and manual flagging of potential “copers” shows that the flags identified from the IV data do not correspond with the areas of the plot that would represent the same types of coping, in the simulator. There are a few points from “Passive” coping that wind up near the correct position, but are random in their scatter. Like with the Figure x, applying the Gaze X flags to the Gaze Y data (and vice versa) has no noticeable effect for the Passive coping points, but the Strain coping points generally maintain their original position on the alternative plots.

Similar plot patterns emerge looking at the change in gaze fixation duration as the variable of interest for performance; again plotted against heart rate. For both on-road and off-road fixation durations, the manually selected Strain gaze points map back and forth. In the same vein as previous plots, trying to predict compensatory mechanism style beginning with the simulator data is not likely.

Figure 48. Change in Fixation Duration with increased Workload: Sim and IV

The last group of gaze variables evaluated were the fixation frequency (on- and off-road) changes as additional mental workload was applied. Like with the other gaze variables, the flagged observations do not show any real pattern when applied to the simulator data. When looking at the differences between on-road and off-road fixation the Strain coping points map back and forth among the IV plots very well; and the Passive copers do not reflect the same ease of identification on the opposite-fixation data set.

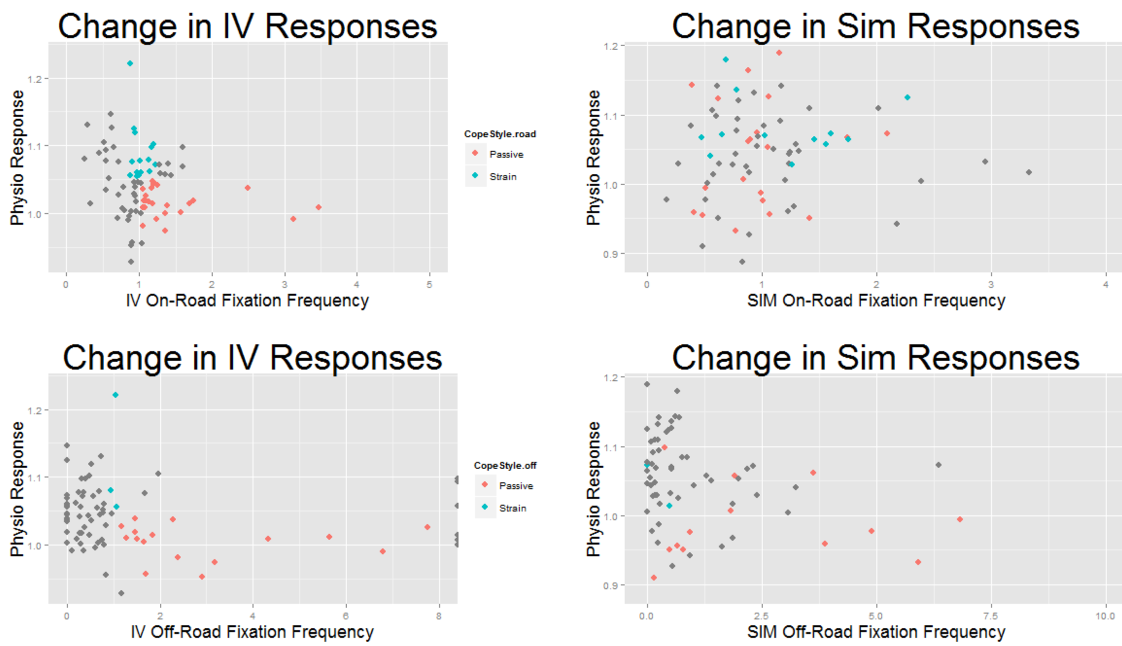

Figure 49. Fixation Frequency and Physio Response, Copers Flagged

DISCUSSION

This study has identified some very interesting aspects of the physiology-performance relationship, through its exploration of multivariate responses as well as the differential analysis comparing changes in driver behavior between different treatment combinations of applied mental workload. A summary of key findings follows:

- Several variables were found to be good detectors for differences in applied cognitive workload: heart rate, HRV, controlled pupil diameter, steering reversal frequency, gaze dispersion, on-road fixation duration and frequency, and self-reported NASA-TLX data
- There is no reciprocal relationship between the simulator and instrumented vehicle for nearly all of the physiological and performance relationships as workload increases
- NASA-TLX and driving performance metrics are consistent across environments for “strain copers”
- A reciprocal relationship does exist between simulator and on-road environments for off-road gaze fixation frequency, on-road fixation frequency, and horizontal gaze dispersion
- Relative validity shown between the simulator and instrumented vehicle for heart rate, heart rate variability, horizontal gaze dispersion, off-road fixation frequencies, and most NASA-TLX subscales
- No absolute validity was demonstrated for any study variables

MANOVA Variable Selection

Multivariate analyses performed on simulator variables here showed driver trends that were generally expected, given past literature on human response to workload. Of the physiological variables that were examined, only the mean electrodermal response was not significant for any of the independent variables. The non-significance of a variable known to detect effects under applied workload was due to 1) the long-term nature of each experimental session, and 2) unavoidable motion artifacts from driving disrupting the signal. The advantages of EDA lie with its short-term detection abilities; when used over longer intervals, data processing involves work compensating for irregular EDA signal drift. The processing done on this data was insufficient to detect the real EDA signal and compare it across treatment combinations. The EDA leads were also attached to the driver's left hand, with cabling running down their wrist. While this was appropriate while the drivers were not turning or moving, it did introduce a high level of motion artifacts from standard driver movements (scratching face, waving on pedestrians), which were encountered in both the simulator and instrumented vehicle. Pupil diameter would have been the best metric to measure mental workload, were it not for the constantly changing light levels that were encountered in the real world. The pupil diameter in the dark simulator lab was sufficient to detect differences in task and complexity; but this was lost for the IV drives only. If this study could have been redesigned, North-South routes might have been more appropriate in lieu of losing all morning and evening data to sun-related glare. The broadest application of the data loss relative to pupil diameter for diagnosing mental workload is with unobtrusive driver assistive systems using eye tracking technology to alert drivers regarding different hazards or conditions. If future systems were to rely solely on eye tracking, the signal may be so fraught with light-influenced artifacts to be unusable: instrumented vehicle systems may not be suitable to assess pupil-based workload measures due to light variability inherent in on-road driving. That being said, the simulator was a perfect application of using pupillometry to identify changes in driver workload. The sympathetic HRV component was a good diagnostic tool, but required a substantial amount of preparation before data collection could begin.

The vehicle handling driving performance metrics showed that the drivers with higher workload had more frequent reversals; however this data had reduced capacity to detect signals in the real car. The simulator data output a 60Hz signal that came straight from the vehicle computer, but the vehicle data has to be processed from a 30Hz camera pointing at the steering wheel. Collecting data from the camera instead of directly from the vehicle computer resulted in a massive amount of data loss; a more direct method of data capture would be preferable for other naturalistic data collection projects.

The TLX data were significant in nearly all categories; one telling difference was that the only significant differences detected between the simulator and the real world were in the “performance” and “effort” subscales. The interaction effect was interesting—all of the subscales saw a shift up during the high-complexity simulator drives while performing a secondary task. This interaction suggests that the combination of the high-complexity environment with the secondary task resulted in a substantially higher amount of workload on the drivers. Self-reported performance was not different for the drivers between the roadway complexity conditions—drivers may inappropriately judge themselves to have higher performance in more demanding driving configurations than is actually happening—which underscores the importance of objective performance data to support any findings identified through potentially biased self-reported data.

Gaze-related eye tracking variables are one part of the data associated with this project that were not generated as the drives occurred; but were rather derived following data-processing. Eye tracking metrics may not be a viable way to measure workload, but an understanding of how they change with applied workload will go far in different fields such as novice or commercial driver training. The variables here support literature finding that gaze dispersion tends to decrease with mental workload—whether from scanning side-to-side or looking forward down the roadway, when drivers are distracted they don't engage in scanning behaviors necessary for safe driving. Drivers tended to fixate on the road for longer and more frequently in the real world than they did in the simulator—one potential explanation could be the novelty of the simulator world to most drivers. Studies incorporating eye tracking into their study design as a measure of distraction or driver ability should keep this in mind while evaluating different tasks expected to distract the driver from eyes-on-road in the simulator; it is not a measure that has been absolutely validated. The act of performing a secondary task can be compared to the more frequently researched fields looking at testing and design of in-vehicle driver assistive systems. The data here show that simulators are a suitable proxy for detecting differences in workload due to these types of secondary tasks; perhaps providing a slightly conservative estimate of the effect of distraction, since these differences were not always detected in the on-road driving segments.

While completing secondary tasks, drivers tended to look off-road less often. One out-of-place finding relative to mental workload applied by the driving environment was that driver had fewer off-road fixations per minute on the low-complexity roadways. This may have been due in part to the very low levels of roadside objects and scenery present in this study, in both the real-world and simulator drives.

Another potentially contributing factor toward the interaction between the secondary task and environment in many variables is the importance that each individual driver assigned to the secondary task. Drivers in the simulator tended to fixate on the secondary task to a much higher degree than they did in the real world; simulator drivers ran red lights in five separate drives, only during secondary task performance—this was never observed in the real world. Mitigation of this effect in the driving simulator has not been explored. Potential approaches could include driver performance-based incentives. Other studies may consider a short interview with the participant after each drive, so that the relative importance or difficulty of the tasks could be recorded for each driver and taken into account in the statistical analysis. A second contributing factor toward differences in behavior between the simulator and instrumented vehicle could be due to ambient traffic—while the traffic levels were similar in both environments, ambient traffic on-road may apply an additional amount of workload due to the real-world risks and consequences of collision that are not present in the simulator.

Response Variable Rate of Change with Increasing Workload

The approach taken here to examine the rate of driver response variable change as mental workload increases was successful in assessing the likelihood of a model for prediction of specific driver behaviors at excessive mental workloads. While unfortunate that a predictive model was not possible here, the focus of this dissertation on different patterns and trends between simulators and instrumented vehicle driver performance and physiology did have many beneficial findings.

Figure 46 in the Analysis section illustrates how drivers who are believed to belong to “Passive” or “Active” coping categories in the real world cannot be mapped to that position starting from simulator data. The same is also true trying to predict “Passive” copers in TLX data by using objective performance data (and vice versa). However, the “Strain” copers (shown in blue) nearly perfectly map back and forth between both methods of measuring driver performance. Figure 50 below shows the IV plots, with the identified “coping” driver flags created from the reversal magnitude performance data, but applied on the TLX performance plot. This suggests that we are seeing drivers with an actual “Strain Coping” style of handling increases in mental workload, where the drivers’ reaction to the applied workload is largely physiological in nature.

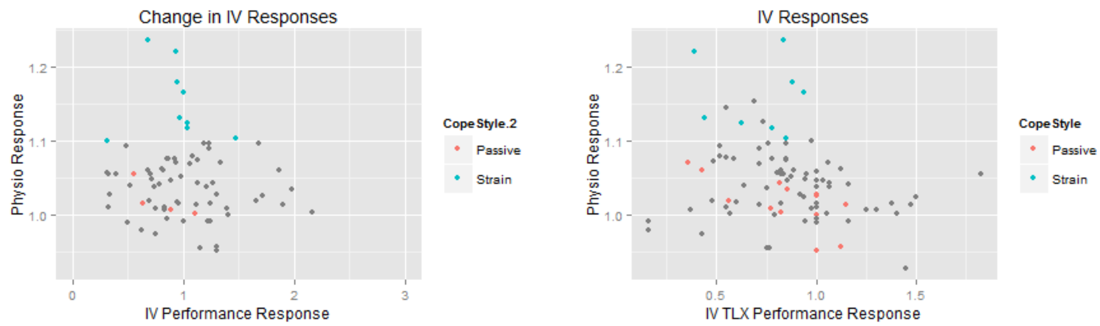

Figure 50. IV Performance and Physiological Change; Coping Flags from other variable

The same scatter distribution happens repeatedly when looking at the instrumented vehicle data: a set cluster of random distribution around the “origin” which corresponds to no change in either the performance or physiological metric, with cluster “arms” reaching out along each neutral axis. Because the shape consistently appears across different performance-based metrics, there is merit to the idea that the relationship between performance and physiology relative to applied cognitive workload is real. What warrants further study is the non-reproducing lower “arm,” where the anticipated passive coper cluster should lie. In contrast, the points that are believed to be Hockey’s strain coping points are still strain points across multiple variables; lending credence to the theory that strain coping is a visible type of compensatory mechanism for increasing mental workload. Unfortunately, this pattern is not seen in the driving simulator. One hypothesis is that the amount of risk presented by the driving simulator is so low, that drivers are not forced into the automatic feedback loop that coerces either a strain or passive coping strategy. Driver behavior was observably different between the two environments; multiple drivers became so distracted by the secondary task in the simulator that they altogether failed to notice traffic control; this never happened in the instrumented vehicle. Whether or not risk is the cause, drivers subsequently seemed to place higher emphasis on the secondary task while in the simulator than they did in the instrumented vehicle; this explains the lack of visible compensatory strategy clusters in the simulator data.

CONCLUSION

This study initially set out to explore simulator validation, in hopes of finding both independent and dependent variables that could be used to reproduce realistic driver behavior from the real world in a simulator. The nature of the validity issue was clear in this study; at the mental workload levels, validity at high workload was not seen in the simulator, reflective of real world behavior. Fortunately, a host of dependent variables were collected and described, for researchers to use as a reference when conducting future work in heavy workload environments. This thesis established the relationships between physiological response during increasing workload and driver performance; both gaze- and vehicle handling-related. The findings here support past work proposing specific passive and strain compensatory mechanisms, but future work will need to apply additional cognitive loading to achieve the same findings in the simulator. Currently, simulator-elicited driver behavior still has no specific realistic method that can be used to adequately model real world driving. However, relative validity will serve the research and education community—driving simulators are still a valuable research tool, and merit further study so that an accurate model of driver behavior can be built.

The eye tracking variable findings demonstrate that there is a set need for eye tracking data collection in simulators, which has historically been limited. The on-road and off-road fixation frequencies were the only variables that demonstrated relative validity in the performance-physiology relationships. Since the same people are showing similar types of visual scanning behavior in these performance-physiology relationships, this lends additional foundation to research using eye tracking variables as a method of workload; which can be visible even when driver physiology has no apparent increase.

To summarize, 1) variables found to be adequate to detect differences in mental workload in different study environments were identified: heart rate, HRV, controlled pupil diameter, steering reversal frequency, gaze dispersion, road fixation duration and frequency, and self-reported NASA-TLX self-reported variables, 2) a non-reciprocal relationship was found between the simulator and IV for most study variables, indicating that the compensatory mechanisms visible in the simulator were generally not observed on-road, 3) relative validity between the simulator and instrumented vehicle was found in several of the dependent variables measured in this study: heart rate, HRV, horizontal gaze dispersion, and off-road gaze fixation frequency.

While a completely valid simulator system is not currently available to serve as a proxy for on-road driving, there is certainly still a place in the scientific community for simulator studies. Simulators provide a unique environment where study variables can be explicitly controlled and artificially induced; saving time and money over naturalistic studies. Since they are not a perfect proxy, it is reasonable to assume simulators' place in the research world is to be used alongside on-road and naturalistic studies, as a pilot environment prior to research moving up the ecological validity spectrum.

Future Benefits

The findings from this study are applicable towards the field of transportation; this study has the potential to benefit other domains beyond transportation.

Understanding how driver response variables change between simulators and real world driving is important because nearly all simulator studies are performed in order to have some type of real world application. Research needs to have a strong justification for variable selection to identify mental workload, understanding the relationship between driver physiology and performance in the real world can apply towards simulation is important.

The medical field has begun using virtual training techniques for complex training in procedures that may not otherwise be encountered outside of occasional unscheduled patient experiencing that specific illness. He et al. (2011) proposed a new surgery simulation tool designed to aid in training for knee arthroscopic surgery—the findings from this study show that drivers in similar settings (simulator versus on-road) may require additional elements in virtual training in order to reach the same levels of mental stimulation as encountered in the real world. If this can be explored in virtual worlds outside of transportation, then the results can be applied towards He et al.'s study, creating a starting point for effective virtual surgical simulation.

The medical field is not the only domain looking into virtual training—Poeschl and Doering (2012) are studying the use of virtual audiences as a tool in cognitive behavioral therapy for patients with a fear of public speaking. Virtual maintenance repair programs have been used to train workers (Li et al., 2012), training orientation and mobility in disabled users (Seki & Sato, 2011), and in combat training or combat stress resilience (Reitz & Seavey, 2012; Rizzo et al., 2012). This is not nearly an exhaustive list of the different fields that have keyed into virtual simulation, but shows the potential applications that the findings from this study may have in validating and promoting the use of virtual training tools. All of these tools have the same basic premise: educate and train humans in an environment that is cheaper and less risky than the real world, ideally up to the same standards as a real-world training program. The tools and findings from this thesis will serve to assist program managers and research teams in the future hoping to gather similar metrics in developing virtual education to the workforce.

REFERENCES CITED

Åkerstedt, T., & Gillberg, M. (1990). Subjective and objective sleepiness in the active individual. *International Journal of Neuroscience*, *52*(1-2), 29-37.

Akinwuntan, A.E., Weerdt, W.D., Feys, H., Pauwels, J., Baten, G., Arno, P., et al. (2005). Effect of simulator training on driving after stroke: A randomized controlled trial. Neurology, 65, 843-850.

Allen, R., Rosentha,, T. & Cook, M. (2011). A Short History of Driving Simulation. In D. Fisher, M. Rizzo, J. Caird& J. Lee (Eds.), *Handbook of Driving Simulation for Engineering, Medicine, and Psychology* (2.1-16). Boca Raton, FL: CRC Press.

Allen, T. A. & Tarr, R. W. (2005).Driving Simulators for Commercial Truck Drivers – Humans in the Loop. Proceedings of the Third International Driving Symposium on Human Factors in Driver Assessment Training and Vehicle Design, p1-7.

Anderson, G.J. (2011). Sensory and Perceptual Factors in the Design of Driving Simulation Displays. In D. Fisher, M. Rizzo, J. Caird & J. Lee (Eds.), *Handbook of Driving Simulation for Engineering, Medicine, and Psychology* (8.1-11). Boca Raton, FL: CRC Press.

Arnedt, J. Todd, et al. "How do prolonged wakefulness and alcohol compare in the decrements they produce on a simulated driving task? " *Accident Analysis & Prevention* 33.3 (2001): 337-344.

Balk, S. A., Bertola, A., & Inman, V. W. (2013). Simulator Sickness Questionnaire: Twenty Years Later. In *7th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*.

Beck, J.G., Palyo, S.A., Winer, E.H., Schwagler, B.E. & Ang, E.J. (2007). Virtual reality exposure therapy for PTSD symptoms after a road accident: An uncontrolled case series. Behavior Therapy, 38, 39-48.

Bedard, M. (2008). Driving Simulators; Are they useful tools? Paper presented at the STISIM users group meeting, Quebec City, Canada.

Bedard, M., Dahlquist, M., Parkkari, M., Riendau, J. & Weaver, B. (2010). Assessment of driving performance using a simulator protocol: validity and reproducibility. *American Journal of Occupational Therapy*, 64, 2, 336.

Bedard, M., Parkkari, M., Weaver, B., Riendeau, J. &Dahlquist, M. (2010). Assessment of driving performance using a simulator protocol: Validity and reproducibility. American Journal of Occupational Therapy, 64, 336-340.

Beirness, D.J. (1995).The relationship between lifestyle factors and collisions involving young drivers. In *New to the Road: Reducing the Risks for Young Motorists International Symposium*, 8–11 June, p. 17. UCLA Brain Information Service/Brain Research Institute, Los Angeles.

Bella, F. (2005).Validation of a driving simulator for work zone design. Transportation Research Record, 1937, 136-144.

Bella, F. (2008).Driving simulator for speed research on two-lane rural roads. Accident Analysis & Prevention, 40, 1078-1087.

Bieliauskas, L. A. (2005). Neuropsychological assessment of geriatric driving competence. Brain Injury, 19, 221-226.

Bittner , A.C., Simsek, O., Levison, W.H. & Campbell, J.L. (2002). On-road versus simulator data in driver model development: Driver performance model experience. Transportation Research Record, 1803, 38-44.

Blaauw, G.J. (1982). Driving experience and task demands in simulator and instrumented car: A validation study. Human Factors, 24, 473-486.

Bradley, M. M., & Lang, P. J. (2000). Emotion and motivation. *Handbook of psychophysiology*, *2*, 602-642.

Bridger, R. (2003). *Introduction to ergonomics*. CRC Press.

Burnett, G. E. (2011). On-the-Move and in Your Car: An Overview of HCI Issues. *Human-Computer Interaction and Innovation in Handheld, Mobile and Wearable Technologies*, 60.

Caldwell Jr, J. A., Caldwell, J. L., Brown, D. L., & Smith, J. K. (2004). The Effects of 37 Hours of Continuous Wakefulness on the Physiological Arousal, Cognitive Performance, Self-Reported Mood, and Simulator Flight Performance of F-117A Pilots. *Military Psychology*, *16*(3), 163.

Cantin, V., Lavalliere, M., Simoneau, M. & Teasdale, N. (2009). Mental workload when driving in a simulator: Effects of age and driving complexity. *Accident Analysis and Prevention*, vol. 41, No. 4, p763-771.

Carroz, A., Comte, P. A., Nicolo, D., Deriaz, O., & Vuadens, P. (2008).Interet du simulateur de conduite pour la reprise de la conduite automobile en situation de handicap [Relevance of a driving simulator in the assessment of handicapped individuals]. Annales de Readaptationet de Medecine Physique, 51, 358-365.

Charlton, J.L., Fildes, B., Oxley, J., Keeffe, J., Odell, M., Verdoon, A., et al. (2008). Driving performance and visual search associated with visual field loss. Paper presented at the International Conference on Traffic and Transport Psychology, Washington, DC.

Chelette, T.L., Albery, W.B., Esken, R.L., and Tripp, L.D. (1998). Aviation, Space, and Environmental Medicine, 69(9): 862-868.

Classen, S., Bewernitz, M., Shechtman, O. (2011).Driving Simulator Sickness: an evidence-based review of the literature. *American Journal of Occupational Therapy*, 65, 2, 179-188.

Collet, C., Petit, C., Champely, S. & Dittmar, A. (2003). Assessing Workload through Physiological Measurements in Bus Drivers Using an Automated System During Docking. *Human Factors*, 45(4), 539-548.

Cox, D.J., Singh, H. & Cox, D.M. (2010).Effectiveness of acupressure and acustimulation in minimizing driving simulation adaptation syndrome. Poster presented at the Association for the Advancement of Automotive Medicine Annual Scientific Conference, La Vegas, USA.

Devos, H., Akinwuntan, A.E., Nieuwboer, A., Tant, M., Truijen, S., De Wit, L.,Weerdt, W.D. (2009). Comparison of the effect of two driving retraining programs on on-road performance after stroke. Neurorehabilitation and Neural Repair 23(7) 699-705.

Devos, H., Akinwutan, A., Nieuwboer, A., Tant, M., Kiekens, C. & De Weerdt, W. (2011). Short and long term effects of two driving retraining programmes after stroke: a randomised controlled trial. *Physiotherapy*, 97, 283-284.

Donovan, D.M, Umlauf, R.L & Salzberg, P.M. (1988). Derivation of personality subtypes among high-risk drivers. Alcohol, Drugs and Driving, Vol. 4, pp. 233–244.

Ekman, P., Friesen, W.V. & Ellsworth, P. (1972). Emotion in the human face: Guidelines for research and an integration of findings. Oxford, England: Pergamon Press.

Endsley, M.R. (2000). Direct Measurement of situation awareness: Validity and use of SAGAT. In M.R. Endsley& D. J. Garland (Eds.), Situation awareness and measurement, Mahwah, NJ: Lawrence Erlbaum.

Federal Highway Administration (2007).Drivers' evaluation of the diverging diamond interchange. FHWA TechBrief. Retreived July 16, 2013 from http://www.fhwa.dot.gov/publications/research/safety/07048/index.cfm.

Fisher, D.L, Pradhan, A.K., Pollatsek, A. &Knodler, M.A., Jr. (2007).Empirical evaluation of Hazard anticipation behaviors in the field and on driving simulator using eye tracker. Transportation Research Record, 2018, 80-86.

Flad, N., Nieuwenhuizen, F. M., Bülthoff, H. H., & Chuang, L. L. (2014). System Delay in Flight Simulators Impairs Performance and Increases Physiological Workload. In *Engineering Psychology and Cognitive Ergonomics* (pp. 3-11). Springer International Publishing.

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

Granda, T.M., Davis, G.W., Inman, V.W. & Molino, J.A. (2011).The Use of High-Fidelity Real-Time Driving Simulators for Geometric Design. In D. Fisher, M. Rizzo, J. Caird & J. Lee (Eds.), *Handbook of Driving Simulation for Engineering, Medicine, and Psychology* (34.1-14). Boca Raton, FL: CRC Press.

Hakamies-Blomqvist, L., Ostlund, J., Henriksson, P. &Heikkinen, S. (2001). Elderly car drivers in a simulator – A validation study. Linkoping, Sweden: Swedish National Road and Transport Research Institute.

Hancock, P. & Sheridan, T. [eds Fisher, D., Rizzo, M., Caird, J. & Lee, J.). (2011). The Future of Driving Simulation. Handbook of Driving Simulation for Engineering, Medicine, and Psychology. 4.1-11.

Harrison, A., Derwent, G., Enticknap, A., Rose, F. D., & Attree, E. A. (2002). The role of virtual reality technology in the assessment and training of inexperienced powered wheelchair users. *Disability & Rehabilitation*, *24*(11-12), 599-606.

Hart, S.G. & Staveland, L.E. (1988). Development of the NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.), Human mental workload (pp. 139-183). Amsterdam, the Netherlands: North Holland.

Heino, A., van den Molen, H.H. & Wilde, G.J.S. (1992).Risk homeostasis process in care following behavior: individual differences in care following and perceived risk. Report VK 92-02, Rijksuniversiteit Groningen Haren, The Netherlands.

Hoffman, J.D., Lee, J.D., Brown, T.L. & McGehee, D.V. (2002).Comparison of driver braking responses in a high-fidelity simulator and on a test track. Transportation Research Record, 2, 59-65.

Hogema, J. (1997). Compensation for delay in the visual display of a driving simulator.*Simulation*, 69, 27-34.

Horberry, T., Anderson, J., Regan, M. A., Triggs, T. J. & Brown, J. (2006). Driver distraction: The effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance. *Accident Analysis and Prevention*, 38, 185-191.

Hutchinson, C. (1958). Automobile driving simulator feasibility study (Rep. YM-1244-V-6). Buffalo, NY: Cornell Aeronauticla Laboratory, Inc. (CAL).

Jamison, K. & McGlothlin, W.H. (1973). Drug usage, personality, attitudinal, and behavioral correlates of driving behavior. The Journal of Psychology, Vol. 83, pp. 123–130.

Jamson, H. [eds Fisher, D., Rizzo, M., Caird, J. & Lee, J.). (2011). Cross-Platform Validation Issues. Handbook of Driving Simulation for Engineering, Medicine, and Psychology. 4.1-11.

Jonah, B.A. (1997). Sensation seeking and risky driving: a review and synthesis of the literature, Accident Analysis & Prevention, Vol.29, Issue 5, pp 651-665.

Jones, D.G. & Kaber, D.B (2005).Situation Awareness measurement and the situation awareness global assessment technique. In N. Stanton, A. Hedge, K. Brookhuis, E. Salas, & H. Hendrick (Eds.), Handbook of human factors and ergonomics methods (pp. 42-1 – 42-8). New York: CRC Press.

Kaber, D. B., Riley, J. M., Endsley, M. R., Sheik-Nainar, M., Zhang, T., & Lampton, D. R. (2013). Measuring situation awareness in virtual environment-based training. *Military Psychology*, *25*(4), 330.

Kafrissen, M. E., & Biederman, J. (2006).Using self-reported data to assess the validity of driving simulation data. Behavior Research Methods, 38, 314-324.

Kennedy, R., Lane, N., Berbaum, K. & Lilienthal, M. (1993). A simulator sickness questionnaire (SSQ): A new method for quantifying simulator sickness. *International Journal of Aviation Psychology,* 3, 203-220.

Klee, H., Bauer, C., Radwan, E & Al-Deek, H. (1999). Preliminary validation of driving simulator based on forward speed. Transportation Research Record, 1689, 33-39.

Kramer, Arthur F., and John L. Larish. Aging and dual-task performance. *Aging and skilled performance* (1996): 83-112.

Lansdown, T. C., Brook-Carter, N., & Kersloot, T. (2004). Distraction from multiple in-vehicle secondary tasks: vehicle performance and mental workload implications. *Ergonomics*, *47*(1), 91-104.

Lee, H. C., & Lee, A. H. (2005). Identifying older drivers at risk of traffic violations by using a driving simulator: A 3-year longitudinal study. American Journal of Occupational Therapy, 59, 97-100.

Lee, H. C., Lee, A. H., Cameron, D., & Li-Tsang, C. (2003).Using a driving simulator to identify older drivers at inflated risk of motor vehicle crashes. *Journal of Safety Research*, 34, 453-59.

Lee, H., Falkmer, T., Rosenwax, L., Cordell, R., Granger, A., Viera, B., et al. (2007).Validity of driving simulator in assessing drivers with Parkinson's disease. Advances in Transportation Studies, Special Issue, 81-90.

Lew, H. L., Poole, J.H., Lee, E. H., Jaffe, D. L., Huang, H. C., & Brodd, E. (2005). Predictive validity of driving-simulator assessments following traumatic brain injury: A preliminary study. Brain Injury, 19, 177-188.

Li, Y., Chen, X., Zhang, S., & Lv, H. (2012, June). Studies on subjects design in equipment virtual maintenance training system. In *Quality, Reliability, Risk, Maintenance, and Safety Engineering (ICQR2MSE), 2012 International Conference on* (pp. 513-516). IEEE.

Loo, R. (1979). Role of primary personality factors in the perception of traffic signs and driver violations and accidents. Accident Analysis and Prevention, Vol. 11, pp. 125–127.

Manly, T., Lewis, G. H., Robertson, I. H., Watson, P. C., & Datta, A. K. (2002). Coffee in the cornflakes: time-of-day as a modulator of executive response control. *Neuropsychologia*, *40*(1), 1-6.

Mazer, B., Gelinas, I., Vanier, M., Duquette, J., Rainville, C. & Hanley, J. (2005). Effectiveness of retraining using a driving simulator on the driving performance of clients with a neurologic impairment.Archives of Physical Medicine and Rehabilitation, 86, E20.

McGehee, D.V., Mazzae, E.N. & Baldwin, G.H.S. (2000). Driver reaction time in crash avoidance research: Validation of a driving simulator study on a test track. Paper presented at the meeting of the Human Factors and Ergonomics Association/International Ergonomics Association, Santa Monica, CA.

McMillen, D.L., Smith, S.M. & Wells-Parker, E. (1989). Brief report: the effects of alcohol, expectancy, and sensation seeking on driving risk taking. Addictive Behaviors, Vol. 14, pp. 477–483

Moe, D. & Jenssen, G.D. (1990).*Ungeforere, risikotakingogpedagogiskekonsekvenser* (Young drivers, risk-taking and pedagogical consequences). SINTEF Transport Engineering, Trondheim, Norway.

Moe, D. & Jenssen, G.D. (1993).*Ungdom, risikotakingogbilkjoring* (Young drivers, risk taking and traffic accidents). SINTEF Transport Engineering, Trondheim, Norway.

Moore, D., Cheng, Y., McGrath, P., & Powell, N. J. (2005). Collaborative virtual environment technology for people with autism. *Focus on Autism and Other Developmental Disabilities*, *20*(4), 231-243.

Mourant, R. R., Ahmad, N., Jaeger, B. K. & Lin, Y. (2007). Optic flow and geometric field of view in a driving simulator display. *Displays*, 28, 145-159.

Mullen, N., Charlton, J., Devlin, A. & Bedard, M. (2011). Simulator Validity: Behaviors observed on the Simulator and on the Road. In D. Fisher, M. Rizzo, J. Caird & J. Lee (Eds.), *Handbook of Driving Simulation for Engineering, Medicine, and Psychology* (13.1-13.18). Boca Raton, FL: CRC Press.

Palinko, O. (2010). Estimating cognitive load using remote eye tracking in a driving simulator. *In Proceedings of the 1010 Symposium on Eye-Tracking Research and Applications*, p141. New York, NY, USA: Association for Computing Machinery.

Patomella, A.-H., Tham, K., & Kottorp, A. (2006). P-drive: Assessment of driving performance after stroke. Journal of Rehabilitation Medicine, 38, 273-279. Reimer, B., D'Ambrosio, L. A., Coughlin, J. E.,

Pauzie, A. (2008). A method to assess the driver mental workload: The driving activity load index (DALI), IET Intelligent Transport Systems, 2(4), 315-322.

Pertaub, D. P., Slater, M., & Barker, C. (2002). An experiment on public speaking anxiety in response to three different types of virtual audience. *Presence: Teleoperators and virtual environments*, *11*(1), 68-78.

Philip, P., Sagaspe, P, Taillard, J., Valtat, C., Moore, N., Akerstedt, T., et al. (2005). Fatigue, sleepiness, and performance in simulated versus real driving conditions. Sleep, 28, 1511-1516.

Poeschl & Doering. (2012). Designing Virtual Audiences for Fear of Public Speaking Training–An Observation Study on Realistic Nonverbal Behavior. *Annual Review of Cybertherapy and Telemedicine 2012: Advanced Technologies in the Behavioral, Social and Neurosciences*, *181*, 218.

Ream, E., & Richardson, A. (1996). Fatigue: a concept analysis. *International journal of nursing studies*, *33*(5), 519-529.

Reed, M.P. & Green, P.A. (1999).Comparison of driving performance on-road and in a low-cost simulator using a concurrent telephone dialing task. *Ergonomics*, 42, 015-1037.

Reimer, B., Mehler, B., Coughlin, J. F., Godfrey, K. M. & Tan, C. (2009).An on-road assessment of the impact of cognitive workload on physiological arousal in young adult drivers. Proceedings of the First International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Sep 21-11 2009, Essen, Germany.

Reimersma, J.B.J., van der Horst, A.R.A, Hoekstra, W., Alink, G.M.M. &Otten, N. (1990). The validity of a driving simulator in valuating speed-reducing measures. Traffic Engineering and Control, 31, 416-420.

Reitz, E. A., & Seavey, K. P. (2012, January). Using Virtual Environments to Improve Real World Performance in Combat Identification. In *The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC)*(Vol. 2012, No. 1). National Training Systems Association.

Ricard, G., Cyrus, M., Cox, D., Templeton, T. & Thompson, L. (1978). *Compensation for transport delays produced by computer image generation systems* (IH-297/AFHRL-TR-78-46). Orlando, FL: Human Factors Laboratory, Naval Training Equipment Center.

Rimmo, P.A. &Aberg, L. (1996). Do sensation seekers make more violations and errors as drivers. Paper presented at the *International Conference on Traffic and Transport Psychology*, Valencia, Spain.

Rizzo, A., Buckwalter, J. G., John, B., Newman, B., Parsons, T., Kenny, P., & Williams, J. (2012). STRIVE: Stress Resilience in Virtual Environments: a pre-deployment VR system for training emotional coping skills and assessing chronic and acute stress responses. *Studies in health technology and informatics*, *173*, 379.

Rizzo, A., Difede, J., Rothbaum, B. O., Reger, G., Spitalnick, J., Cukor, J., & Mclay, R. (2010). Development and early evaluation of the Virtual Iraq/Afghanistan exposure therapy system for combat-related PTSD. *Annals of the New York Academy of Sciences*, *1208*(1), 114-125.

Seki, Y., & Sato, T. (2011). A training system of orientation and mobility for blind people using acoustic virtual reality. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, *19*(1), 95-104.

Shinar, D. & Ronen, A. (2007).Validation of speed perception and production in a single screen simulator. Advances in Transportation Studies, Special Issue, 51-56.

Slater, M., Pertaub, D. P., & Steed, A. (1999). Public speaking in virtual reality: Facing an audience of avatars. *Computer Graphics and Applications, IEEE*,*19*(2), 6-9.

Slick, R.F., Evans, D.F., Kim, E. & Steele, J.P (2006).Using simulator o train novice teen drivers: Assessing psychological fidelity as a precursor to transfer of training. Paper presented at the Driving Simulation Conference – Asia/Pacific, Tsukuba, Japan.

Son, J., Reimer, B., Mehler, B., Pohlmeyer, A. E., Godfrey, K. M., Orszulak, J., ... & Coughlin, J. F. (2010). Age and cross-cultural comparison of drivers' cognitive workload and performance in simulated urban driving. *International Journal of Automotive Technology*, *11*(4), 533-539.

Stoner, H.A., Fisher, D.L., Mollenhauer, M. (2011).Simulator and Scenario Factors Influencing Simulator Sickness. In D. Fisher, M. Rizzo, J. Caird& J. Lee (Eds.), *Handbook of Driving Simulation for Engineering, Medicine, and Psychology* (14.1-14.24). Boca Raton, FL: CRC Press.

Tepper, Oren M., et al. "Use of virtual 3-dimensional surgery in post-traumatic craniomaxillofacial reconstruction." *Journal of Oral and Maxillofacial Surgery*69.3 (2011): 733-741.

Ting, C. H., Mahfouf, M., Nassef, A., Linkens, D. A., Panoutsos, G., Nickel, P., ... & Hockey, G. (2010). Real-time adaptive automation system based on identification of operator functional state in simulated process control operations. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, *40*(2), 251-262.

Tomasevic, N., Regan, M.A.,Duncan, C.C. &Desland, M.L. (2000).Use of advanced simulation to treat a driving-related phobia. Paper presented at the SimTecT Conference and Exhibition. Sydney, Australia.

Tornros, J. (1998). Driving behavior in a real and a simulated road tunnel – A validation study.Accident Analysis and prevention, 30, 497-503.

Toxopeus, R. (2007). Differences between elderly and young drivers: Driving vigilance in two tasks. Unpublished master's thesis, Lakehead University, Thunder Bay, Ontario, Canada.

Verwey, W. (2000). On-line driver workload estimation: Effects of road situation and age on secondary task measures. *Ergonomics*, 43(2), 187-209.

Wade, M.G. & Hammond, C. (1998). Simulator validation: Evaluating driver performance in simulation and the real world (Report No. 28). Minneapolis, MN: Minnesota Department of Transportation.

Wagner, J., Kim, J. & Andre, E. (2005). From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification. In Proceedings of IEEE International Conference on Multimedia and Expo, 2005, pp 940-943.

Wang, Y., Mehler, B., Reimer, B., Lammers, V., D'Ambrosio, L. A., & Coughlin, J. F. (2010). The validity of driving simulation for assessing differences between in-vehicle informational interfaces: a comparison with field testing. *Ergonomics*, *53*(3), 404-420.

Weir, D. & Wojcik, C. (1971). Simulator Studies of the driver's dynamic response in steering control tasks. *Highway Research Record*, 364, 1-15.

Wells-Parker, E., Anderson, B., Pang, M. &Timken, D. (1993). An examination of cluster-based classification schemes for DUI offenders. Journal of Studies on Alcohol, Vol. 54, pp. 209–218.

Willcutt, Erik G., et al. "Validity of the executive function theory of attention-deficit/hyperactivity disorder: a meta-analytic review." *Biological psychiatry*57.11 (2005): 1336-1346.

Wilson, M., McGrath, J., Vine, S., Brewer, J., Defriend, D., & Masters, R. (2010). Psychomotor control in a virtual laparoscopic surgery training environment: gaze control parameters differentiate novices from experts.*Surgical endoscopy*, *24*(10), 2458-2464.

Wilson, P. N., Foreman, N., & Tlauka, M. (1996). Transfer of spatial information from a virtual to a real environment in physically disabled children. *Disability & Rehabilitation*, *18*(12), 633-637.

Wilson, R.J & Jonah, B.A.(1988). The application of problem behavior theory to the understanding of risky driving. Alcohol, Drugs and Driving, Vol. 4, pp. 173–191

Wilson, R.J. (1988). Subtypes of DWIs and high-risk drivers: implications for differential intervention. Alcohol, Drugs and Driving, Vol. 7, pp. 1–12.

Yan, X., Abdel-Aty, M., Radwan, E., Wang, X. & Chilakapati, P. (2008).Validating a driving simulator using surrogate safety measures. Accident Analysis & Prevention, 40, 272-288.

Yang, C., Guo, J., Han, J., Liao, X., & Yuan, Z. (2013). Coupled tissue bleeding simulation in virtual surgery. In *Intelligent Computing Theories* (pp. 331-338). Springer Berlin Heidelberg.

Young, M.S. & Stanton, N.A. (2005).Mental workload. In N. Stanton, A. Hedge, K. Brookhuis, E. Salas & H. Hendrick (Eds.), Handbook of human factors and ergonomics methods (pp. 39-1 – 39-9). Boca Raton: CRC Press.

Zijlstra, F.R.H. (1993). Efficiency in work behavior. A design approach for modern tools. Doctoral dissertation, Delft University of Technology. Delft, the Netherlands: Delft University Press.

Zuckerman, M. & Neeb, M. (1980).Demographic influences in sensation seeking and expressions of sensation seeking in religion, smoking and driving habits. Personality and Individual Differences, Vol. 1, pp. 197–206

Zuckerman, M. (1994) *Behavioural Expressions and Biosocial Bases of Sensation Seeking*. University of Cambridge Press, Cambridge.