Driving Simulator and Scenario Effects on Driver Response

Jessica Mueller and Laura Stanley
Western Transportation Institute, Montana State University
Bozeman, MT

Taylor Martin
Embry-Riddle Aeronautical University
Daytona Beach, FL

Conor Gallagher
Virginia Polytechnic Institute and State University
Blacksburg, VA

Abstract

Driving simulators are becoming a common research tool for investigating driver behavior in safe, controlled environments. Wide ranges of simulator configurations are available for research—this wide range, as well as and the complexity of the virtual worlds being driven, has not been studied for specific effects on driver response. This study examined driver behaviors in virtual scenarios with varying complexity (difficulty of driving maneuvers) while using driving simulators of different levels of fidelity to assess how complexity and fidelity can affect physiological and subjective mental workload ratings. Participants were equipped with physiological data collection equipment and drove each scenario twice on each simulator; as a baseline drive and again while completing a one-back recitation task. Participants completed NASA-TLX surveys to rate self-reported overall mental demand after each scenario. ANOVA analysis showed that scenario complexity and secondary task were significant factors for all NASA-TLX subscales as well as for driver pupil radius; simulator complexity had no effect. These findings indicate that lower-fidelity simulators may be as useful a research tool as more expensive higher-fidelity machines for physiological and subjective mental workload ratings; a useful finding for research labs without access to high-fidelity simulators.

Keywords
Driving simulators, simulator fidelity, scenario complexity, mental workload, physiological driver response

1. Introduction

Transportation in the United States is something that is widely taken for granted and the issues that arise with it can easily be overlooked. Although driving has become increasingly habitual over the years, it is important to realize it is not an automated task. There is still a considerable amount of thought that should go into driving. According to the Department of Transportation’s National Highway Traffic Safety Administration [1], the number of fatalities due to motor vehicle crashes was estimated around 34,080 in 2012 alone. This was a 5.3 percent increase from the previous year. Also, the Official U.S. Government Website for Distracted Driving stated that in 2011 distracted driving was the cause of 3,331 fatalities and additionally 387,000 injuries [2]. These data show an important need for research to improve driving safety. Driving simulators can provide knowledge and training for drivers to improve traffic safety. This report will discuss characteristics of driving simulators important when considering the use of a simulator. Fidelity is the degree to which the simulator reproduces or mirrors real world situations [3]. High-fidelity simulators cost more, but in return tend to produce more realistic responses from the three main sensory channels [4]. Based on the costs, simulators are generally put into the following three categories from least expensive to most expensive, respectively: low-level, mid-level, and high-level [4]. Another characteristic that helps define these levels of simulators is the degree to which the simulator stimulates the three main senses—visual,
kinesthetic, and auditory. Low-level simulators are associated with limited utility for providing feedback to these three senses. A high-level simulator provides greater stimulus of all three senses [4]. Overall, it is evident that training programs that use simulation have shown a positive effect on driver behavior, but it is still uncertain whether it will significantly reduce vehicle accidents [5-7].

The overall objective of this study is to better understand driver cognitive state in specific to see how these different types of simulators and scenarios can affect mental workload. Understanding how a driver’s cognitive state changes in different simulators is important in understanding the wide variety of simulators and configurations available to research institutions. In order to validate the use of simulators to increase driver safety, it is important that specific characteristics of the simulators be evaluated for fidelity and validity across platforms, as well as between simulators and the real world. This study focuses on the comparisons of different fidelity levels of simulators in different research applications. Potential applications include research evaluating road geometry design, human–computer interface testing, and training groups of at-risk drivers.

Over several years, researchers have studied whether positive transfer of higher-level cognitive skills can occur between a simulator and the real world [8]. The Swedish research project called TRAINER began in 2000. The project’s goal was to develop new ways for driver training using simulators and computer-based training (CBT) methods [6]. The project used two types of simulators, a Low Cost Simulator (LCS) and a Mean Cost Simulator (MCS). These two simulators differed in field of view and motion availability, with the MCS consisting of greater field of view and some motion capabilities and the LCS with a smaller field of view and no motion capabilities. Three groups of subjects received different treatment for the study. The first group had computer-based training followed by simulator training on the LCS, the second group had computer-based training followed by simulator training on the MCS, and the third group received no training via any of the methods. After the training, participants were presented with six scenario types in a high-level motion-based simulator for testing. Results showed some positive effects on performance of young drivers who experienced computer-based training and simulator training, and a greater effect was evident for those drivers who experienced training in the Mean Cost Simulator than those in the Low Cost Simulator or those without any training [6]. These findings support previous [9] that there is a correlation between success of training in simulators and the financial expense of the simulator.

There are three major categories for measuring mental workload—physiological, subjective, and performance measures. Physiological indicators can be used to assess cognitive workload. Pupillary has been used in numerous research efforts as an indicator of mental workload [10]. Pupillary dilation occurs quickly at the onset of processing the task at hand and diminishes immediately following the processing. Not only does dilation of the pupil occur during processing, there is also a direct correlation between the size of the pupils and the amount of workload (processing) present at the time. The results of a study on cognitive load in relation to pupil size found that the more complex tasks resulted in a higher value of pupil size [11, 12]. According to The Handbook of Human Factors Testing and Evaluation [13] heart rate variability, breathing rate, and skin conductance have all been validated as methods to measure driver workload. In research done on these physiological measurements it has been shown that all three will increase as the complexity of the task increases [14].

A second major category for measuring workload is subjective assessment of mental workload. There are several options for subjectively measuring workload. The NASA Task Load Index (NASA-TLX) is a widely used, multifaceted scale that was developed to assess workload of operators over a broad spectrum of jobs [15]. It is usually performed during the task or immediately following the task being observed. The scale covers six main variables that were considered after very extensive research. These six variables are physical, mental, and temporal demands, frustration, effort, and performance [15].

Performance measures provide a third effective means to assess cognitive workload; there are predominantly two types of performance measures that can be used to assess cognitive workload of drivers [13]. These performance measures focus on a primary task or a secondary task [13]. Performance on primary tasks for drivers has been studied using measures such as standard deviation of lateral position and variation in speed [16]. Secondary tasks have been included in some studies to provide more data on the cognitive workload of drivers in certain scenarios [17].

Although none of these measures alone ensure the validity of driving simulators for training purposes, when used together they can provide a solid explanation on the implications of cross-platform validity and training. By
measuring mental workload of drivers in different scenarios and using different measurements one can better understand the cognitive demands placed on drivers [16]. Research in similar areas will continue to provide new data on cross-platform validation of simulators and the implications of training on driver behaviors in the real world.

2. Methods

2.1 Participants
Participants were recruited using advertisements, fliers, and posters around the Bozeman, MT, area. Eligibility criteria for participation was that participants needed to be between 18 and 55 years old, have a valid driver license, have normal or corrected-to-normal vision, and show no susceptibility to motion sickness. Participants were compensated $10 per hour, and the experimental session lasted approximately two hours. Twenty people participated in the study (45% female) and the average participant age was 23.3 years. This study was conducted with Institutional Review Board approval.

2.2 Equipment
This study used two simulators from the Western Transportation Institute’s (WTI) Advanced Driving Simulation Lab at Montana State University—a fixed-base Drive Safety simulator and an advanced motion-capable high-fidelity Realtime Technologies Incorporate (RTI) simulator. The fixed base simulator has a 155° horizontal field of view with a quarter-cut Saturn sedan cab, with the graphics viewed on five adjacent monitors (Figure 1).

![Figure 1. Fixed-Base Driving Simulator (left) and Advanced Driving Simulator (right)](image)

The advanced simulator has a 240° horizontal field of view with graphics projected on a wraparound projection screen with a 2009 Chevy Impala body. The Moog motion base allows for approximately 18 inches of movement with six degrees of freedom to replicate the motions associated with driving. The advanced simulator was used with two configurations: one with motion enabled, and again without motion capability. Two different scenario development software tools were used to author the simulator scenarios; Hyperdrive (1.9 35) was used to develop scenarios for the DriveSafety simulator and Internet Scene Assembler (2.0) was used for the advanced simulator.

A Zephyr BioHarness™ BT was used to track the participant’s heart and breathing rate, and the BioHarness data were manually synchronized to the simulator data. An Applied Science Laboratory MobileEye was used to collect pupil diameter.

3.3 Procedure
Upon arrival at the Advanced Driving Simulation Lab, participants read and signed informed consent paperwork, in accordance with Institutional Review Board policy. Participants were given a vision test and a demographic survey to confirm their eligibility and gather demographic information. Researchers conducted practice n-back tests to ensure that participants were comfortable with 1-back tests. Once the n-backs trials were completed, participants were equipped with both a BioHarness and a head-mounted MobileEye device.
Participants drove specifically designed driving routes, consisting of three distinct driving scenarios requiring different levels of complexity: a low-complexity scenario (going straight on a rural two-lane road), a mid-complexity scenario (going straight through a signaled intersection), and a high-complexity scenario (performing an unprotected left-hand turn at a signaled intersection). The low-, mid-, and high-complexity scenarios were combined in random order to create a route that the participant travelled. The scenarios consisted of two-lane rural roads with no foliage or roadside objects along the road. The low-complexity scenario was a straight roadway stretching 1,000 meters, the mid-complexity scenario was a straight roadway stretching 1,000 meters going straight through a four-way intersection controlled by a traffic light. The high complexity level was 1,000 meters long with a four-way intersection controlled by lights, where the participant was required to make an unprotected left-hand turn at the light. When the low-, mid-, and high-complexity scenarios were combined into the driving routes, a 1000-meter section of straight two-lane road was added between each section to avoid any carryover effects for driving behaviors between the scenarios in each route.

After each scenario (three times per drive), the participants were asked to pull over to the side of the road, where they completed a NASA-TLX survey, followed by a simulator sickness questionnaire to make sure that they were not experiencing simulator sickness symptoms.

These drives were repeated on three different simulator configurations: the fixed-base simulator, the advanced simulator, and the advanced simulator with motion capability disabled. In each simulator configuration, the participants drove through a brief practice drive to familiarize themselves with the vehicle dynamics and then drove through two of those randomized routes. The first route was a baseline drive where no secondary tasks were administered, followed by the second route, participants engaged in a 1-back verbal response secondary task.

4. Findings

4.1 NASA-TLX

Once data was collected and reduced, a repeated-measures ANOVA was run for each NASA-TLX subscale—mental demand, physical demand, temporal demand, performance, effort, and frustration. The analysis showed that for all six of the NASA-TLX subscales and the rating scale for mental effort there were significant findings for the scenario complexity and secondary task.

The same general trends were observed for all subscales: there was generally no significant difference in NASA-TLX scale scores for simulator, but on the performance subscale, participants reported decreased performance in the fixed base simulator ($\mu_{FixedBase} = 16.03$) compared to the advanced simulator with motion and the advanced simulator without motion ($\mu_{AdvMotion} = 17.42$, $\mu_{NoAdvMotion} = 16.95$). Performance scores are reverse-scaled in Figure 1, and were subjectively scored out of a maximum possible performance score of 20 by the participant.
Table 1: Summary of NASA-TLX ANOVA Results

<table>
<thead>
<tr>
<th>NASA-TLX Subscale</th>
<th>Simulator Type</th>
<th>Secondary Task</th>
<th>Scenario Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frustration</td>
<td>NS</td>
<td>F=49.99, p&lt;0.0001</td>
<td>F=6.36, p=0.0022</td>
</tr>
<tr>
<td>Effort</td>
<td>NS</td>
<td>F=107.06, p&lt;0.0001</td>
<td>F=3.73, p=0.0261</td>
</tr>
<tr>
<td>Performance</td>
<td>F=5.21, p=0.0064</td>
<td>F=39.75, p&lt;0.0001</td>
<td>F=4.28, p=0.0154</td>
</tr>
<tr>
<td>Temporal Demand</td>
<td>NS</td>
<td>F=26.39 p&lt;0.0001</td>
<td>F=4.52, p=0.012</td>
</tr>
<tr>
<td>Physical Demand</td>
<td>NS</td>
<td>F=8.46, p&lt;0.0041</td>
<td>F=4.59, p=0.0115</td>
</tr>
<tr>
<td>Mental Demand</td>
<td>NS</td>
<td>F=119.94, p&lt;0.0001</td>
<td>F=3.63, p=0.0286</td>
</tr>
</tbody>
</table>

Secondary task assessment showed the same trend across all subscales. As the secondary task was added, self-reported workload increased, and performance decreased. Looking at scenario complexity, there was generally no significant difference between the low- and mid-complexity scenarios; but the higher-complexity scenario had significantly higher self-reported workload and lower performance scores. These trends are evident in Figure 1, below.

4.2 Physiological Measures

The dependent physiological variables that were evaluated include driver heart rate, breathing rate, and pupil diameter. Heart rate variance was measured but due to the short intervals of interest for this study was not long enough for a reasonable HRV measurement, and so was not included in the analysis. An initial general model ANOVA was performed with independent variables being the Scenario Complexity, Secondary Task, and Simulator Type; with the dependent variables including heart rate, pupil diameter. Pupil diameter data were normally distributed; however, ANOVA analyses were conducted for all measures as ANOVA analyses are robust to slight departures from normality.

**Heart Rate.** Significant factors affecting heart rate were the presence or absence of a secondary task (F=13.94, p=0.0003), and the task complexity (p=10.57, p<0.0001); there was no effect from simulator type. When a secondary task was applied to the drivers, heart rate was on average 4.01 beats per minute faster than when no secondary task was applied. Heart rate was significantly higher when navigating through the high-complexity scenario (µ= 76.61 bpm) compared to mid- and low- (µmid = 70.63, µlow = 69.96 bpm), but not significantly different between the mid- or low-complexity scenarios.

**Breathing Rate.** Significant factors affecting participant breathing rates were the type of simulator (F=6.62, p=0.0022) and whether or not a secondary task was applied (F=4.02, 0.0484). Breathing rates were significantly higher in the fixed-base simulator (µFixedBase= 11.40 bpm); compared to the advanced simulator both with and without motion (µadvMotion = 9.31, µAdvNoMotion= 9.12 bpm); however breathing rates in the advanced simulator with motion was not significantly different from the advanced simulator without motion enabled. Breathing rates were
significantly higher without an applied secondary task (µ\text{NoTask} = 10.71 \text{ bpm}) than with a 1-back test (µ_{1-back} = 9.29 \text{ bpm}). There were no significant differences in breathing rate between the different scenario complexities.

**Skin Temperature.** The only significant factor affecting driver skin temperature was the secondary task (F=11.37, p=0.0009). When a secondary task was applied, skin temperature was slightly higher than when no task was applied (µ_{1-back} = 35.34, µ\text{NoTask} = 35.14 \text{ °C}).

**Pupil diameter.** All three independent variables significantly affected pupil diameter. Pupil diameter was significantly different among simulator types (F=90.0, p<0.0001), with the fixed-base simulator showing significantly smaller pupil diameters (µ\text{FixedBase} = 8.54mm) compared to the advanced simulator (µ\text{advanced} = 10.25mm) and the advanced simulator without motion (µ\text{AdvNoMotion} = 9.98mm). Looking at the effects of secondary task (F=11.57, p=0.0009); pupil diameter was larger when performing the secondary task (µ_{1-back} = 9.83mm) compared to no task (µ\text{NoTask} = 9.44mm). Complexity was also significant (F=11.98, p<0.0001), with the higher-complexity scenario resulting in a significantly larger pupil diameter (µ\text{highComplex} = 10.05mm) than the mid- and low-complexity scenarios (µ\text{MidComplex} = 9.45; µ\text{LowComplex} = 9.41mm), which were not significantly different from each other.

5. Conclusions
The main differences between the different types of simulators explored in this study is involving self-reported performance scores. Drivers generally scored themselves higher in the more advanced simulator. The only physiological parameter that showed differences was pupil diameter—but this could be due to the different rooms and ambient light levels found in both simulator labs. If this is not the case, there is a possibility that the larger pupil diameters found in the advanced simulator could be due to a higher cognitive load [6]. There was a significant decrease in performance between the advanced simulator with motion and the fixed-base simulator when the secondary task was present. This could mean that the advanced simulator is a better tool for self-reported driving performance; it could be easier for participants to become used to the advanced simulator due to its higher degree of physical validity. These results support research that showed a significantly greater positive effect on performance of young drivers who had received training on a median-cost simulator versus young drivers who received training on a low cost simulator [7]. There were no statistically relevant differences between the advanced simulator with motion compared to without motion. This is similar to results suggesting that motion may not be a large contributor to driver workload in simulator research [19, 20].

Manipulating the complexity of the driving task showed general similarities between the low- and mid-complexity scenarios: the NASA-TLX scores show a significant increase in overall demand between the low- and mid-complexity scenarios and the hard scenario. These findings are similar to a study that used similar road scenario complexity levels and found that the only significant difference in performance occurred when the participants were asked to turn at an intersection rather than proceed straight on the road [21]. The scenario complexity showed significant results in all six NASA-TLX subscales.

As far as the applied secondary task, these results were generally expected. The NASA-TLX surveys showed a significant increase in overall demand when a secondary task was presented to the driver rather than no secondary task. The secondary task also showed significant results in all six NASA-TLX subscales and the rating scale of mental effort, as well as for the physiological variables.

This study contributes to gaining more knowledge of different types of simulators in order to provide evidence towards cross-validation of simulators. The measures collected in this study provide a comprehensive idea of the different types of simulators and the effects that each have on drivers in certain scenarios. Further analysis of the already existing data could help gain a deeper insight to the differences between simulators as well as the individual differences of participants. Future research could also be done to build upon the findings of the current research to support the validity of driving simulators for training purposes.

**Acknowledgements**
This material is based upon work supported by the National Science Foundation under Grant No. 1116378, and assisted by funding from the Research Experience for Undergraduates program at Montana State University.
References