2007

Evaluation of virtual reality snowplow simulation training

Christopher Michael Masciocchi
Iowa State University

Evaluation of virtual reality snowplow simulation training

by

Christopher Michael Masciocchi

A thesis submitted to the graduate faculty

in partial fulfillment of the requirement for the degree of

MASTER OF SCIENCE

Major: Psychology

Program of Study Committee:
Derrick J. Parkhurst, Major Professor
Veronica J. Dark
William S. Robinson

Iowa State University

Ames, Iowa

2007

UMI Number: 1443061

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

The effectiveness of virtual reality snowplow simulator training for current Iowa Department of Transportation employees was examined. Operators received approximately two hours of training, which included several drives in a driving simulator designed to emulate a snowplow. Performance on a test scenario was compared for operators who had received this training versus those who were trained after the test scenario. Participants also completed a variety of personality and training questionnaires designed to measure personality tendencies, as well as their opinions of training and the realism of the simulator. Responses to these questionnaires were generally positive: operators reported that the features of the simulator mimicked those of a real snowplow, and that they enjoyed all aspects of training. Moreover, several performance differences (e.g., number of collisions, average speed and fuel consumption) were found between trained and untrained operators. They suggest that snowplow simulator training improved the driving performance of trained operators.

CHAPTER 1. INTRODUCTION AND LITERATURE REVIEW

Introduction

Each winter, Iowa Department of Transportation (IDOT) maintenance operators are responsible for plowing snow off federal and state roads in Iowa. Drivers typically work long shifts under treacherous conditions. In addition to properly navigating the vehicle, drivers are required to operate several plowing mechanisms simultaneously, such as plow controls and salt spreaders. Furthermore, there is little opportunity for practicing these skills in real-world situations. During snowfalls, when training would be most realistic and effective, all available vehicles, drivers, and potential instructors are required to be plowing on the roadways. Consequently, novice operators often do not undergo as comprehensive a training regimen as desired, and experienced operators do not have opportunities to improve their current practices or test new ones. Additionally, conducting novice snowplow operator training on roadways may present an unnecessary hazard for the trainee as well as other drivers.

Virtual reality training is an option when real-world training would be prohibitively high-priced, inappropriate, or hazardous. For example, because they provide a safe yet realistic environment in which students can be taught how to operate aircraft, practice in flight simulators is a basic requirement of pilot training programs. A similar training program for snowplow operators would provide experienced operators with a chance to practice their skills under realistic yet safe conditions and would supply basic training to novice or less-experienced operators. Training could be conducted during any time of year, which would be especially beneficial during the summer when vehicles and drivers are not pre-engaged. In order to provide such training to snow plow operators in Iowa, the IDOT purchased a snowplow simulator, a TranSim VS III, in 2005 from L3 Communications.

Research Using Driving Simulators

Driving simulators have recently become a desirable training and research tool due to technological advances in display quality and research advocating their applicability. Several studies have shown that individuals tend to operate driving simulators in virtual environments similar to real vehicles, implying that they are high in relative behavioral validity. In particular, a wide field of view and the ability to modify individual pieces of equipment increase their realism. Due to advances in simulator fidelity, driving simulators are often used in lieu of real vehicles in basic and applied settings, especially when conducting a particular study or training regimen in real life would be prohibitively difficult or expensive. This suggests that researchers and trainers are generally confident that driving simulators provide an accurate approximation of real-world conditions.

Because driving simulators provide a realistic yet non-hazardous driving experience, several studies have examined trainees' eye movements as they navigate simulated environments, without the risk of causing a serious accident. Eye movements can be used to determine two important aspects of driving behavior: (1) the typical adaptive scan pattern employed by experienced drivers during normal driving conditions and (2) the effects of certain distractions on people's scanning behavior.

Finally, one area that has received a lot of attention is the comparison between driving behaviors of novice and expert drivers. Research in this area demonstrates how eye movement and behavioral measurements can both be used for comparing the performance of groups at different training levels. In addition to the above topics, this chapter summarizes current literature on driving simulator research and training practices, discusses possible causes and preventions of cybersickness, and provides a review of personality measurements related to job performance.

Simulator Fidelity

For driving simulator training to be considered a useful technique for improving drivers' skills, the experience of operating it should approximate that of a real vehicle. Sparse and inaccurate visual information may reduce participants' level of perceived presence in the simulator, the extent that they feel like they are a part of the virtual environment. This feeling of presence is crucial both for inducing individuals to properly perform the assigned tasks within the simulator and for facilitating transfer of the knowledge gained in the virtual environment to real-life conditions (Witmer & Singer, 1998). Moreover, the operation of the virtual vehicle must also be similar to conventional vehicles. Otherwise, skills learned in the simulator may not transfer to real driving situations. Thus, it is imperative that the visuals, as well as the location and responsiveness of the vehicle controls, resemble those of a real vehicle.

Kemeny and Panerai (2003) reviewed the current literature on perception within driving simulators to determine which factors are critical for realism. In particular, they investigated which visual and haptic cues are present in driving simulators that are also apparent in real vehicles. The primary feature of driving simulators that accounts for their similarity to real vehicles is that they provide immersive optic flow, or a sense of motion, from the movement of objects in the scene. All objects in the scene—even those in the periphery that are not the immediate focus of attention—move in relation to the driver and provide visual feedback concerning the driver's speed. Consequently, training in immersive virtual reality simulators is superior to training with standard computer displays, as simulators provide encompassing visual feedback in an immersive environment. However, Kemeny and Panerai stated that most simulators lack other types of cues that are present in real vehicles. For instance, motion parallax, the phenomenological experience that pairs of objects move in different directions depending on their location and distance from the observer, is not present in most driving simulators. This is because head position is typically not calibrated, preventing the real time updating of relative object position. Still, Kemeny and Panerai concluded that simulators with a large field of view, greater than 120 degrees, provide participants with enough information that they can estimate speed relatively accurately.

Kinesthetic (body movement) and haptic (touch) cues are also important in driving simulators but are only found in motion-based simulators. Kemeny and Panerai (2003) stated that kinesthetic cues have been shown to decrease reaction time to external disturbances (e.g., simulated strong winds) and to affect drivers' lane position and speed, especially while turning. For instance, drivers tend to take wider turns when haptic cues are available versus when they are absent. Consequently, motion-based simulators are considered to be superior to fixed simulator models, as they provide these cues and thus offer a greater degree of presence and realism.

Appropriate visual and haptic information, then, is critical to the fidelity of driving simulators. In addition, the operation of the equipment and functionality of the peripherals in the simulator—the pedals, steering wheel, and mirrors—must correspond to those of a real vehicle to facilitate the transfer of skills to real driving. While it is certainly not necessary that these features perfectly mimic the performance in an vehicle, they should have comparable functionality and exhibit similar responsiveness. Note that the TranSim VS III fulfills these criteria, as the mirrors, seat, gear shifts, and transmission can be manipulated, allowing a trainer to tailor the simulator to meet task requirements. These features may be adjusted to mimic the driver's real-life vehicle or to provide the operator with a novel driving experience.

Overall, it is critical that driving simulators approximate real vehicles, which would be demonstrated if drivers' behavior in the virtual environment is similar to their behavior in real vehicles. In other words, for a driving simulator to be considered effective, one needs to demonstrate that it has sufficient validity.

Blaauw (1982) proposed two types of validity related to simulator validation: physical validity and behavioral validity. Physical validity refers to the physical similarity between the layout and dynamics of the simulator compared to a real vehicle. This type of validity is often accounted for with a description of the simulator and its similarities to the modeled vehicle. Behavioral validity concerns the similarity of the operators' behaviors as he or she interacts with the environment in the virtual reality simulator compared to a real vehicle (e.g., speed, following distance, and lane position). According to Blaauw, there are two types of behavioral validity: absolute validity and relative validity. Absolute validity is attained when a behavioral measurement from the simulator and a real life environment are highly similar. Not surprisingly, this type of validity is rarely achieved. Relative validity, on the other hand, is achieved when the differences in the two conditions are in the same direction and have relatively similar magnitudes. Finding relative validity is more feasible, and it can still provide strong support for the high fidelity of a simulator. In fact, Tornros

(1998) suggested that relative validity is sufficient for simulators to be considered valid approximations of real vehicles.

Godley, Triggs, and Fildes (2002) examined the behavioral validity of the Monash University Accident Research Centre driving simulator for speed research. In their experiment, two groups of drivers operated a real car and a driving simulator. Each condition contained six critical sites: stop sign, left turn, and right turn, each of which included both a rumble strip present and rumble strip absent scenario. Driving speeds were compared for the real vehicle versus the simulator as participants drove over the rumble strips. Absolute validity for speed in this study would be achieved if the operators' speeds for both the rumble strip present and absent conditions were equivalent in the simulator and real-life environments.

Two different types of relative validity were examined: interactive relative validity and average relative validity. Interactive relative validity was measured by comparing the speed profile of each site at three instances for the simulator and real vehicle conditions. Specific comparisons were made of the vehicle speed as drivers approached the rumble strips, went over them, and then either stopped or completed the turn at each site. This was designed to determine how and when the presence of a rumble strip affected driver speed in the simulator versus the real vehicle. Average relative validity was determined by taking the mean speeds of the vehicle for each site. Relative validity would be represented if the disparity between the mean speeds in all sites were similar in direction and magnitude for the simulator and real-vehicle conditions.

Overall, Godley et al. (2002) found that individuals reduced their speeds where rumble strips were present earlier than at those sites without rumble strips. The correlations at each section of the course were similar for the real car and simulated vehicle, and thus interactive relative validity for speed in this driving simulator was confirmed. However, the mean speed differences between the sites with and without rumble strips were larger for the simulator than the road trials. Because participants drove faster in the simulator when rumble strips were not present versus when they were compared to the real vehicle. This was the case when the rumble strip in the treatment condition preceded a right or left turn. When the rumble strip appeared before a stop sign, however, mean speeds were highly correlated between the simulator and road tests. Average relative validity was therefore achieved for trials in which participants reacted to the presence of a stop sign, but was not found for the left and right turn conditions. Furthermore, absolute speeds were generally different in the various sites between simulator and road tests. Absolute validity comparing the speed individuals drove the driving simulator and the model vehicle, therefore, was not found for this study. Importantly, however, participants' speed adjustments to the presence of rumble strips in the simulator generally mimicked those made in the real car.

Panerai et al. (2001) conducted a similar study using speed and following distance to measure simulator fidelity. Four professional truck drivers participated in two separate tasks. The first task was to drive a real and a simulated vehicle back and forth along a controlled course, obeying speed information either from signs or an instructor. Critically, the instrumentation panel was masked in both the simulator and real-road conditions. The second task involved maintaining a safe—but not predetermined—following distance with respect to a lead vehicle. Additionally, in a separate experiment, 30 non–professional truck drivers completed the simulator portions of the experiment for the vehicle-following task.

The course for the speed control task was divided into five sections, and comparisons were made between the corresponding sections for the professional drivers in the simulator and real-world course. Overall, there was a 0.85 correlation for the average speed in each section between the road and simulator courses. This suggests that even though the drivers had no speed information from instrumental devices, their speed perception and performance was quite in the simulator and real vehicle conditions.

The same course was then used for the lead car following task. Drivers were instructed to maintain an appropriate and safe distance behind the lead vehicle in the simulated and real-world conditions. When values were averaged for all drivers, safety distance from the lead car was approximately twice as large in the simulator compared to the real world test. Note, though, that only professional drivers participated in the real world condition, while professional truck drivers and non–truck drivers drove in the simulator portion. As expected, the difference between simulator and real-world performance was much smaller when only the data for professional drivers were examined, as the non-professional drivers had a 47% greater following distance than professional drivers. The authors suggest that non–truck drivers may have had difficulty adapting to the handling of a truck and the raised viewpoint relative to that of a car. Overall, Panerai et al. (2001) found support for absolute validity of speed for driving simulators, but failed to find support for relative validity for following distance. They attributes this result to the lack of motion parallax cues in their driving simulator, as discussed previously.

Tornros (1998) sought to demonstrate behavioral validity for driving simulators by comparing people’s driving behaviors in a real tunnel versus a simulated one. Again, the goal was to provide converging evidence for relative validity of driving simulators for speed and lane position by comparing driving behavior in the real and simulated vehicles. Both the real and simulated tunnels were free of traffic to allow participants to have maximum control over their environment. Twenty participants participated in both the real and virtual tunnel driving conditions, and they drove through each tunnel twelve times. Speed and lateral position were measured in both conditions. The effects on speed were manipulated by denying participants' access to the speedometer for half of the test runs, and by comparing the speed of participants on various portions of the tunnel where one lane was narrower than the other. Lateral position effects were measured based on which side of the wall was closest to the car. Participants were predicted to position their vehicle farther from the wall when it was on their left and they were driving on the left. Additionally, curvature was believed to affect lateral position, as previous research suggested that people have a tendency to follow curves on the inner side (Harms, 1996).

The data showed that drivers' speed was higher overall in the simulator, which shows a lack of absolute validity, as participants' behavior was different in the simulator and real-world conditions. There was also a difference in speed across lanes, with participants tending to drive slower in the right lane compared to the other two lanes. However, the effect of lane position was equivalent in the simulator and real world conditions with respect to driving speed. Additionally, while participants drove significantly faster when speedometer information was present, this was the same the simulated and real tunnels. Overall, the effects of driving lane and presence of speed information were consistent in the simulator and real life conditions. In other words, individuals who drove in the simulator and real world conditions reacted similarly to the presence or absence of speed information and their current driving lane. Therefore, although absolute validity was not confirmed for speed, Tornros (1998) found evidence for relative validity of speed for this particular driving simulator.

Lateral position was also calculated in this study, as the distance between the middle of the car and the center of the current driving lane. For the straight sections of the tunnel, participants tended to position themselves further from the wall in the real world condition than in the simulator condition. Participants also tended to position themselves further from the wall when it was on their left than when it was on their right, as predicted by previous research. This factor did not interact with driving condition, which is another positive sign of relative validity. Finally, although there was a three-way interaction between driving condition (simulator or real), tunnel wall (nearest wall on left or right side) and curve (left or right), it was quite weak, accounting for less than 1% of the total variance. While this showed that participants drove differently in the simulator and real life conditions, depending on whether the nearest tunnel wall was on their left or right and whether they were navigating a left or right curve, the magnitude was trivial. Overall, Tornros (1998) found evidence for relative validity of lateral position and speed for this driving simulator.

Another method of assessing validity is to have participants perform a secondary task and compare its effects on various conditions. Possibly, if the secondary task affects performance in the different conditions similarly, similar skills may be used across conditions. Santos, Merat, Mouta, Brookhuis, and De Waard (2005) compared performance on a visual search task while participants completed a driving task in three different environments: a driving simulator, a laboratory computer, and a real vehicle. The researchers were interested in how the difficulty level of the visual task affected participants' mean speed and self-reported driving performance in the various environments. The secondary task consisted of identifying whether a target arrow was present in a display of other arrows. The arrow task had three levels of difficulty, with a greater display size increasing the complexity of the task. There were also two baseline conditions, one in which participants completed the driving task without the secondary task, and one where participants performed the secondary visual search task without driving. The aim of the study was to determine if performing the task in either the driving simulator or laboratory computer condition would approximate performance in the real vehicle, thus providing evidence that people behave similarly in real-life and virtual environments.

Overall, the results showed several differences on the performance measurements in the three conditions as a result of increased secondary task difficulty. Participants in both the simulator and real vehicle conditions reported consistent decreases in subjective performance as secondary task difficulty increased, while reported laboratory subjective driving performance was sporadic. A similar pattern appeared for mean speed across difficulty for the secondary task: mean speed tended to decrease in the simulator and real vehicle conditions, while participants in the laboratory condition performed as fast in the most difficult secondary task condition as they did when they did not have to complete the visual search task. Finally, changes in performance across the three difficulty levels of the secondary task were comparable in the simulator and real vehicle conditions. For both conditions, response time for detection of the visual target increased as difficulty increased, while there was no increase in the laboratory condition. Thus, participants' performance in this study tended to follow similar trends in the simulator and field conditions, lending further support to the relative validity of driving simulators.

Driving simulator validation studies show, therefore, that while equal numerical values in performance between simulator and field conditions are rare, variables that affect performance in one condition are likely to have a similar impact in the other. To summarize, individuals in driving simulators and real vehicles react similarly to stop signs and curves, maintain similar speeds even in the absence of speedometer feedback, select similar lane positions in tunnels, and are affected similarly by secondary tasks. Thus, while it would be difficult to accurately predict quantitative results for real-life driving based on a study conducted in a simulator, one would expect to see similar performance trends. Similarly, individuals trained in a simulator on specific driving circumstances would be expected to perform comparably in real life, even if scores on their performance in the two conditions were not exact.

Applications and Uses of Driving Simulators

Driving simulators are valuable for aiding researchers in investigating applied and basic research problems that are too hazardous or difficult to control in real environments. Recently, they have been used to explore the effects of cell phones on driving performance (Schneider & Kiesler, 2005), of conditions that improve performance (Bullough & Rea, 2001), and of devices to help mitigate accidents, such as in-vehicle warning systems (Enriquez, Afonin, Yager, & Maclean, 2001; Lee, Hoffman, & Hayes, 2004). Researchers have also utilized driving simulators to investigate fundamental processes of human attention and visual perception while driving (Readinger, Chatziastros, Cunningham, Bulthoff, & Cutting, 2002) or to test the efficacy of driving aids for individuals with visual impairments (Peli et al. 2005).

Engstrom, Johansson, and Ostlund (2005) studied the effects of visual and cognitive demand on driving performance in real life and in both motion and stationary driving simulators. Their goal was to investigate whether or not the two qualitatively different distractions would have a similar impact on driving performance. They recorded speed, lane keeping performance, eye movements and self-report driving performance. Two types of secondary tasks—the previously described arrow task used by Santos et al. (2005) and an auditory memory task, in which participants had to count and remember a number of sounds—were chosen to maximize the demand on a driver's visual or cognitive load, respectively. Each task had three difficulty levels, as well as a baseline level where no task was given.

The secondary tasks were designed to mimic common real life driving distractions, such as talking on a cell phone or being distracted by a passenger. The visual task can be compared to an individual operating a cell phone and averting his or her eyes from the road. As portable technology becomes more advanced, and more devices become available, people will inevitably spend more time looking away from the road while driving. Thus, it is important to ascertain how operating these increasingly complex devices will affect drivers' performance. Similarly, the cognitive demands of the auditory memory task may be akin to those of holding a conversation with a passenger. Listening is a common distraction for drivers, but its effects on driving performance are rarely taken into consideration. This study sought to determine whether or not visual and auditory distractions affect drivers similarly, and whether one type of distraction is more detrimental to performance than the other.

The results suggested that participants' driving performance was affected differently based on the type of processing required to complete the secondary task. Participants tended to reduce their speed relative to the baseline condition (no secondary task) while performing the more difficult levels of the visually demanding arrow task. This task also affected participants' lateral control of their vehicle and their self-reported driving performance. Again, increased variation in lane position was observed for the more difficult levels. This was true in the driving simulator as well as in the real road conditions, adding further evidence for relative validity for driving simulators. Victor, Harbluk, and Engstrom (2005), who examined eye movement data, noted that participants had longer dwell times in the center of the roadway after completing the visual distraction task in their periphery. The significance of this phenomenon will be discussed later. However, this demonstrates that performing a secondary visual task had an adverse effect on participants scanning behavior.

Conversely, there were no significant changes in driving speed across difficulty levels for the auditory memory task. Somewhat surprisingly, lane position variation also reduced, suggesting that participants actually had better control over their vehicle while performing more difficult memory tasks. Participants continued to report poorer driving performance, however, as task difficulty increased. Finally, eye movement data (Victor et al., 2005) suggested that participants had different eye movement patterns while performing the auditory task than they did following the visual arrow task. In particular, for the auditory task, percent center gaze was not different from baseline in most of the conditions for the various difficulty levels. Participants only showed significantly less horizontal scanning in one of the simulator conditions, suggesting that for the most part, they continued to employ regular, safe scanning behaviors while performing the auditory memory task. Again, this finding is in contrast to the visual task where participants demonstrated reduced scanning practices immediately after completing the arrow task in the periphery. Thus, the authors found differences between the effects of cognitively and visually demanding distractions—differences which were observed both in driving simulators and in real vehicles.

Charlton (2004) used a driving simulator to investigate the effects of certain road signs on drivers' behaviors. Previous studies showed that approximately 90% of drivers either ignore (Chowdury, Warren, Bissell, & Taori, 1998) or do not recall seeing (Drory & Shinar, 1982) various types of road signs. Other research suggests that even drivers who cannot recall having seen a particular sign still respond to it appropriately; for example, they will reduce speed before going around a curve compared to conditions where the sign is completely absent (Fischer, 1992). This study sought to determine which types of road curve warning signs would cause drivers to decrease their speed in accordance with the posted recommendation. Three 45-degree curves were examined, with posted speed suggestions of 45 km/h, 65 km/h, and 85 km/h (1 mile = 1.6 km). Moreover, three different types of road signs were examined, as well as a baseline condition where no warning signs were posted. Signs were defined as perceptually low or high in highlighting the upcoming curve, depending on where they were placed in relation to the curve. Drivers were asked to identify different markings that appeared on the road signs after completing the curve as a measure of how effective the signs were at capturing a person's attention. Finally, at random intervals throughout the scenario, a cell phone noise was presented, and drivers were asked to remember one or five words that followed the second ring.

Charlton (2004) found that the road signs designated as highest in perceptual feedback were most effective at reducing speed in the less severe turning sites of 65 km/h and 85 km/h. Additionally, detection and memory of these signs were less affected by the verbal distraction task than the other signs. While having to remember a set of words, participants demonstrated increased ability to detect and remember the markings on signs high in perceptual feedback, as compared to the other road signs. With regards to simulator fidelity, these results suggest that perceptual information presented in the scene does have an impact on drivers' behavior in the simulator. This is demonstrated by all three road sign conditions producing slower speeds compared to the no-sign baseline condition for the severe 45 km/h curve site. Thus, drivers reacted appropriately to the information provided by the warning signs as if they were driving a real vehicle. Additionally, a general reduction in detecting the road sign markings while performing the cell phone tasks complements other studies that showed the demands of talking on cell phones lower drivers' performance in real and simulated driving conditions (Sodhi, et al., 2002). Therefore, participants in the current study appeared to treat the driving simulator as if it were a real vehicle, and adjusted their speeds appropriately at critical points in the scenario in response to perceptual feedback.

As mentioned above, virtual reality research is often conducted when real world investigations would be prohibitively costly or otherwise impossible. In regards to driving simulators in particular, researchers have begun using results from driving simulators in lieu of real world data. This suggests that they are confident that people behave similarly in driving simulators as they do in real world vehicles, as discussed above. Several investigations have recently been reported in which data from driving simulator studies were ultimately used to make decisions about potential modifications to real roadways.

One such study was conducted by Godley, Triggs and Fildes (2004), who explored the impact of different lane widths on drivers' speed. Lane widths were set at 2.5, 3.0, or 3.6 meters, and the lane markings were either of normal length or extra wide. Past research on real roadways suggests that although tighter space encourages individuals to drive more slowly, the reduced space nevertheless increases the accident rate. The present study sought to alleviate this problem by co-varying the width of the lane markings and the lane itself. For instance, wide lanes can be combined with wide lane markings to simultaneously maintain safe traveling conditions (i.e., wide shoulders) while subjectively reducing drivers' perception of the width of the lane, therefore creating a condition where drivers elect to slow down yet still have a wide safety cushion. Note that altering the width or lane markings of real roads would be expensive and time consuming.

Drivers' average speeds were lower in the narrow road condition compared to the medium road condition, suggesting that drivers decreased their speed when the road was narrower. Similarly, lateral position within the lane was less variable in the narrow road condition compared to the other two conditions, implying that drivers also maintained greater control over the vehicle for narrow roads. Finally, subjective and objective ratings of participants' cognitive workload on the various road conditions were higher in the narrow road condition. Overall, participants' performances in the driving simulator were consistent with their ratings; they drove slower and more controlled on the narrow roadways, which they judged to be more difficult. The importance of this study is that it demonstrated that driving simulator research can be used to enhance understanding of drivers' behaviors in lieu of real world manipulations.

Horberry, Anderson, and Regan (2005) conducted a similar study, using the same driving simulator, to investigate the effects of standard versus enhanced road markings on driving behavior. They created scenarios to simulate night driving on wet roads when visibility of lane markings is poorest. Road markings are particularly important under these conditions as they provide perceptual information about the center and edges of the road when other information is not available (Godley, 1999). Real world studies suggested that the presence of lane markings reduces accidents (Miller, 1992); thus, making lane markings more prominent in low visibility driving conditions may result in fewer nighttime accidents. In this study, an independent panel of road marking experts confirmed the realism and accuracy of the lane markings used in the scenarios. Participants were asked to drive the course with the standard and enhanced lane markings while performing a secondary task half of the time.

The results showed that participants drove closest to the target speed of 100 km/h and were more consistent with their speed in the enhanced lane marking condition. Additionally, drivers' lateral positions' were more constant, and they were less likely to cross the center or edge lines at inappropriate times. Finally, subjective and objective workload measurements revealed once again that drivers preferred the enhanced marking condition. Overall, then, enhanced lane markings were determined to be superior to the standard markings. Again, the significance of this study is that a driving simulator was used to determine the impact of real-life influences on driving behavior, thus demonstrating that experts deem them to have sufficient validity. Furthermore, the results of the study support previous research from real-world driving studies and expert predictions, namely that increasing the prominence of lane markings would positively affect drivers' performance under poor visibility conditions.

Hulst, Meijman, and Rothengatter (2001) conducted a study to assess behavioral indicators of fatigue in driving simulators. Previous studies showed that in a following task, significant deviations between the speed of the lead and following cars were found after approximately 2.5 hours (O'Hanlon & Kelley, 1977). One possible explanation for this finding is that participants' judgment of the lead car's speed was impaired after prolonged driving. Conversely, participants' may have adopted a compensatory strategy for dealing with increased fatigue while driving, specifically leaving a wider gap between their vehicle and the next vehicle.

Additionally, Hulst et al. (2001) sought to determine how reaction time to critical events was affected by fatigue, and whether or not pressure to continue driving would lead to more adaptive performance strategies. The scenario they used was a 32-km circuit with an imposed speed limit of 80 km/h. Participants drove this scenario before and after completing a monotonous driving task of memorizing urban routes. Participants were divided into two groups: an experimental group that was instructed to complete the course in 30 min, and a control group that was not given explicit completion time requirement. A lead vehicle, which participants were instructed to follow, was present approximately half the time during the scenario. At certain instances this vehicle decreased its speed to 55 km/h, either gradually or abruptly and at predictable (heavy traffic) or unpredictable (no contextual cues) times. Headway, steering control, and lane position were measured for these instances. Additionally, once per circuit, drivers approached a vehicle traveling 40 km/h and were unable to pass it due to heavy traffic. Drivers’ responses to the presence of this vehicle were also measured. Finally, before and after the two test drives, participants completed several questionnaires that assessed fatigue, driving aversion, and effort.

Results showed that lane position variation was larger during the last ten minutes of each ride than the first ten minutes for the first and second times through the course, suggesting that steering became gradually worse throughout the course of the experiment and each scenario. Additionally, fatigue ratings were highly correlated with larger variations in lane position and headway between the participant and lead vehicle, although they were also correlated with effort ratings. Thus, as drivers became more fatigued they chose to increase the gap between themselves and the lead vehicle, and even though participants performed more poorly as they became fatigued, they actually reported that they were trying harder. Performance decrements were not associated with reduced effort, therefore, and were seemingly the result of fatigue.

In regards to the car following task, minimum headway between the lead car and the participant was shortest in the unpredictable conditions, showing that drivers were not as fast to react to the lead car's deceleration when its behavior was unpredictable. This implies that participants must have adopted some sort of expectancy for other vehicles in the driving simulator, in that they expected the simulated cars to behave similarly to real cars. This is another indication that driving simulators seem to have high fidelity and promote a sense of presence. Finally, there were several interesting differences between the timed experimental group and the un-timed control group. First, participants in the timed group had reduced headway to the lead car compared to the un-timed group, even when traffic restrictions prevented them from passing. Additionally, this trend continued even when these drivers were fatigued (i.e., in the final ten minutes of the experiment), suggesting that drivers who are in a hurry or on a tight schedule are hesitant to increase safety margins when they do become weary. Thus, the combination of drowsiness and a deadline is precarious for drivers, as they may choose not to rely on compensatory measures that they would normally apply when they become fatigued.

Greenberg et al. (2003) conducted an interesting study using a driving simulator that compared the ability of adult and teenage drivers to detect dangerous vehicles on the highway while distracted with other tasks. Distraction tasks included dialing and answering normal and hands-free cell phones, retrieving voice mail, tuning the radio, and adjusting the temperature controls. Participants were instructed to engage their turn signals when they noticed a driving error made by another vehicle. These vehicles appeared either in the lane next to the lead car or in the drivers' rear view or side mirrors. Participants' lane violations and headway were recorded, along with the percentage of lane violations made by other drivers that they detected.

Several of the distraction tasks resulted in lower detection rates of other drivers' lane violations for teenage and adult drivers, especially for the cell phone tasks. Teenage participants missed over 50% of the violations while performing the phone dialing task, compared to approximately 2% when no distraction task was performed. Thus, it appears that they were directing a significant amount of their attention to the phone task at the expense of monitoring the environment. Combined with the fact that teenage drivers drove almost twice as close to the lead vehicle as adult drivers, this study raises serious concerns for the use of handheld cell phones by teenage drivers. It also adds another example to the growing body of research conducted in driving simulators that would be difficult or dangerous to replicate in real world environments.

Studies that are designed to replicate real world experiments tend to provide converging evidence in favor of the conclusions drawn by past researchers. In other words, participants perform similarly in driving simulators and in real vehicles. The next section discusses evidence of transfer of skills learned in a driving simulator to real driving situations.

Driving Simulators and Training

Driving simulators are commonly used for training for the reasons described above. Specifically, they offer a realistic yet safe environment for students to learn everything from basic driving to advanced vehicle-handling skills. Many studies have looked at different types of training regimens, or ways to maximize training effectiveness such as minimizing the time needed to acclimate individuals to driving in a simulator. Results from a few of these studies that have investigated simulator training are discussed below.

When conducting research on or providing training for a large number of individuals, it is advantageous for sessions to be as short as possible while remaining effective. One way of minimizing training length when using driving simulators is to make the simulator acclimation portion of training as short as possible. McGehee, Lee, Rizzo, Dawson, and Bateman (2004) sought to determine the shortest amount of time needed for participants to adapt to the steering mechanisms in the driving simulator, and to determine if there were differences between older and younger drivers. To test this, they analyzed drivers' lane position and steering wheel deviation at three specific intervals during a 25-minute driving scenario.

As expected, all drivers showed a reduction in steering variation in the final segment compared to the first two, suggesting that their handling of the simulator improved over time. Similar results were obtained for lane position, with all drivers also showing a reduction in the variability of their lane position across segments. Age differences, though present, were relatively minor. No differences were found between young and old drivers in steering or lane position variability for the first or second segments. The only difference was found in the last segment, where older drivers tended to show more variability than younger drivers.

These results suggest that older drivers take longer to become acclimated with driving simulators than younger drivers. This familiarity process was generally complete, however, within five minutes. Some drivers even showed normal amounts of variability after two minutes in the driving simulator. While absolute adaptation time for this simulator may not necessarily generalize to other driving simulators, these data do suggest that drivers adapt quickly to the handling of driving simulators. McGehee et al (2004) stated that variables such as high simulator fidelity and number of physical perturbations in the simulator, which encourage individuals to regard the simulator as a real vehicle, may also reduce this adaptation period. Overall, drivers appear to adapt to driving simulators relatively quickly.

Ivancic and Hesketh (2000) conducted an interesting study on the effects of different types of training on individuals' performance in a driving simulator. In particular, they compared two types of strategies: error training (where participants actually made errors) and guided error training (where drivers explicitly learned from other peoples' errors). These training types were compared to errorless training, in which participants completed a driving scenario that was designed not to elicit errors or provide feedback of their performance. In error training, trainees perform difficult activities that invariably lead to mistakes, and they are encouraged to actively learn from instructor feedback regarding their errors. Previous research suggests that this leads to better generalization from training to performance in the field (Ivancic & Hesketh, 1995). Guided error training, on the other hand, provides systematic and controlled feedback to all participants. It may also be less prone to reducing participants' motivation (Ivancic, 1997).

Training and testing were conducted in a driving simulator (Ivancic & Hesketh, 2000). Experiment 1 compared error training to errorless training, while Experiment 2 compared guided error training to errorless training. Each participantcompleted two training scenarios. The first training scenario was designed to familiarize participants with the simulator controls and provide feedback, in the error conditions, on their performance. The second scenario was the test scenario and it contained six critical events that were identical in both experiments. Five of these critical events were analogous to conditions that drivers saw during training. For instance, in both the training and testing phase, drivers encountered a situation where the safe maneuver was to stop and let other vehicles pass before steering around an obstacle. Thus, Ivancic and Heketh were interested in seeing which training regimen led to the highest rate of transfer to the test scenario, exemplified by fewer crashes or offenses in the second scenario. In the error training condition in Experiment 1, participants were either given a ticket or caused a collision when they used an incorrect strategy at a critical event. The errorless group, on the other hand, did not receive any sort of feedback after making a poor decision. In Experiment 2, the guided error group watched a video of drivers receiving negative feedback when they made an error, while the errorless group watched the same video without the inclusion of driver feedback. Thus, in the guided error training, all participants viewed the same mistakes and were given similar feedback, while those participants in the error training group only received feedback when they made an error.

For Experiment 1, the error-notification group made fewer errors during the test scenario than the corresponding errorless control group. They also drove significantly slower as they approached and maneuvered around obstacles. The difference in number of errors between the guided error group and control group in Experiment 2 was smaller, suggesting that guided error training was not as effective at reducing participants’ errors and that those drivers' strategies did not transfer as well from training to test phases. Additionally, no speed differences were found between the two groups. Overall then, error training, in which participants were not discouraged from making errors and instead received individual feedback based on their performance, appeared to be the most effective training strategy. Participants who received error training made fewer critical errors and showed greater speed reduction in hazardous environments during the test scenario than participants who received errorless training. Also, although reported self-confidence following training was lower for the error group than the errorless group in Experiment 1, this trend reversed after the test phase. In other words, although the error group seemed more discouraged than the errorless group immediately after their first drive, they were apparently more encouraged by their superior performance in the test phase, and thus did not show any adverse effects of receiving negative feedback. Therefore, training methodologies that require the student to perform challenging tasks with feedback appear to be more effective than not providing feedback or requiring trainees to watch other drivers make mistakes.

Eye and Head Movements

Several studies have used eye and head movement data to investigate various aspects of driving performance as participants operate driving simulators and real vehicles. Typically, the research focuses on two components of driving: (1) the effects of distractions and (2) the characteristics of adaptive driving behavior. The latter line of research refers to determining effective search strategies used by experienced drivers in distraction-free conditions to examine the environment around the vehicle. Consequently, it is often beneficial to determine what effects, if any, distractions, instructions, or training have on drivers' eye movements. For example, as discussed previously in Victor et al. (2005), eye and head movements can be used to assess the impact of secondary tasks on driving performance.

Sodhi et al. (2002) examined drivers' eye movements on a real world course while they completed various distractive tasks. Eye movements are considered useful for investigating driving behavior, as they are a good indicator of the processing that occurs for a given task. In other words, they provide an indication of where an individual's cognitive resources are allocated, and thus can be used to measure the impact of distractors. Previous research suggested that drivers typically employ a time-sharing method to monitor all necessary vehicle devices (Wierwille, 1993). Drivers tend to focus mainly on the road in front of them while periodically glancing towards areas in their periphery (e.g. rear view mirror, side mirror, speedometer, and either side of the vehicle). Generally these checks are limited to 1.6 seconds or less to allow the eyes to return to the area in front of the vehicle where hazards are most likely to occur. However, more complex tasks, such as dialing a cell phone or finding a specific radio station, may require additional time on task processing. In other words, they may require participants to fixate for a longer a period of time away from the center of the road, increasing the likelihood that the driver may miss a critical event. Sodhi et al. investigated the impact on eye movements of performing several common tasks that require the driver to avert his or her eyes from the center of the road: changing the radio, glancing at the rear view mirror and odometer, and talking on a cell phone.

Participants appeared to utilize the time-sharing method in the radio, rear view mirror, and speedometer tasks. Typically, they would glance towards the device for a short period of time before returning their gaze to the road ahead. This procedure would continue until the task was complete (i.e., the radio was set correctly or the driver was accurately able to report his or her speed). A different pattern of eye movements was observed for the cell phone task, however. Drivers tended to fixate only on the middle of the roadway while talking on the cell phone, a condition known as visual tunneling, where the useful field of view is reduced. Interestingly, this pattern of reduced peripheral glances persisted even after the participant hung up the phone. Redelmeier and Tibshirani (1997) attributed this tendency to people maintaining afterthoughts of the phone conversation, as if they were rehearsing or replaying the dialogue. Thus, Sodhi et al. (2002) concluded—along with Engstrom et al. (2005)—that cognitive and perceptual tasks affect drivers' scanning patterns differently.

Campagne, Pebayle, and Muzet (2005) examined the effects of prolonged driving and fatigue on drivers' eye movements and blink frequency in a driving simulator. Previous research suggested that reduced blinking is associated with the performance of more difficult tasks and that attention is oriented to critical stimuli in the scene (Drew, 1951; Veltman & Gaillard, 1996; Wilson, 1993; Goldstein, Bauer, & Stern, 1992). Campagne et al. sought to determine whether drivers' blink patterns changed as a function of the amount of time they spent driving, and whether critical events in the environment would restore typical blinking patterns. The number of fixations that participants made on the speedometer was also measured. Participants drove the same 50-km circuit in a moving-base simulator five times to induce boredom. Each lap contained 18 road signs and seven critical events that were important for proper navigation of the vehicle. To measure the impact of these events on blinking, three periods were defined: a main period extending from the time the event was first perceived until the driver passed the obstacle, and two other periods comprising an equal amount of time preceding (pre-period) and succeeding (post-period) the main period.

The results showed that blink frequency and duration increased significantly with the number of laps. Thus, as drivers became more accustomed to the course, and presumably became more bored with the task, they tended to blink more. Moreover, participants made fewer glances to the speedometer as the amount of time they spent in the vehicle increased, suggesting a reduction of attention to in-vehicle conditions. However, in regards to the restoration of normal blink patterns to certain stimuli, the data showed a mixed pattern of blink frequency based on the type of critical event. For instance, after the first lap, participants showed an increase in blink activity when they encountered either a moving vehicle or certain road signs, suggesting that they more or less ignored them. However, they did continue to decrease their blink activity during the speed limit sign and truck stopped in emergency lane events. This suggests that certain events can restore drivers to a heightened state of vigilance even after prolonged driving. Still, fatigue did have an effect on participants' blinking behavior, especially during periods of relative monotony. Note too that this study was conducted in a motion-based driving simulator that has the majority of visual and haptic cues found in real cars.

Novice Versus Expert Drivers

Recently, Underwood and his colleagues (Crundall & Underwood, 1998; Crundall, Underwood, & Chapman, 1999; Underwood, Crundall, & Chapman, 2002; Crundall, Underwood, & Chapman, 2002; Underwood, Chapman, Bowden, & Crundall, 2002; Chapman, Underwood, & Roberts, 2002; Underwood, Chapman, Berger, & Crundall, 2003; Underwood, Chapman, Brocklehurst, Underwood, Crundall, 2003; Crundall, Shenton, & Underwood, 2004) have conducted extensive research, using behavioral and eye movement data, on the effect of experience on driving performance. Although they primarily tested participants using real vehicles or while they watched video clips of other people driving, their results are applicable to predicting novices' and experts' driving performance in simulators.

Crundall and Underwood (1998) began by investigating novice and experienced drivers’ eye movements as they drove three different roads: rural, suburban and urban. Similar categories were used in subsequent studies. Urban roads contained a higher volume of traffic, and therefore were designated as higher in cognitive demand than the other two roadways. Suburban roads were classified as more challenging than rural roads. The authors hypothesized that as driving conditions became more difficult, novice drivers would undergo perceptual narrowing and focus more of their attention, as measured by eye fixations, at the center of the road. Critically, as this window of attention narrows, drivers receive less information about events occurring in their periphery. Additionally, novice drivers may show a tendency to fixate more often on lane markers if they are unable to maintain lane position by using information from their mirrors or extract enough information from their periphery while glancing forward. This would suggest that novice drivers have more difficulty operating a vehicle than expert operators, which may account for the fixation differences.

The results showed that experienced drivers had a wider horizontal and vertical search on urban roads than the other two roads. Novice drivers, on the other hand, did not show any differences in horizontal or vertical search variance for any of the conditions. This suggests, therefore, that experienced drivers compensated for the more demanding urban roads by increasing their spread of search. This strategy would be quite effective for anticipating potential hazards from peripheral locations, which would naturally increase for heavily populated environments. Novice drivers, on the other hand, showed little difference in search variation across the three road types, suggesting that they were not using the expert drivers' strategy of scanning as much of the roadway as possible.

Crundall and Underwood (1998) found that novice drivers' mean eye fixation durations were longest on the most demanding road (urban). They pointed out that longer fixations, similar to fewer blinks, are typically associated with extra processing. Thus, novice drivers required more time to process information on the urban roads. However, long fixations limit the amount of scanning that an individual can accomplish. While novice drivers appear to take more time processing potentially hazardous information on the challenging roadways, they consequently perceive less information from other areas in the scene, particularly in the periphery. Experienced drivers, on the other hand, compensate for the increased demands of urban roads by attending to a greater amount of information, as evident by shorter fixation lengths. Of course, this strategy may only be effective because experienced drivers can better predict potential hazards, or have an easier time processing information based on their superior understanding of driving environments. Overall, this suggests that experienced drivers tend to adapt their scanning behavior to the complexity of the road, while novice drivers are too inflexible or inexperienced to alter their scanning behavior at appropriate times.

Crundall, Underwood, and Chapman (2002) investigated whether similar results could be obtained in a laboratory setting. Novice and experienced drivers were instructed to watch short clips of other drivers from the perspective that they were the driver. Each clip contained one to four hazardous events, and participants were instructed to respond whenever they noticed a potentially hazardous event. Additionally, participants were asked to detect peripheral targets that occurred once during every five second segment of a clip at eccentricities of less than 5, 5-6, 6-7 or greater than 7 degrees of eccentricity from their current fixation point. Each video segment was classified as high or low in demand depending on the number of hazard responses elicited from participants during the five second segment. For instance, a target appearing 6.5 degrees from the participant's current fixation location would be classified as more difficult if two hazards appeared within the five-second window of its presentation than if only one hazardous situation occurred. Critically, the task of driving the vehicle was eliminated in this study. Thus, if any differences in performance were found between expert and novice drivers, they would likely be due to a poor understanding of ideal scanning behavior and not due to different levels of cognitive demand for novice and expert drivers.

Overall, participants detected fewer peripheral targets during high-demand clips versus low-demand clips. Also, percentage of target detection was significantly lower for peripheral targets presented farther than 7 degrees from participants' fixation, compared to all other onset eccentricities. Novice drivers detected fewer targets overall than experienced drivers, and this did not interact with eccentricity or level of demand. In other words, novice drivers were poorer than expert drivers at identifying targets at every distance from their current fixation for each level of difficulty. This suggests that inexperienced drivers have inferior performance regardless of the level of demand or eccentricity of target. Interestingly, though, there was no interaction between task difficulty and eccentricity, suggesting that the different levels of demand had the same impact on identifying targets at each eccentricity. Thus, the drop in performance as targets were presented at farther eccentricities was similar in both the high-demand and low-demand conditions. Consequently, no evidence was found for drivers' adopting a tunnel vision strategy on the higher demand tasks, which would be evident by a severe decline in target detection on the further eccentricities in the high demand compared to the low demand condition.

Additionally, inexperienced drivers took longer to respond to the presence of peripheral targets than experienced drivers. Since the differences in performance cannot be attributed to tunnel vision or to novice drivers allocating more cognitive resources to driving than experienced drivers, these results imply that novice drivers do not have a fully developed understanding of driving conditions. This pattern of result was replicated by Crundall, Underwood, and Chapman (1999) using a less attention-demanding primary task, where participants had to rate the danger and difficulty of driving through various scenes.

In a related study, novice and expert drivers' eye movements were tracked as they watched video clips taken from urban, rural, and suburban roads (Underwood, Chapman, Bowden, & Crundall, 2002). Again, participants were not required to operate the vehicle and only needed to direct their eye movements to whatever they thought were the most important locations. Thus, if there were differences between fixation locations of novice and expert drivers, it would likely be due to novice drivers having a lesser understanding of critical areas in the environment to look at. Each group of drivers also completed a separate questionnaire regarding either the location in the scenes that they thought they were looking at the most (experts), or what they believed experienced drivers tended to look at the most (novices).

Glance fixation duration for both groups of drivers was found to be longer on the least demanding video clips, specifically, the rural road condition. Also, no differences in fixation duration were found based on drivers' level of experience, suggesting that novice and expert drivers used similar fixation timing strategies for investigating the scenes. However, the overall horizontal scan variance was smaller for novices than expert drivers, implying that they made fewer glances to the periphery. This difference was most pronounced for the highly demanding conditions. Additionally, questionnaire responses showed that experienced drivers actually underestimated the number of glances they made to the various objects in the different scenes, suggesting that they were more cognizant of the environment than they thought they were. Novice drivers also underestimated the amount of time that expert drivers looked at the critical objects in the scene.

Overall, these data suggest that even when novice drivers do not have to control a vehicle, they show qualitative differences in their scanning patterns compared to expert drivers. Specifically, they tend to fixate less on objects or vehicles in the periphery, especially under demanding or hazardous conditions. The results from Crundall and Underwood (1998) and Underwood, Chapman et al. (2002) also suggest that this may result from novice drivers' poor understanding of proper scanning behavior. Not only were novices' scan paths different from experienced drivers when the cognitive demands of driving the vehicle were removed, but novice drivers also underestimated the number of glances that experienced drivers made to objects outside their vehicles. Taken together, these studies suggest that novice drivers are less aware of events occurring in the periphery than are experienced drivers. These findings are consistent with the hypothesis that novice drivers have not developed an appropriate driving scheme of scanning all areas of their environment. Instead, novice drivers appear to fixate to a greater extent on the area directly ahead of them at the expense of information in their periphery.

The critical question, then, is whether it is possible to train novice drivers to use more appropriate search strategies, or if proper scanning while driving is a skill that can only be learned through real world experience. Chapman, Underwood and Roberts (2002) devised a training intervention to inform novice drivers of their typical scanning patterns and to encourage them to implement a more adaptive strategy. Two groups of novice drivers were tested on three occasions during their first year of driving independently. The training intervention, administered to the experimental group immediately before the second test, involved tracking participants’ eye movements as they drove on real roads and while they watched video clips of hazardous situations. The control group did not receive training, and instead simply completed a questionnaire during that time. The authors sought to improve three key factors in novices’ cognitive driving strategies: knowledge, scanning, and anticipation. Critically, the purpose of the training was not simply to demonstrate more adaptive search techniques, but also to teach the skills and strategies implicit in effective scanning techniques. Successful training would be identifiable by reduced fixation times and greater scanning variances, implying that novice drivers were able to process information faster and gain more information from their environment.

Participants were first tested (phase 1) immediately after passing their driving test. The training intervention (phase 2) occurred approximately three months later, where participants were divided into the experimental group (received training) and a control group (no training). After receiving the actual training or answering questions, both groups completed tasks similar to those from the first session. The final phase of testing (phase 3) took place three to six months after the training intervention, allowing for an investigation of the long-term effectiveness of training.

Overall, there were no differences of average speed for the experimental and control group across the different phases of training, suggesting that the intervention did not affect drivers’ speed. However, mean speeds in all conditions did not exceed posted speed limits by more than 5 mph, implying that these drivers did not show a tendency to speed even in the pre-training condition. On-road eye movement measurements appeared to change, however, as a result of training. Immediately after training, drivers in the experimental condition showed a greater horizontal spread of search than control drivers who did not receive training. This effect disappeared in the final testing phase, though, where there was no difference between trained and untrained drivers' horizontal scanning in the road courses. This suggests that although training caused novice drivers to change their scanning behavior right after they received training, this pattern did not persist until the last phase of testing. However, novice drivers successfully demonstrated wider horizontal scan patterns while watching the video clips both immediately after training and during the follow-up testing phase, compared to the untrained drivers in the control condition. Therefore, while the cognitive demands of driving may continue to adversely affect novice drivers' eye movements even after receiving training, the fact that they demonstrate wider scanning patterns while viewing the video clips suggests that they did benefit from the intervention. In other words, novice drivers apparently retained knowledge of adaptive scanning techniques, even though they were not able to demonstrate this on the real world course.

Other studies have specifically looked at differences in performance in a driving simulator between expert and novice drivers in a particular domain. Dorn and Barker (2005) investigated trained police officers compared to non-police drivers on two tasks: overtaking a slow-moving bus in a rural environment and trailing a fast-moving lead vehicle in an urban environment. In particular, researchers were interested in looking at variables—other than reduced accident rates—that demonstrated improvements with regards to training; accident involvement was not considered, since it is typically not fully under the driver's control.

Moreover, given that the police group and the control group had notable differences in training prior to this study, another topic of interest was whether these differences would manifest themselves in a driving simulator. In other words, would differences in real-life training be reflected in simulator driving performance?

Non-police drivers were significantly more likely than police drivers to overtake the bus at unsafe locations (e.g., at double yellow lines). Thus, police drivers showed more restraint than the control group of drivers and tended to pass the bus at safe, legal opportunities. For the car following task, police drivers tended to drive closer to the center division between forward and oncoming traffic, suggesting that they utilized available lane space differently than civilian drivers. Speed differences were also found during that task. At a critical point, when a bus was parked in the right lane, police drivers drove significantly slower than non-police drivers. Given these two differences in police and civilian drivers during the car following task, it seems that police drivers were more cognizant of other vehicles on the highway (as demonstrated by slowing down for the parked bus), and they also used their training when selecting an appropriate lane position. The former was also demonstrated during the rural drive, when trained police officers exercised more discretion when passing a slow-moving bus. Interestingly, Dorn and Baker (2005) reported that this strategy mimics what is taught to recruits during actual police training programs. Thus, instructions that police offers were given during real-life training were reflected in their performance in the driving simulator task.

Utah DOT Snowplow Study

The Utah Department of Transportation (UDOT) recently conducted a study, along with the University of Utah, to test the effectiveness of snowplow simulator training on operators' driving performance during actual plowing (Strayer, Drews, & Burns, 2004). The investigators designed a training program to instruct drivers in fuel management, proper scanning techniques, shifting techniques, and space and speed management. Forty current UDOT snowplow operators received approximately four hours of training. Additionally, participants completed a questionnaire about the quality and usefulness of simulator and classroom training. Their driving performance during the subsequent winter season was compared to an additional set of 40 operators, matched with the experimental group in age, years with a driving license, experience operating a snowplow, and experience driving a truck. In particular, the study examined differences in the two groups' accident rates, including number and severity, as well as fuel efficiency. The authors hypothesized that simulator training would lead to a reduction in the number of accidents and amount of fuel consumed by drivers.

The questionnaire data suggested that drivers found the simulator and classroom training very useful, and thought that training should be mandatory for all UDOT snowplow operators. This was true for operators at all levels of experience; that is, experienced operators found the simulator and classroom training to be just as useful as novice operators did.

Accident rates were relatively low during the six-month winter season, with three accidents reported for operators in the experimental condition and four accidents for operators in the control condition. Moreover, two of operators who were involved in an accident in the experimental condition were determined not to be at fault, and thus their accidents were disregarded in the final analysis. Consequently, the results of the experimental group (who received the simulator training) approach a statistically significant number of fewer accidents than the control group, which suggests that training was effective at reducing the number of accidents. Additionally, the accidents reported for the control drivers were more severe than the one at-fault accident of the driver in the experimental group. Thus, it seems as if snowplow simulator training did have a positive impact on the performance of operators. However, to achieve statistically significant results, the authors concluded that approximately 20 more participants were needed in each group.

Finally, fuel and maintenance costs were compared for the control and experimental groups. Although there were difficulties in obtaining precise data for each participant, the experimental group showed a 6.2% improvement in fuel efficiency compared to the control group. These data, as well as similar results reported by Strayer and Drews (2003), suggest that simulator training can also lead to a measurable improvement in fuel efficiency. Overall, Strayer et al. (2004) concluded that not only did operators rate simulator training as relatively positive, but that virtual reality simulator training appeared to cause a positive improvement in operators' driving performance and fuel consumption during the subsequent winter.

Cybersickness

One of the major drawbacks of virtual reality environments is their tendency to cause discomfort in some individuals, a condition known as cybersickness. Studies have reported (Stanney & Salvendy, 1998) that up to 95% of participants experience some form of cybersickness, and as many as 30% of those participants elect to end participation early. Cybersickness tends to mimic symptoms of motion sickness, with the paradox being that individuals are typically stationary in these environments. However, the physiology of the visual system and the vestibular system (which helps control balance, movement, and orientation and is located in the inner ear) create a sense of self-motion for individuals in high-fidelity simulators, resulting in the perception of motion. The obvious consequence of implementing a simulator training program that causes cybersickness in the trainees is that students will either not benefit from it or will simply refuse to participate in that portion of training. Moreover, these effects can last for hours (LaViola, 2000), potentially affecting the trainee when he or she leaves the training facility. Thus, it is in the best interest of the trainer to ensure that all precautions are taken to avoid cybersickness.

There are several theories as to the cause of cybersickness, such as the sensory conflict theory, the poison theory, and the postural instability theory. The sensory conflict theory stipulates that cybersickness is the result of conflicting inputs from the vestibular and visual systems. Basically, the student experiences motion from the optic flow patterns of the environment, resulting in a sense of vection (i.e., an illusional experience of motion) from the visual system. However, due to the fact that the individual is not actually moving, he or she receives a conflicting message from the vestibular system. In other words, the visual system experiences motion while the vestibular system maintains that the participant is stationary. This discrepancy between the two neural systems results in cybersickness. From a theoretical standpoint, however, this explanation has difficulties, as it does not explain why some individuals experience cybersickness while others do not, and it does not account for why such a conflict would necessarily result in feelings of discomfort.

The poison theory attempts to explain cybersickness from an evolutionary perspective. It suggests that ingesting poison often affects sensory systems, such as the vestibular system, and an adaptive strategy of combating the intake of poison is to vomit. Consequently, when the above conflict arises between the visual and vestibular system, the brain misinterprets the source of the discrepancy and responds as it would if the individual digested poison. However, similar to the sensory conflict theory, this explanation cannot clarify why some individuals experience cybersickness while others do not.

Finally, the postural instability theory states that humans intrinsically attempt to maintain postural stability in their environment. Proponents of this theory hypothesize that virtual reality environments produce prolonged postural instability. The longer an individual is immersed in such an environment, the more intense the person's discomfort will become. While this explanation has similar faults to the previous two theories, it does accurately account for the finding that feelings of cybersickness are highly correlated with the amount of time spent in the virtual environment.

There are, however, several contributing factors to cybersickness that can be manipulated to reduce discomfort. One commonly cited cause of cybersickness is lag between participants' actions and the updating of the visual display. In driving simulators, for instance, if the driver turns his or her head and body, along with the steering wheel, as part of a sharp turn, but there is a delay in registering this command in the simulator, the user will have to wait for the vehicle to properly respond. This delay between expectation and execution, especially when head movements are involved, can cause cybersickness. Additionally, screen flicker, which occurs when the refresh rate of the monitor is not fast enough, can exacerbate cybersickness. With advances in virtual reality technology, however, this problem has become less common as visual displays have advanced enough to eliminate perceived flicker.

Adaptation to the virtual environment is also important in reducing cybersickness. Turning especially should be integrated into the simulation gradually. Several researchers have also suggested that providing individuals with rest frames may reduce cybersickness

(LaViola, 2000; Duh, Parker, & Furness, 2004). A rest frame is any object that an individual perceives to be stationary (LaViola, 2000) and can aid people in determining which other objects in the environment are stationary and which are in motion. People who have difficulty identifying a rest frame in a virtual environment are more likely to experience cybersickness. Duh et al. (2004) found that the rest frame—in this case a checker-board wall—could even be presented behind the simulated stimuli and still reduce feelings of cybersickness. Thus, it may be advantageous for trainers to indicate stationary parts of the driving simulator, such as the panels in between the screens or the top and bottom portions of the driving simulator, to trainees who are experiencing cybersickness.

Finally, Rizzo et al. (2003) investigated the effects of braking and steering on cybersickness. Participants drove an uneventful rural scenario for up to 30 minutes, which was interspersed with several critical events that could result in a collision. Participants who dropped out from symptoms of cybersickness were matched with an equal number of participants who managed to complete the study. Interestingly, participants who dropped out of the study used the brakes significantly more frequently than the matched sample, although no steering differences were found. One interpretation of these results is that, especially during the simulator adaptation or familiarity period, superfluous braking requirements should be minimized as much as possible to reduce the likelihood of inducing cybersickness.

One commonly used measure of simulator sickness in virtual environments is the Simulator Sickness Questionnaire (SSQ) developed by Kennedy, Lane, Berbaum, and Lilienthal (1993). The SSQ consists of 16 questions that comprise three subscales: nausea, oculomotor discomfort, and disorientation. Participants respond from “none” to “severe” (on a scale of 0 to 3) for each question, and total scores from three subscales are summed, and multiplied by 3.74 to find the total score. The average level of simulator sickness necessarily differs based on the virtual environment and factors that affect simulator sickness. For instance, So, Lo, and Ho (2001) used head-mounted virtual reality displays to simulate navigating a vehicle through a city. Their participants reported simulator sickness ratings of up to an average of 60 for the fastest speed condition. This would be the equivalent of responding “slight” (i.e., a score of 1) for every question. Similarly, Arms and Cerney (2005) reported simulator sickness scores of approximately 25 for participants between the ages of 28 and 60 in their immersive virtual environment. Even though these scores seem low, approximately 90% of their participants reported experiencing some simulator sickness. Incidentally, this finding provides converging evidence for Stanney and Salvendy’s (1998) finding that the vast majority of individuals report some feelings of simulator sickness after being exposed to a virtual environment.

Personality

Within the past 100 years, personality researchers have set out to develop accurate, reliable, and frugal methods of classifying individuals based on certain personality characteristics. Some of these tests are relatively specific (e.g., Sensation Seeking), while others attempt to succinctly describe an individual’s entire personality (e.g., Big Five personality factors). While researchers are cautious when interpreting the results of these tests (McCrae & Costa, 2003), significant advances have recently been made in agreeing upon a fundamental set of factors underlying personality. Studies of personality trait research tend to focus on one of two different factors: ascertaining whether these measurements are reliable and valid, or determining the predictive power of these tests in real-life domains.

The two main measurement systems described here, the NEO Five Factor Inventory (NEO-FFI) and Zuckerman's Sensation Seeking scale and personality scale, have undergone several updates through the years and have been well-corroborated in the literature (Borgatta, 1964; Hakel, 1974; Zuckerman, 1979; McCrae & Costa, 2003) in a wide range of domains. However, there has been extensive disagreement concerning the applicability of these measurements. For instance, Guion and Gottier (1965), after conducting a meta-analysis of personality measures used in personnel selection, stated that, "It is difficult in the face of this summary to advocate, with a clear conscience, the use of personality measures in most situations as a basis for making employment decisions" (p. 160). More recent research, on the other hand has offered cautious support for the use of personality measurements for employment decisions (Barrick & Mount, 1991; Tett, Jackson, & Rothstein, 1991).

The Big Five

The development of the Five-Factor Model (FFM) and the NEO-FFI personality questionnaire that is commonly used to measure it were based primarily on the work of Costa and McCrae (McCrae & Costa, 1985; Costa & McCrae, 1985). They were inspired by the work of several researchers who employed a technique called "factor analysis" to determine personality characteristics, or traits, that are highly associated with one another. One of the biggest drawbacks to using traits as a basis of personality measurement is that researchers have identified as many as 18,000 (Allport & Odbert, 1936) potential personality traits; basically, factor analysis can be used to group together the traits that are highly correlated with each other, until a workable number of broad, discrete dimensions remain. Then, these domains that represent the combination of many highly correlated traits are examined by personality tests.

There are five dimensions that are explicitly measured by the NEO-FFI: extraversion, emotional stability (neuroticism), agreeableness, conscientiousness, and openness to experience (McCrae & Costa, 2003; Barrick & Mount, 1991). Each dimension can be thought of as a continuum, with its title representing one of the end points. For instance, the extraversion scale is sometimes labeled as its antithesis, introversion. Extraversion is associated with being social, assertive and talkative, and it is thought to consist of ambition and sociability components (Hogan, 1986). The second dimension, emotional stability, is also commonly referred to as neuroticism, the other end of the continuum. Anxiety, depression, insecurity, and the tendency to be emotional are traits associated with this dimension. The third dimension, agreeableness, is associated with cooperation, flexibility, trust, and tolerance. The fourth dimension, commonly known as conscientiousness, is not as agreed-upon. Researchers tend to accept, though, that it has some relationship to being dependable, hardworking, compliant, and persevering. The final dimension is the most disputed characteristic, and its name differs based on the traits used to define it. Costa and McCrae (1985) termed this dimension “openness” and suggested that traits such as being imaginative, creative, cultured, and original are related to this characteristic. Table 1 shows the means and standard deviations for each domain as reported by Rolland, Parker, and Stumpf (1998) in their study of 500 normative American males, and it presents the means from Saucier (1998) on his study of 732 American men and women. Note that scores are on a scale from 0 to 48.

Table 1.

Normative means and standard deviations for the NEO-FFI.

Authors	Mean and Standard Deviations by Domain Neuroticism	Extraversion	Openness	Agreeableness	Conscientiousness
Rolland, Parker & Stumpf (1998)	17.6 (7.46)	27.2 (5.85)	27.1 (5.82)	31.9 (5.03)	34.1 (5.95)
Saucier (1998)	18.0	26.7	23.8	33.8	33.9

Once a set of domains is determined, however, it is important to ascertain whether the questions used to measure these domains are reliable. The most common measurement of reliability is test/re-test reliability, where people's scores on the same test are compared at different times. Correlations for the separate domains on the NEO Personality Inventory (NEO-PI), a longer version of the NEO-FFI, across administrations of the test, even over several years, have been found to be as large as 0.70 to 0.85 (Costa, Herbst, McCrae, & Siegler, 2000; Costa & McCrae, 1988; McCrae & Costa, 2003). These results suggest that the test itself is quite reliable and also imply that personality is relatively stable over time. Still, some might question the use of self-report tests as a valid measure of an individual's personality. In particular, it seems as if test takers may lie to make themselves appear more socially desirable. However, McCrae and Costa (2003) reviewed several studies which concluded that trying to correct for these types of responses does not improve the validity of participants' scores and may actually impair it (Dicken, 1963; McCrae & Costa, 1983; Piedmont, McCrae, Riemann, & Angleitner, 2000). Thus, they recommend accepting individuals' self-report scores as long as they have no motivation to fake them.

The Big Five and Job Performance

As mentioned above, there has been some reluctance to utilize personality tests as indicators of job performance. This was mainly due to relatively weak correlations between job performance ratings and personality scores. Recently, however, two literature reviews that examined the correlations between personality scales and job performance demonstrate that improvements have been made in personality battery construction. Barrick and Mount (1991) examined 117 studies that specifically compared the relationship between job performance and responses to the Big Five personality test.

Barrick and Mount (1991) categorized the professions studied in these articles into five occupational groups: professionals, police, managers, sales, and skilled/semi-skilled workers. Truck drivers, for instance, were a component of the skilled/semi-skilled group. Additionally, three types of data categories were identified as part of job performance: job proficiency, training proficiency and personnel data. Job proficiency referred to performance ratings and employee productivity, training proficiency was comprised of training performance ratings, and personnel data included salary, turnover, and status information. Thus, this meta-analysis comprised a wide variety of occupations and job performance measurements, while focusing on a set of relatively specific personality measurements.

Overall, the conscientiousness dimension was the best predictor of job performance across all occupations. Although the correlations were relatively low, between 0.20 and 0.23, they were very consistent across occupational groups. Moreover, conscientiousness scores were also correlated with all types of categories that were said to comprise job performance. Thus, the dimension of conscientiousness appears to be an important factor for predicting people’s job performance across a wide variety of professions and employment functions.

Additionally, openness and extraversion were also positively correlated with training proficiency, suggesting that individuals who tend to be outgoing and accepting of new experiences respond better to training opportunities.

Surprisingly, the emotional stability/neuroticism dimension in particular was not found to be positively correlated with job performance in this meta-analysis. Based on some of the traits it is hypothesized to comprise (e.g., anxiety and depression), emotional stability seems to be an important trait for success in the workforce. However, the authors suggest that a “selecting out” process (Barrick & Mount, 1991, 20) may have occurred on this dimension. Specifically, working individuals may require a minimal amount of emotional stability to secure a job, be productive, and retain it. People high in neuroticism may not be able to function in the workforce, and thus their job performance would not be gauged by these studies. Consequently, as long as someone has enough emotion stability to hold a job, additional variance in this domain does not appear to be predictive of job performance.

Tett et al. (1991) also conducted a meta-analysis of the job performance and personality literature to assess the validity of using personality measurements, such as the Big Five, as predictors of job performance. Their analysis covered 86 studies conducted between 1968 and 1991. Unlike Barrick and Mount (1991), Tett et al. found evidence that emotional stability, openness, agreeableness and conscientiousness were all moderately correlated with job performance, with correlations ranging from 0.18 for conscientiousness to 0.33 for agreeableness. Thus, this study adds converging evidence that personality tests are becoming better at predicting job performance. Differences in results between Tett et al. and Barrick and Mount can be attributed to different selection criteria for the studies that comprised their respective meta-analyses and different classification and weighting techniques to determine correlations. Both studies agreed, though, that the Big Five personality test can be used as a moderate predictor of job performance.

Additional evidence for the predictive power of the Big Five personality test comes from its high correlation to the Holland RIASEC (Realism, Investigative, Artistic, Social, Enterprising and Conventional) vocational interest model, also called the Big Six Interests (Holland, 1985). Holland's personality test was designed to match a person's interests with an appropriate category of vocations. His theory of vocational interest states that interests are an expression of an individual's personality and that people tend to select environments and careers that are compatible with their interests. Holland identified six main personality types that can be associated with vocational interests: realistic, investigative, artistic, social, enterprising, and conventional (Holland, 1979). Similar to the Big Five, these personality types have unique traits associated with them. Given that the Big Five is one of the most recognized personality assessments, support for its use as a measure of job compatibility would result from a high correlation with scores on Holland's test.

In an attempt to uncover evidence for this prediction, De Fruyt and Mervielde (1997) conducted a study to assess the association between the Big Five and the RIASEC. They sampled a large number of individuals (934 participants) from a wide variety of college majors and professions. They found several moderately high correlations between Big Five personality domains and interests on Holland's test (see Table 2).

Table 2.

Correlations between Big Five and Big Six personality dimensions (De Fruyt & Mervielde, 1997)

NEO-FFI Big Five Personality Domains	Holland's RAISEC Big Six Interests	r
Openness	Artistic	0.56
Extraversion	Enterprising	0.48
Extraversion	Social	0.29
Emotional Stability	Enterprising	0.33
Agreeableness	Social	0.29
Conscientiousness	Conventional	0.42

Note: See De Fruyt and Mervielde (1997, p. 94) for a list of all correlations.

Additionally, Larson, Rottinghaus and Borgen (2002) conducted a meta-analysis of 12 studies that specifically recorded peoples' scores both on the Big Five and on Holland's vocational interest assessment. They found several moderate relationships between the dimensions on the two tests, such as a correlation of 0.48 between artistic and openness, 0.41 between enterprising and extraversion, and 0.31 between social and extraversion. Several other significant, albeit smaller, correlations were also revealed. (For a list of correlations, see Larson et al., 2002, p. 223). However, neither De Fruyt and Mervielde (1997) nor Larson et al. found evidence for any meaningful correlations between Holland's realistic and investigative traits and any Big Five domains. Thus, while the Big Five has been shown to be a good predictor of job performance, it does not seem to be useful for uncovering those individuals with either realistic or investigative interests.

Holland's (1985) vocational interest model was designed to match people's interests with appropriate careers, not to predict success or achievement in those vocations. Thus, although the finding of some strong correlations between the Big Five and Holland models is promising, one must be cautious in determining the relevance for the Big Five personality assessment and job performance. Nevertheless, the parallels do suggest that the Big Five assessment has some relevance for measuring vocational appropriateness. For instance, based on the moderate correlations between Big Five personality dimensions and some of Holland's six interests, it may be reasonable to use the NEO-FFI to assess the likelihood of someone enjoying a particular career, provided that some of the Big Five dimensions are related to the interests in question. However, Holland's RIASEC inventory seems to be distinct in its ability to match an individual's interests with a particular career.

The Big Five and Driving Performance

Several interesting studies have found evidence that driving performance has a strong relationship with several of the Big Five personality domains. For instance, Arthur and Graziano (1996) tested the relationship between accident involvement and individuals' Big Five personality scores. Because conscientiousness is associated with traits such as compliance, the authors hypothesized that conscientiousness should be highly correlated with driving performance. People who are high in this dimension should be more aware of traffic laws and other drivers and thus be involved in fewer collisions. There should be a negative correlation, then, between a person's conscientious score and the number of accidents that he or she precipitated. To test this prediction, Arthur and Graziano used two different measurements of the Big Five personality domains, including the NEO Five-Factor Inventory, with two different samples of participants. Participants completed both personality tests, as well as a driving behavior questionnaire (Arthur, 1991) where they reported the number of at-fault and not-at-fault accidents in which they had been involved.

The results showed that for both samples of participants, individuals who reported being involved in one or more at-fault accidents had lower conscientious scores than individuals who did not report being involved in an accident. In other words, the researchers' hypothesis was confirmed: there was a negative correlation between number of at-fault accidents and conscientious scores. Moreover, conscientiousness was also found to be inversely related to the number of moving violations cited. Thus, individuals who score higher in the conscientious domain appear to cause fewer accidents and receive fewer tickets. Note that for professions that require large amounts of driving, conscientiousness may therefore be an even greater predictor of job performance than suggested by previous studies.

Other Methods of Predicting Job Performance

Several measurements other than the Big Five are also used for assessing an individual's potential to excel at a given career. As discussed previously, Holland's RIASEC scale is used to determine which types of jobs are associated with a person's interests. However, congruence between one's interests and vocational choice does not necessarily guarantee success. Other research has investigated the validity of Criterion-focused Occupational Personality Scales (COPS), personality scales that were developed by industrial psychologists for the specific purpose of predicting various criteria related to job performance and aiding in personnel selection decisions (Ones & Viswesvaran, 2001a). Examples include integrity tests, stress tolerance, and violence scales. Thus, unlike traditional personality tests, COPS are explicitly designed to predict job performance measures and differences in employees' work behavior.

Ones and Viswesvaran (2001a) discussed the validity of sveral of these measures as they relate both to the criteria that they were designed to measure and to overall job performance. Integrity tests, for instance, were shown to be highly correlated with supervisors' ratings of employee job performance, and negatively correlated with counterproductive behavior at work, and on-the-job accidents (Ones, Viswesvaran, & Schmidt, 1993). Not surprising, integrity scores were also found to have the highest correspondence to the conscientiousness rating of the Big Five personality dimensions (Ones, 1993), which, as discussed above, is often cited as the best predictor of job performance out of the five dimensions. Other COPS, such as stress tolerance and violence scales, have also been demonstrated to be highly predictive of (1) the criteria that they were designed to measure, (2) counterproductive work behavior, and (3) overall job performance. (See Ones & Viswesvaran, 2001a, p. 35 for a list of correlations.) These scales have also been shown to moderately correlate with the conscientiousness domain (Ones & Viswesvaran, 2001b).

Evidence that many of these COPS are highly correlated with well-established personality dimensions lead Ones and Viswesvaran (2001a) to conclude that

> For organizations aiming to maximize worker performance, one suggestion is to use integrity tests and other COPS… However, if personality testing is being considered as part of a selection system for the purpose of reducing counterproductively only, there may not be a clear advantage to integrity tests and other COPS over conscientiousness measures. (p. 37)

In other words, their recommendation of job performance measurements depends largely on the rationale for assessing current or prospective employees' attitudes. If one desires an overall indication of an employee's personality, then a test such as the NEO-FFI should be preferred, as it can serve as a reliable measure of personality as well as a moderate indicator of job performance. However, if the only motivation for assessing personality is to predict other job-related tendencies, such as counterproductive behavior, then occupational scales may be favored, as they provide a strong indication of personnel tendencies as well as overall job performance. It is also worth noting that integrity, stress tolerance, and violence scales are moderately correlated with the agreeableness and emotional stability dimensions of the Big Five test. Based on the fact that COPS appear to be quite similar to these three personality domains, Ones and Viswesvaran (2001a) also suggested that COPS likely measure the characteristics that are required for functioning in accordance with social rules. Thus, the Big Five seems quite attractive as a catch-all measure of employees' personality and job performance ratings.

Robertson and Smith (2001) reviewed several different types of methods used for personnel selection. They began by reasserting the gains that have been made during the past two decades in developing reliable and valid selection methods, as well as advances in conducting meta-analyses to explore the ability to generalize and apply various measurements to job performance. Based on their review, they extoled the use of other measurements in addition to personality inventories for personnel selection. For instance, structured interviews tend to be highly correlated with job performance, and are generally superior to non-structured interviews. (See Salgado, 1999, for a review of interview types and job performance correlations.) Interviews are thought to measure characteristics such as social skills, experience, and job knowledge that are not necessarily captured on personality questionnaires.

In regards to traditional personality assessments, Robertson and Smith (2001) agreed with the results of many other studies that conscientiousness, or integrity, appears to be the strongest predictor of job performance. This is especially true across a large sample of job types. However, when it comes to predicting job performance for a specific occupation, broad personality measures may not be as valid (Robertson, Baron, Gibbons, MacIver, & Nyfield, 2000). The most effective method of predicting employee job performance and compatibility, therefore, may be to use a variety of assessment techniques, including personality and job interest questionnaires, as well as vocation-specific occupational tests (i.e., COPS) and structured interviewing techniques.

Other Measures Relevant to Simulator Training

Sensation Seeking Tendencies

Sensation seeking is a concept championed by Zuckerman during the 1960s and 1970s. It is commonly defined as the pursuit of novel and intense experiences or the willingness to take risks for the sake of such experiences (Zuckerman, 1994). Sensation seeking measurements of personality are based on the notion that individuals differ in their optimal levels of arousal and stimulation, which influence their choices in activities. For instance, individuals who have a high level of arousal may need to seek out dangerous or risky activities to fulfill their pleasure-seeking requirements. People who score low in sensation seeking, on the other hand, are generally appeased by more mundane activities.

Form V of Zuckerman’s Sensation Seeking Scale is comprised of 40 total questions from four subscales: thrill and adventure seeking (TAS), experience seeking (ES), disinhibition (Dis), and boredom susceptibility (BS). Some common findings include an overall negative correlation with age as well as some minor cultural differences (Zuckerman, Eysenck, & Eysenck, 1978). Positive correlations were discovered with drug use (Forsyth & Hundleby, 1987) and sex differences (Zuckerman, 1979), with males tending to score higher than females across the culture. As reported in Zuckerman, Eysenck, and Eysenck (1978), males’ total sensation seeking scores tend to be around 20 during their 20s, and scores decline linearly to approximately 12 during their 60s. Females’ score were consistently around four points lower across this range.

Immersion and Presence

Witmer and Singer (1998) authored the prominent paper on immersion and presence tendencies for individuals in virtual reality environments. They defined immersion as “A psychological state characterized by perceiving oneself to be enveloped by, included in, and interacting with an environment that provides a continuous stream of stimuli and experiences” (p. 227). Presence was defined as “The subjective experience of being in one place or environment, even when one is physically situated in another” (p. 225). These concepts are critical to the effectiveness of virtual reality environments. Specifically, presence is important for involving oneself in the virtual environment while becoming detached from the physical location of the apparatus, while immersion is a more general measure of involvement and inhibition of distractions in everyday life. With regards to driving simulators, individuals high in presence would perceive themselves as actually driving a vehicle through a roadway surround by other drivers, rather than inside a building, for instance. Similarly, virtual environments that differentiate themselves from the physical environment, for instance by providing a wide field of view or requiring the individual to perform a difficult task, will lead to increased feelings of immersion. Immersive tendencies outside of virtual reality environments include strong feelings of connection with prominent characters in books or movies, becoming highly involved in activities, and becoming mentally detached from one’s current environment.

Witmer and Singer (1998) developed two scales to measure immersion and presence, with 29 and 32 questions, respectively. The immersion questionnaire consists of three subscales—involved, focus, and games—designed to measure how well people can concentrate and apply themselves to general tasks. Other questions deal with one's ability to block out or ignore distractions. The presence questionnaire is comprised of four subscales—control, sensory, distraction, and realism factors—that measure discrete qualities of participants' virtual reality experience. Thus, virtual reality environments that are intuitive to navigate, provide familiar sensory feedback, and are highly realistic provide a high degree of presence. Just as levels of cybersickness can differ based on the device being used, average presence scores are also unique for each virtual environment. Simulators that can isolate the individual from the physical world better, or offer feedback from all sides, may offer a higher degree of presence (Ooms, 2004). Several studies have identified mean total presence scores of around 95 out of a maximum of 224. (Usoh, Catena, Arman, & Slater, 2000; Ooms, 2004). Both of these studies used head-mounted virtual reality displays, which Ooms (2004, p. 1) stated, "are considered the ultimate tool to get taken away into a virtual world." Thus, presence scores of around 100 may be taken to reflect a moderate degree of immersion.

Conclusion

Driving simulator research and training with modern equipment is quite common. Equipment modifications and technological advances in visual displays allow trainers to tailor the virtual environment for the requirements of the trainee or participant. Realism, visual quality, and control responsiveness are not only important for ensuring skill transfer to real-life applications; they may increase trainees' presence within the virtual reality environment and reduce feelings of cybersickness. In addition to performance measurements recorded by most driving simulators, examining people's eye movements is an effective method of determining the impact that distractions have on normal vehicle navigation. Eye movement data can also be used to judge whether or not individuals adopt similar scan patterns in the virtual environment as they do in real life.

Applying the results of previous studies, the literature on driving simulators suggests that simulator training should be an effective method for training snowplow operators. Several authors using different driving simulators have found convergent relative behavioral validity for driving simulators. In other words, these studies have shown that participants tend to behave similarly in driving simulators as they do in real vehicles on several performance measurements such as speed and lane position. Thus, there is no reason to believe that skills or strategies snowplow operators learn or refine in a snowplow driving simulator will not transfer to real world snowplowing. The present study examines how current Iowa DOT snowplow operators respond to simulator training and whether trained operators show any improvement in driving performance over non-trained operators in the driving simulator.

With regards to personality measurements, recent research has cautiously advocated the use of personality questionnaires and other measurements for predicting job performance. Job performance predictions can be measured using broad personality tests, such as the NEO-FFI, or criterion-specific tests, known as COPS. Generally, while criterion-specific tests are more accurate at measuring the intended variable, tests like the NEO-FFI also provide personality information. Overall, simulator training and the use of personality and criterion-specific questionnaires to predict job performance are well-regarded by applied researchers.

CHAPTER 2. PARTICIPANTS, MATERIALS AND METHODS

Participants

Two-hundred full time IDOT snowplow operators were randomly selected from a list of 1098 current maintenance workers, with a representative number chosen from the six districts in Iowa. Half of the operators were randomly chosen to be in the control group, and the other half comprised the experimental group. These groups were matched in terms of age, district, number of accidents in the past five years and approximate amount of snowplowing and truck driving experience. Operators were informed that they would be receiving mandatory training, but that they had the option of participating in the study to assess training effectiveness. Overall, 174 operators, 84 in the experimental and 90 in the control group, elected to participate in the study. Scheduling conflicts on the part of the researchers or IDOT trainers prevented 24 individuals from participating. Only two operators elected not to begin or complete the experiment once they arrived at the facilities. There were two females in the sample. The estimated mean age of the sample was 47.6 years (actual age was not provided, rather the age was reported in 10-year increments) and the average years of experience as a snowplow operator was 12.6 (SD = 7.41).

The Simulator

Participants received training in the TranSim VS III truck and snowplow simulator. The VS III has a 180 degree horizontal and 37 degree vertical viewing area at 34 inches from the center screen, and is comprised of three 1024 x 768 monitors with a refresh rate of 70 Hz. The driving apparatus is similar to what would be found in a typical truck: functional brake, clutch and accelerator pedals, radio, transmission, and digit gauges, all of which can be tailored to the requirements of the trainer. Although a clutch pedal was present, all simulations used an automatic transmission. The simulator also outputs digital sound designed to mimic normal operating sounds, such as engine and exterior noise. Tactile transducers under the drivers' seat provided simulated road vibration. Finally, two graphical rear view mirrors were displayed on the bottom corners of the central screen (spot mirrors) along with two graphical adjustable side mirrors on the outer half of the left and right screens (wing mirrors). A picture of the interior of the simulator is shown in Figure 1.

Figure 1. Photograph of the simulator.

Three weeks into the study, certain parts of the driving simulator were upgraded: more snow was added to the edges of the windshield, and a few additional cars were added to the test scenario. Also as part of the upgrade, some of the tests for the performance measures were changed or excluded. Unless otherwise stated, the performance statistics reported were only included from the 136 participants who completed the training *after* the upgrade.

In order to track the participants' eyes and head movements, cameras were mounted on the very top of the simulator, approximately six inches above the central monitor (see Figure 1), so as not to interfere with participants’ performance in the simulator. The tracking was accomplished by running FaceLab (version 3.2) software. Only head and eye movements for the experimental scenarios were recorded.

Procedure

Training was conducted primarily at the main District One facility in Ames, Iowa. Typically, participants were trained in groups of two in a session that lasted for approximately four hours (8 a.m.-noon or 12:30 p.m.- 4:30 p.m.). Upon arriving at the facility, the participants were introduced to the trainer, who was an experienced IDOT snowplow operator, and shown the simulator. The trainer then drove a 3-minute scenario down a sparsely populated rural road to acquaint the participants with the visuals and with the auditory feedback. Next, participants were asked to sign an informed consent document if they wished to participate in the experiment. As mentioned above, only one individual declined to participate, and one operator later elected to withdraw from the study. Operators who gave their consent to participate then drove an introductory 3-minute scenario through a snowy highway to become acclimated with the controls and handling of the simulator. No data were collected during this orientation drive.

Immediately after the orientation drive, participants completed their first set of questionnaires. This set included: the NEO-Five Factor Inventory (NEO-FFI), which is based on the Five Factor Model developed by Costa and McCrae (1985); a modified version of

Zuckerman's (1979) Sensation Seeking scale (form V); and Witmer and Singer's (1998) immersive tendencies questionnaire. The questionnaires are described in more detail in a subsequent section of the manuscript, as well as in Chapter 1.

While participants were completing these questionnaires, they took turns driving the *first* experimental scenario. The scenario involved merging onto an interstate in snowy conditions and plowing snow for approximately 10 minutes. The simulated vehicle had a right plow and right wing, although participants did not control this equipment. At various times during the scenario participants had to pass slow moving vehicles in the right lane, while avoiding striking them with the right wing, and then merge back into the right lane to avoid faster moving vehicles approaching from the rear. Participants were instructed to operate the simulator in the same manner that they would operate a snowplow while removing snow on a real highway.

After completing the first set of questionnaires and the first experimental scenario, operators assigned to the control group immediately completed the *second* experimental scenario. The second scenario was identical to the first. Following completion of the second drive, the control group began their training. Participants in the experimental group received training before completing their second experimental scenario. This design, in which the only difference between the control and experimental groups was whether their second experimental drive occurred before or after training, allowed us to determine whether the training had any immediate impact on performance, fixation behaviors, or both. Additionally, this methodology allowed all operators to receive simulator training rather than just those in the experimental condition.

The training consisted of three parts: a lecture that included PowerPoint slides, a computer exercise, and a simulator exercise. The trainer began by giving a 20 to 25 minute lecture, which included a PowerPoint presentation. The lecture focused on the importance of being aware of the space around the vehicle and included a discussion of the S.I.P.D.E. (Scan-Identify-Predict-Decide-Execute) method (see Appendix A). After the PowerPoint lecture, each participant completed a 5 to 10 minute driving scenario in the simulator in which he or she was encouraged to employ the techniques learned during the lecture. The trainer would sporadically ask the participant questions to ascertain whether or not he or she was using the information presented during the lecture. Next, participants completed a self-paced computer exercise in which they watched short video clips that contained information about passing vehicles, speed management, and space management. After each series of clips, participants answered multiple choice questions and received feedback regarding their answers. Finally, participants drove an additional 5 to 10 minute scenario in which they were again instructed to employ the techniques that they learned during the PowerPoint lecture and the computer exercise. Training concluded with the trainer giving a 5 to 10 minute summary of the information covered during training (see Appendix A).

Once training ended, participants in the control group completed their second set of questionnaires. Participants in the experimental group drove the second experimental scenario before completing the questionnaires. The questionnaires included a modified version of Witmer and Singer’s (1998) presence questionnaire, Kennedy, et al. (1993) Simulator Sickness Questionnaire (SSQ) and a modified version of the questionnaire used by Strayer, et al. (2004) for the Utah DOT snowplow study. The questionnaires and modifications are described when the relevant results are presented.

CHAPTER 3. RESULTS AND DISCUSSION

Characteristics of IDOT Snowplow Operators

One of the purposes of the study was to determine the characteristics of IDOT snowplow operators. Because personality variables are sometimes good predictors of success in or satisfaction with one's occupation (Barrick & Mount, 1991; Tett et al., 1991), three separate personality inventories were administered prior to training (the NEO-FFI, the sensation seeking scale, and the immersion scale). One hundred seventy participants completed all three of these questionnaires. In the case of missing data (less than 1% of the data), a participant's mean score for the scale or subscale was substituted for the missing value. For a more in depth discussion on the constructs measured in each questionnaire, see Chapter 1.

Demographics

Just over 1% (2 out of 174) of the participants were women. Figure 2 shows the number of operators from each age category who participated in the study compared to the population of IDOT snowplow operators. As would be expected given the selection procedure, the selected operators were representative of the population of snowplow operators with respect to age.

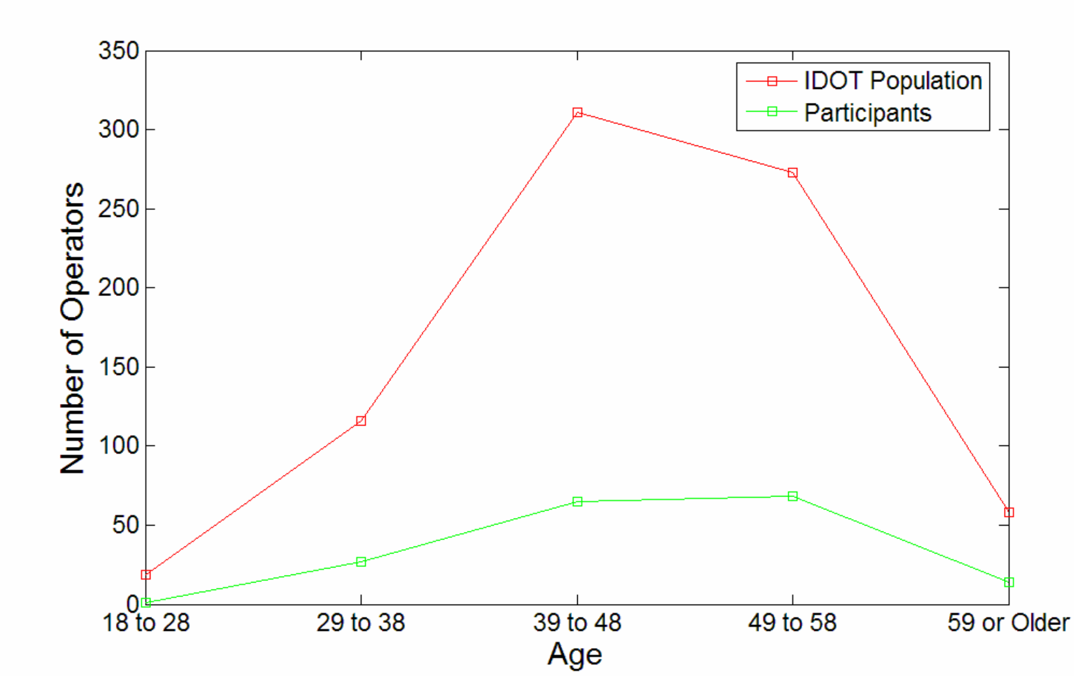


Figure 2. Age of IDOT operators and selected participants.

Years of snowplow experience ranged from 0 to 30 years. For purposes of analysis, participants were divided into three categories based on their years of snowplow experience: 'Low experience' operators were defined as those with between 0 and 5 years of snowplowing experience (43 operators), 'Medium experience' operators were defined as those with between 6 and 15 years of experience (62 operators), and 'High experience' operators were defined as those with 16 or more years of experience (69 operators). Figure 3 shows the number of operators in each of the four age groups for the three levels of snowplowing experience. Experience categorization had no affect on the training that the operator received. Of interest was whether various dependent measures would vary as a function of experience.

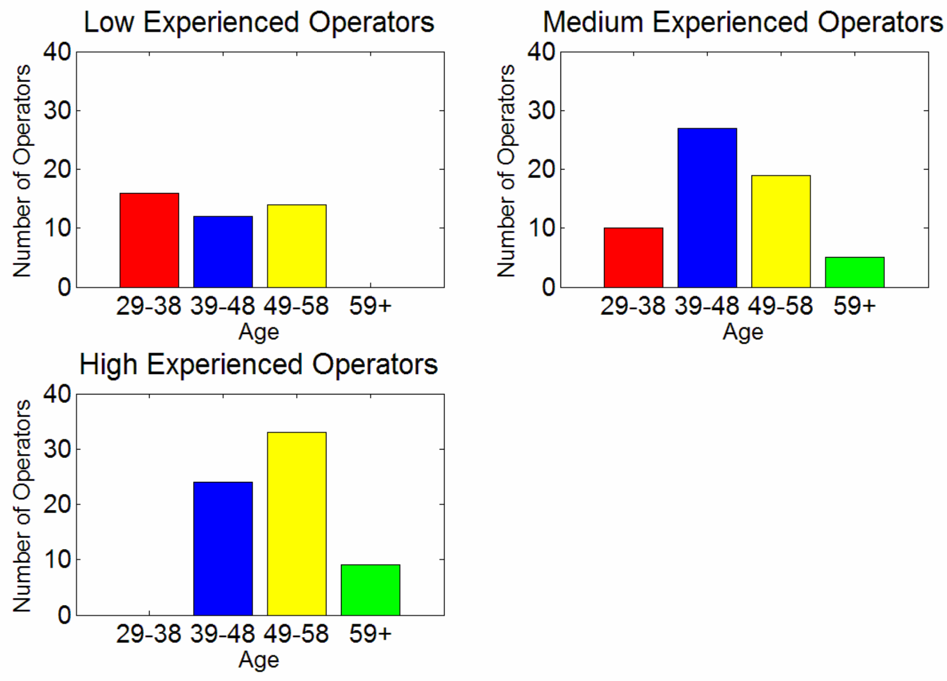


Figure 3. Age breakdown by the three levels of experience.

As would be expected, experience was correlated with age category (r = .39, $p<.001$). Because personality measures often differ as a function of age, whenever a reliable effect of experience was found in the Analysis of Variance (ANOVA), an Analysis of Covariance (ANCOVA) was run with age category included as a covariate in order to determine whether the effect of experience remained when age was statistically controlled in this way.

NEO-Five Factor Inventory

The first personality questionnaire was the NEO-Five Factor Inventory (NEO-FFI), which is based on the Five Factor Model developed by Costa and McCrae (1985). It is designed to measure personality dispositions in five sub scales: neuroticism, extraversion, openness, agreeableness and conscientiousness. The NEO-FFI contains 60 statements (12 for each sub scale). Participants used a five-point scale, ranging from strongly disagree to strongly agree, to indicate the extent to which each statement is a good characterization of their tendencies. The actual questionnaire is in Appendix B. Sample questions include:

- I like to have a lot of people around me (extraversion)
- I often try new and foreign foods (openness)
- I work hard to accomplish my goals (conscientiousness)
- I try to be courteous to everyone I meet (agreeableness)
- I often feel inferior to others (neuroticism)

The first analysis examined whether personality profiles differed as a result of experience level. Of particular interest was whether high experience participants might be different from low experience participants. For instance, any such difference might be useful in determining the type of person likely to remain a snowplow operator for a longer period of time. The second analysis examined whether any particular personality scale or set of scales characterized snowplow operators in comparison to the general population. Again, any such difference might be useful in selecting persons to be hired as snowplow operators.

Figure 4 shows the mean scores on each of the five sub scales for low experience, medium experience, and high experience participants. Separate one-way ANOVAs were run on scores in each of the five domains in order to determine whether there were differences as a function of experience. Only the openness domain showed a reliable difference, ($F(2,167) = 3.43$, $MSE = 29.3$, $p<.05$). Least Significant Differences (LSD) comparisons among the experience groups showed that the low experience participants were more open to new experiences than the medium and high experience participants, who did not differ. However, when age category was included as a covariate in an ANCOVA, it was a reliable covariate and the effect of experience was no longer apparent. This suggests that the observed difference was reflecting age more than experience level. Thus, the personality sub scales measured by the NEO-FFI did not differentiate among individuals in the sample in terms of experience.

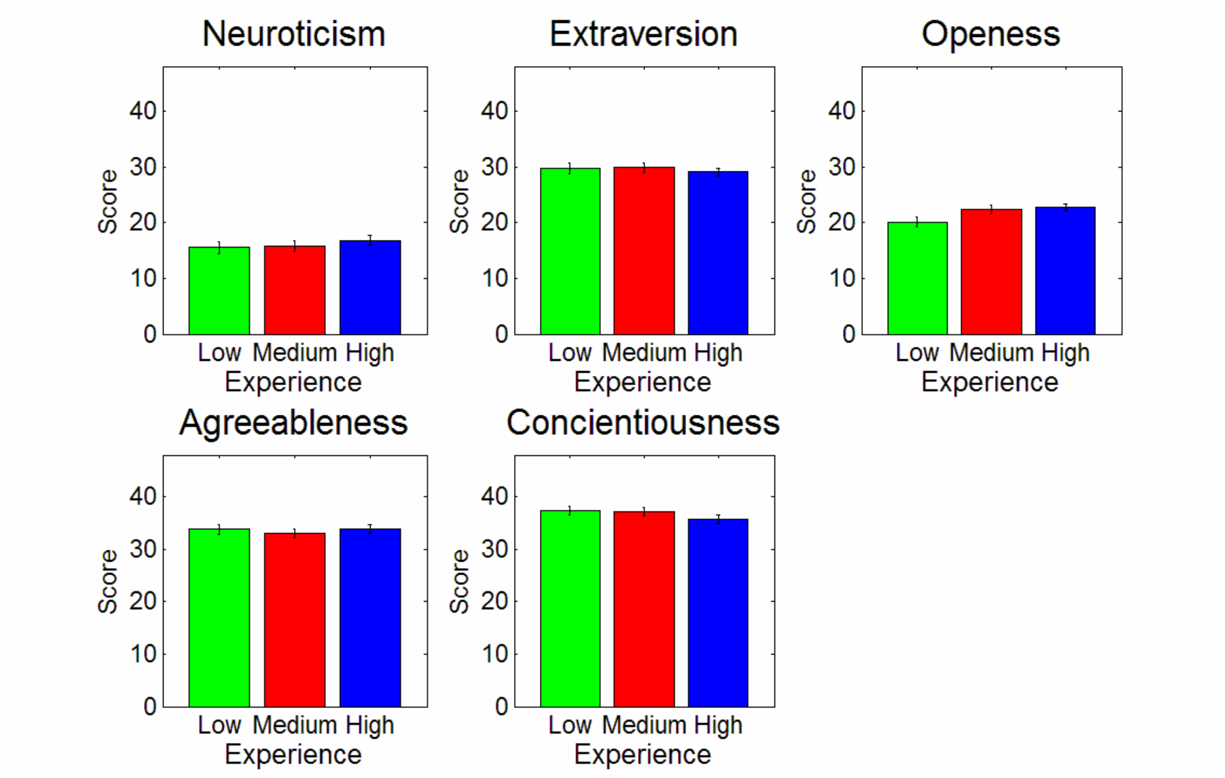


Figure 4. Mean NEO-FFI score on each sub scale for the three experience groups.

The NEO-FFI scores of the snowplow operators measured in our sample were then compared to those of a normative sample of 500 American males as reported in Rolland et al. (1998). Table 3 shows the normative scores and standard deviations reported on the five NEO-FFI domains along with the scores of the IDOT participants collapsed over experience category. Except for the openness trait, the observed differences generally were rather small, but they were reliable. This comparison treated the normative value as a population mean against which the participants' mean was compared. Snowplow operators were less neurotic, more extraverted, less open, more agreeable, and more conscientious (all *t*s > 2.73, all *p*s < .05). Although the data suggest that some NEO-FFI traits might provide information that could be used in the selection of snowplow operators, it is highly likely that the snowplow operators differed from the normative sample in many ways, particularly in terms of demographics. A more appropriate control group would be a sample of IDOT employees who are not snowplow operators. These individuals would be more similar to the sample of snowplow operators on several important factors, particularly age, race and income. This more appropriate matched sample would be needed before the usefulness of NEO-FFI as a selection criterion could be determined.

Table 3.

Normative means and standard deviations (Rolland et al., 1998) for the NEO-FFI and scores on the present study

	Mean and Standard Deviations by Domain				
Authors	Neuroticism	Extraversion	Openness	Agreeableness	Conscientiousness
Rolland, Parker & Stumpf (1998)	17.6 (7.46)	27.2 (5.85)	27.1 (5.82)	31.9 (5.03)	34.1 (5.95)
Present Study	16.1 (7.16)	29.5 (6.06)	23.8 (5.49)	33.8 (6.15)	36.7 (6.08)

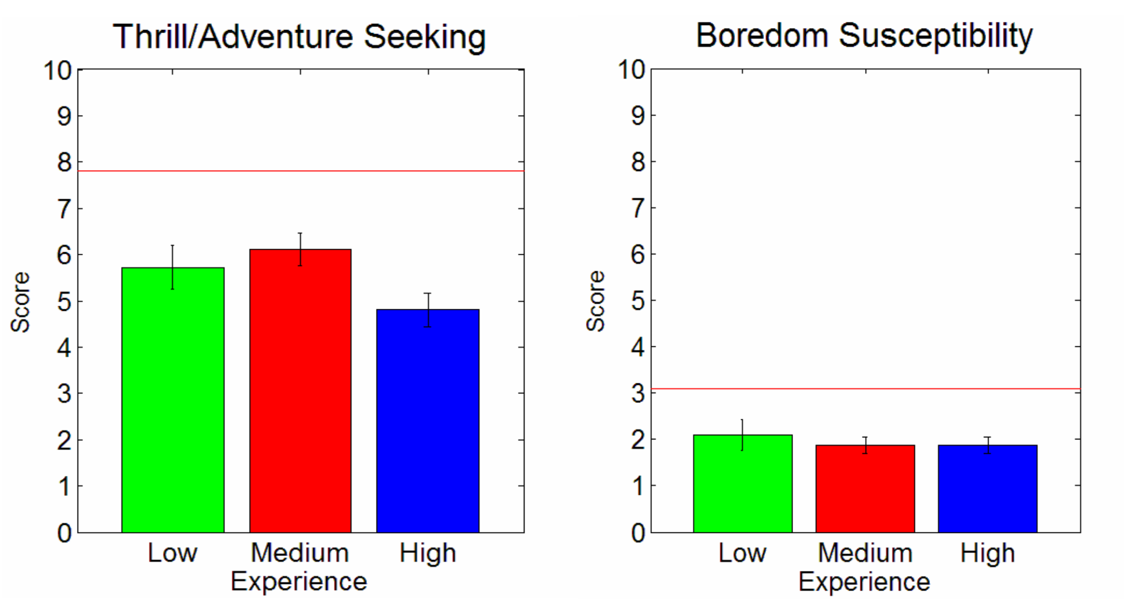
Sensation Seeking

Sensation seeking tendencies were explored by Zuckerman (1979), who described the personality construct of sensation seeking as a tendency to seek out novel or exciting experiences for the sake of enjoying the activity. The sensation-seeking questionnaire has four subscales: thrill and adventure seeking, experience seeking, disinhibition and boredom susceptibility. A modified version of Form V of the questionnaire was used. Form V contains 40 questions, 10 questions for each subscale. However, seven questions were removed from the disinhibition subscale and three were removed from the experience seeking subscale because they asked about drug or alcohol use, or other controversial behaviors. The modified questionnaire is shown in Appendix C. For each question, the participant chose which of two statements is a better description of their behaviors or preferences. Example questions include:

- Thrill Seeking: A) A sensible person avoids activities that are dangerous
  B) I sometimes like to do things that are a little frightening
- Boredom Susceptibility: A) I get bored seeing the same old faces
  B) I like the comfortable familiarity of everyday faces
- Disinhibition: A) I like uninhibited parties
  B) I prefer quiet parties with good conversation
- Experience Seeking: A) I prefer "down to earth" kinds of people as friends
  B) I would like to make unusual friends

The first analysis examined whether more experienced participants have different sensation seeking profiles than less experienced operators. As with the NEO-FFI, any such difference might be useful in determining the type of person likely to remain a snowplow operator for a longer period of time. The second analysis compared snowplow operators to the general population.

The mean scores for the two complete sub scales, thrill and adventure seeking and boredom susceptibility, are show in Figures 5 and 6. A one-way ANOVA on each of the subscale scores with experience as a between-subjects variable showed a difference on the thrill and adventure seeking subscale, $F(2, 167) = 3.33$, $MSE = 8.60$, $p<.05$, in which high experience participants reported less thrill and adventure seeking tendencies than the middle experience participants (LSD Comparisons). Zuckerman (1979) reported that older individuals in general show lower sensation seeking scores than younger individuals. When age category was included as a covariate, it did not produce a reliable difference, but experience effect was no longer reliable (p > .11).



Figures 5 and 6. Mean Thrill and Adventure Seeking and Boredom Susceptibility scores for the three experience groups. Red bars are population means from Zuckerman (1979).

There was also some evidence that IDOT snowplow operators show less sensation seeking tendencies overall than the general population. The red lines in Figure 4 and Figure 5 depict the mean scores for a group of 377 American males on the thrill and adventure seeking and boredom susceptibility subscales, respectively (Zuckerman, 1979); these were the two subscales that were presented in their entirety in the current study. For both traits, all groups of participants, regardless of experience level, reported less sensation seeking than the normative average, all ts > 4.70, all ps < .01 for thrill and adventure seeking and all ts > 3.00, all ps < .05 for boredom susceptibility. These comparisons treated the normative value as a population mean against which the participants' mean was compared. As was indicated in the discussion of the NEO-FFI traits, the current data are only suggestive. An appropriate control group is needed before the usefulness of Sensation Seeking measures as a selection tool can be determined.

Given the correlation between age and years of experience and the fact that the IDOT operators were older in general than the normative sample, the interpretation of this difference must be made with caution. IDOT operators may have lower sensation tendencies than the American males in other research samples, but this difference may also be a function of age. It is likely that the IDOT sample was older than the individuals examined by Zuckerman (1979).

Immersion

Immersion is commonly defined as the ability to become enveloped or involved in an environment or task (Witmer & Singer, 1998). It is a general characteristic not directly tied to virtual reality environments. High immersion scores also imply the ability to block out or ignore task irrelevant distractions. Of particular interest was whether immersion would vary as a function of experience, since individuals with lower immersive tendencies may not become as involved or focused in training, and therefore might not experience as much of a benefit from it.

The Immersion questionnaire (Witmer & Singer, 1998) contained 29 questions, a majority of which are related to one of three domains: focus, involvement and games. For the majority of questions, participants respond on a scale from 1 (not often/much) to 7 (very often/much) on the extent that the statement typifies their behavior. The questionnaire is shown in Appendix D. Sample questions include:

- Do you ever become deeply involved in movies or TV dramas? (involvement)
- How good are you at blocking out external distractions when you are involved in something? (focus)
- How often do you play arcade or video games? (games)

The mean immersion scores for each experience group are shown in Figure 7. One-way ANOVAs with experience as a between-subjects variable showed no differences, all *Fs* $(2, 167) < 1$. Although one must be cautious when interpreting the lack of a difference, the means suggest that older, high experienced operators will be as likely to become immersed in simulator training as low experience operators. To the extent that immersion is a necessary prerequisite to successful simulator training, then, experienced operators should have the potential to benefit as much from training as less experienced operators.

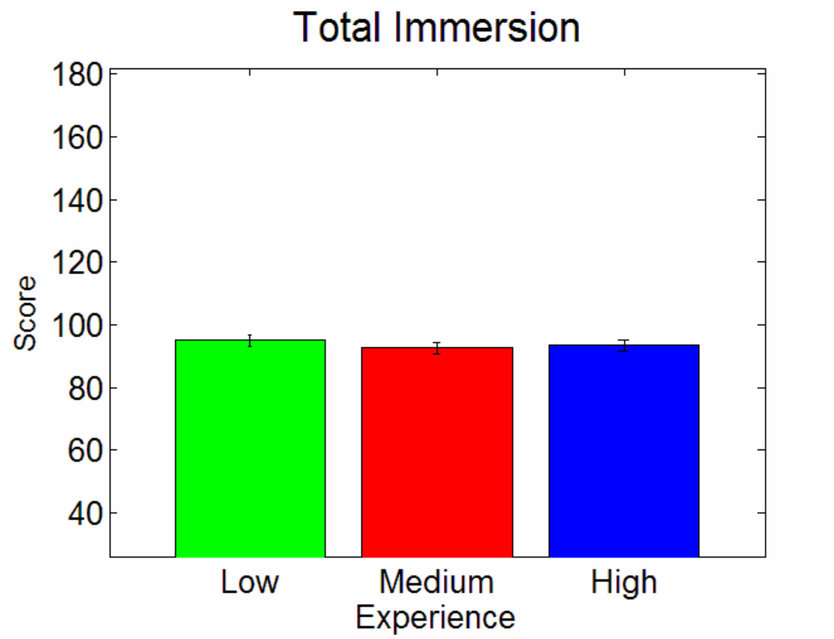


Figure 7. Mean total Immersion scores for the three experience groups.

Summary of Demographic and Personality Variables

Although they varied in terms of years of experience as a snowplow operator, the participants of this study were comparable on three well-researched personality measures when age differences were taken into account: the NEO-FFI, the sensation seeking scale, and a measure of immersion. While none of the measures would currently provide a useful selection tool to determine who might make a good, or at least a long-term, snowplow operator, the immersion measure does suggest that snowplow operators at all experience levels should show similar levels of involvement for simulator training.

Responses to Simulator Training

A major purpose of the study was to determine the effectiveness of simulator training for snowplow operators. In order for the simulator to be a useful tool in training novices and in providing experienced operators with a way to practice their skills throughout the year, it must provide a realistic driving environment. The judged realism of the simulator could vary as a function of experience (e.g., high experience operators may find the simulator to be more or less realistic than low experience operators) making the simulation more or less useful for operators at a particular experience level. Judgments were made about simulator realism quality through participants' responses to the presence questionnaire, the simulator sickness questionnaire, and direct questions about simulator realism. Participants' satisfaction with the entire training experience was also examined.

Presence

Presence is directly related to immersion in a virtual reality environment. Witmer and Singer (1998) defined presence as "the subjective experience of being in one place or environment, even when one is physically situated in another (p. 225)." For example, although operators were physically located in the interior of a trailer during training, those with high levels of presence may have had the experience of actually driving on a real highway. The effect of snowplow experience on amount of presence was examined, with the implication that low levels of presence may result in less satisfaction with the driving simulator or simulator training. Possibly, more experienced operators could be more cognizant of inconsistencies between real and simulated snowplow driving, and therefore might experience less presence.

The presence questionnaire (Witmer & Singer, 1998) contains 32 questions from 4 domains: control, sensory, distraction and realism. Participants respond on a scale from 1 (not often/much) to 7 (very often/much) on the extent that each statement typifies their experience in the simulator. The wording of some items was modified slightly to make them more applicable to the current study involving virtual reality snowplow simulation. The modified presence questionnaire is shown in Appendix E. Sample questions include:

- How natural did the driving simulator seem (control)?
- How much did the scenery of the simulator make you feel like you were really driving a snowplow (sensory)?
- How natural were the driving controls (distraction)?
- How similar was driving the simulator to driving a real snowplow (realism/control)?

The mean presence scores for each experience group are show in Figure 8. A one-way ANOVA with experience as a between-subjects variable revealed no differences between levels of experience. In other words, operators of all experience levels reported experiencing a similar amount of presence. Moreover, a one-sample t-test comparing participants' average response to a score of neutral (i.e., a score of 4 for each question) revealed that scores were significantly greater than this value ($t(169) = 8.14$, $SE = 0.054$ $p<.001$). Thus, participants reported a positive sense of presence in the simulator.

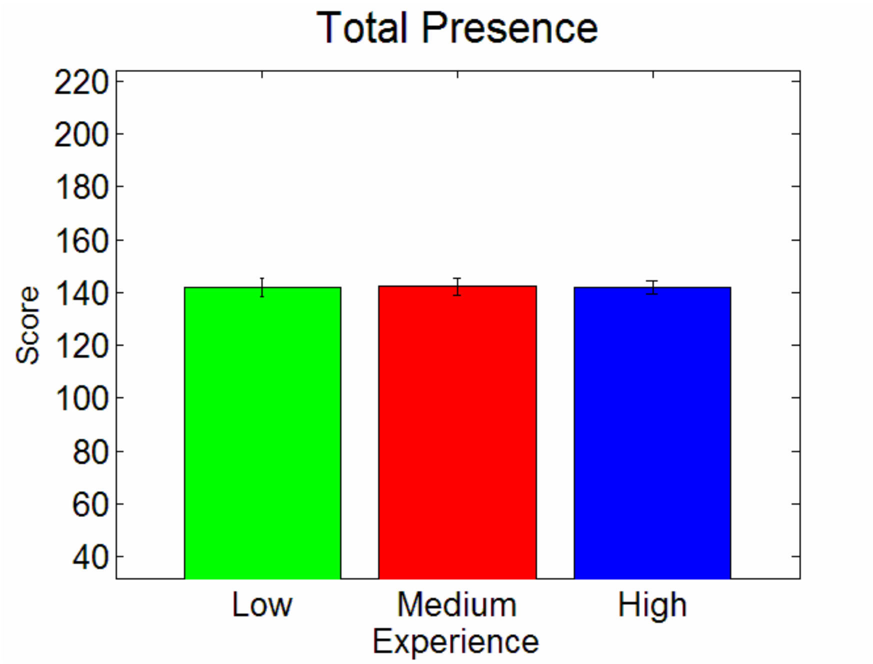


Figure 8. Mean total Presence scores for the three experience groups.

Simulator Sickness

Simulator sickness is always a concern when virtual reality devices are used for training because people who experience simulator sickness may not be able to concentrate as fully on training or they may be unable to finish training. Of particular concern is that older individuals may be at increased risk for simulator sickness (see Arms & Cerney, 2005). The SSQ, which was developed by Kennedy et al. (1993), was used to measure participants' reported level of simulator sickness after completing training. The SSQ contains 16 questions regarding potential symptoms of simulator sickness. Participants were instructed to report, on a 0 (none) to 3 (severe) scale, the extent to which they experienced each of an assortment of symptoms. The reported values are used to compute a total SSQ severity score that ranges from 0 to 180. The full questionnaire and scoring procedures for calculating the total weighted simulator sickness score are shown in Appendix F. Sample symptoms include:

- Headache
- Nausea
- Blurred Vision

The mean SSQ scores for participants from each experience group are shown in Figure 9. A one-way ANOVA with level of experience as a between-subjects variable showed no effect for level of experience, $F(2, 167) = 1.62$, $MSE = 1099.3$, $p > .20$. However, an independent samples t-test between operators with high versus low levels of experience was marginally significant ($t(107) = 1.80$, $p = .075$), suggesting that more experienced drivers did show a tendency to experience more simulator sickness. The most reasonable explanation for this finding is that older drivers experienced more simulator sickness than younger drivers. Thus, there is some indication that simulator sickness was related to age, as has been shown by others (e.g., Arms & Cerney, 2005).

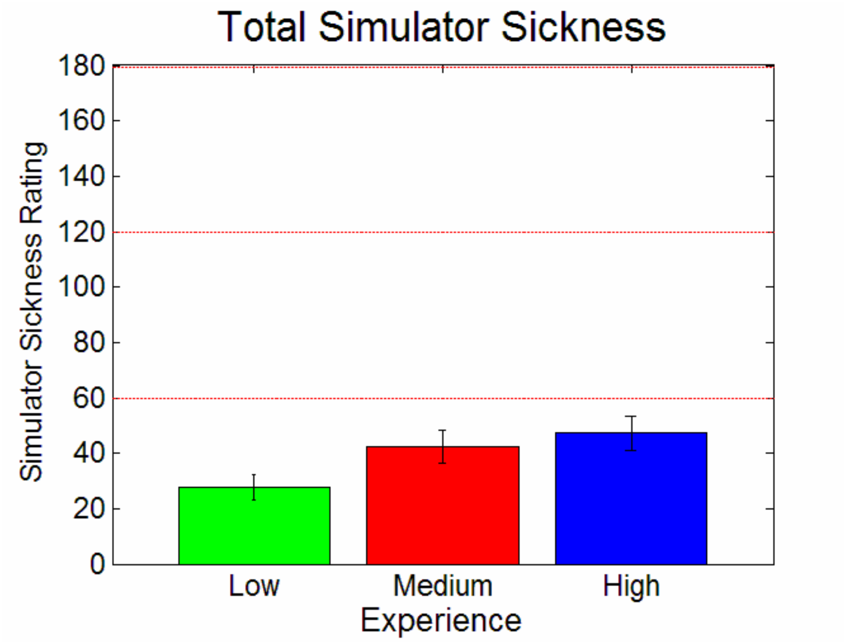


Figure 9. Mean weighted total Simulator Sickness scores for the three experience groups. Red bars refer to ratings of 'slight,' 'moderate,' and 'severe' discomfort for all symptoms.

The most important finding, however, is that participants' simulator sickness ratings were relatively low. A weighted score of approximately 60 corresponds to participants reporting 'slight' discomfort for all questions–a weighted score of approximately 120 corresponds to 'moderate' discomfort for all symptoms, and a weighted score of approximately 180 corresponds to 'severe' discomfort for all symptoms. The means of all experience groups were well below this rank (all ts > 5.86, all ps < .001). The comparisons treated a score of 60 as a population mean against which the participants' mean was compared. Additionally, only 5 out of a total of 174 participants were unable to complete the simulator training due to excessive simulator sickness. Thus, it does not appear that simulator sickness is an obstacle to the use of this simulator in training IDOT operators.

Simulator Realism

Questions regarding the realism of the TransSim VS III were added to the end of the training questionnaire used by Strayer et al. (2004) along with free response questions regarding training satisfaction. The questions addressed the realism of certain aspects of the driving simulator: the scenery (Question 1), the visuals of other vehicles (Question 2), the location (Question 3) and functionality (Question 4) of the mirrors, the behavior of other vehicles (Question 5), the location (Question 6) and responsiveness (Question 7) of the simulators' controls, and whether the operator thought driving the simulator approximated the experience of driving a real snowplow (Question 8). The questions can be found in Appendix G. Realism was rated on a scale of 1 (strongly disagree) to 7 (strongly agree).

The mean responses to each of the eight aspects of the simulator for all operators are show in Figure 10. One-way ANOVAs on the responses with experience as a between-subjects variable showed no differences as a function of experience. Next, the mean response to each question was statistically compared to 4, the neutral point, to determine whether the feature was considered to be realistically or unrealistically captured by the simulator. All aspects of the simulator, except for the overall realism rating were judged to be moderately realistic (all ts >3.80, all ps < .01). Based on the statistical analyses the real snowplow approximation rating was neither above nor below the neutral point, although Figure 10 seems to imply that, based upon one standard error, that the value was significantly less than 4. Thus, the responses to the direct question regarding realism show that participants found the visual qualities, location and functionality of the equipment to be fairly similar to those in a real driving situation.

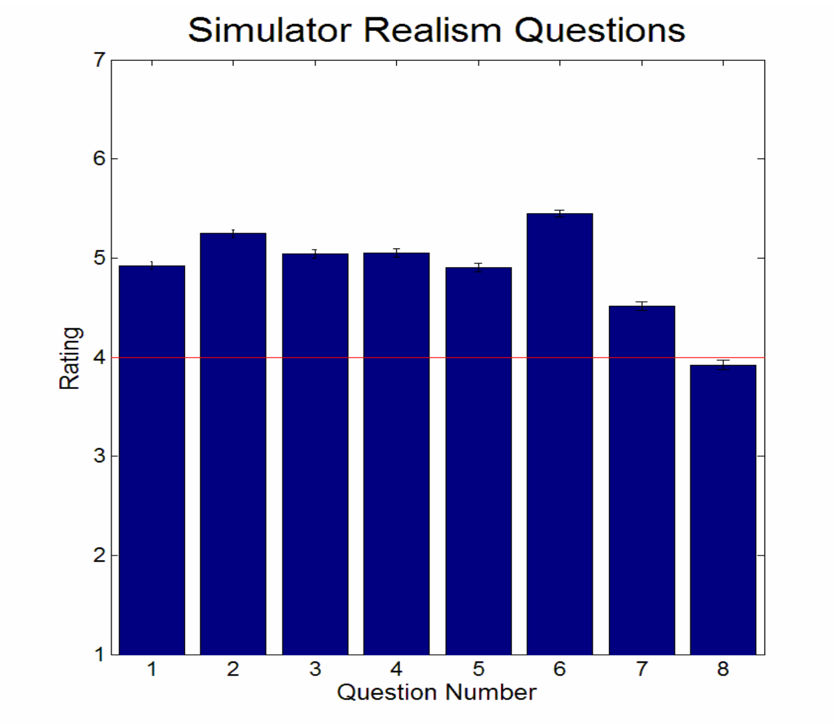


Figure 10. Mean Simulator Realism ratings for the eight realism questions.

Training

The major part of the final questionnaire was based on the questionnaire used by Strayers et al. (2004) in their simulator fidelity study of Utah DOT operators. The complete questionnaire is shown in Appendix G. Participants responded on a scale of 1 (Strongly Disagree) to 7 (Strongly Agree) to statements about five aspects of training: simulator training (5 statements), computer training (5 statements), lectures (4 statements), the trainer (4 statements), and the value of training (4 statements). The mean response to each of the five categories for each experience group are show in Figure 11. A one-way ANOVA on each category with experience as a between-subjects variable showed no differences due to experience. Thus, more experienced operators tended to rate training as highly as less experienced operators, suggesting operators from all levels of experience felt similarly about training.

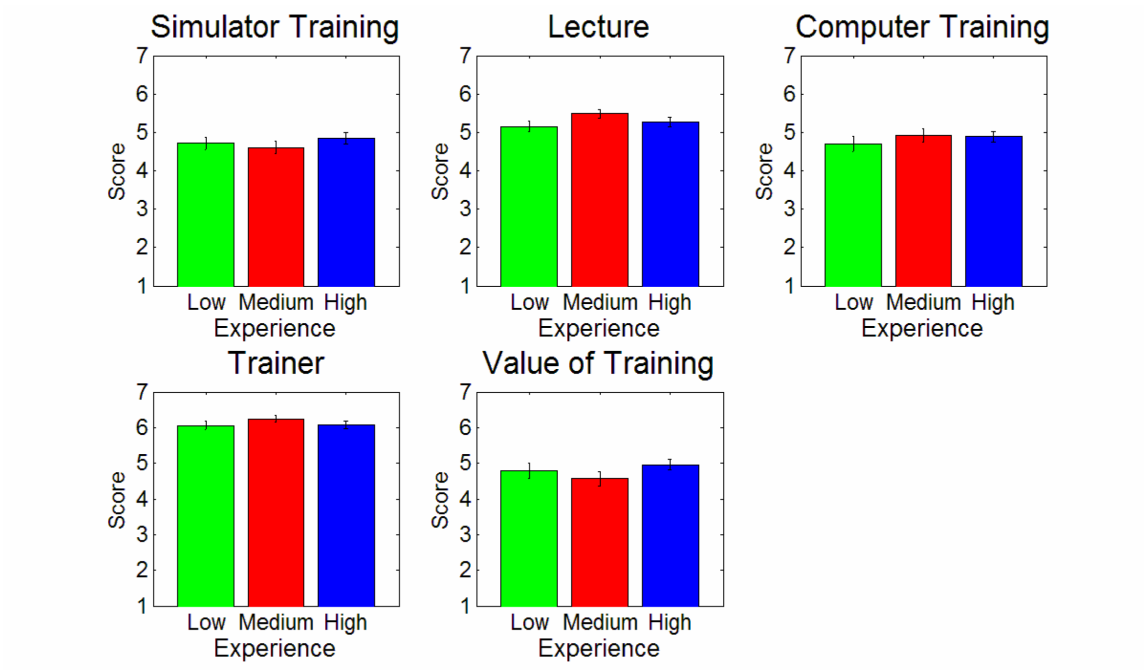


Figure 11. Mean ratings for the five Training categories for the three experience groups.

The mean response to each category was statistically compared to 4, the neutral point to determine whether participants' judgments were positive. All categories received liking ratings (all *t*s >7.18, all *p*s < .001); participants liked all aspects of training, especially those that involved the trainer (lecture and trainer). The uniformity of positive responses is conveyed in Figure 12, which presents the average proportion of participants who responded with dislike scores (1 to 3), a neutral score (4), or like scores (5 to 7) to each items comprising the five aspects of training. In all cases, over 55% of respondents 'liked' that aspect of training or found it valuable.

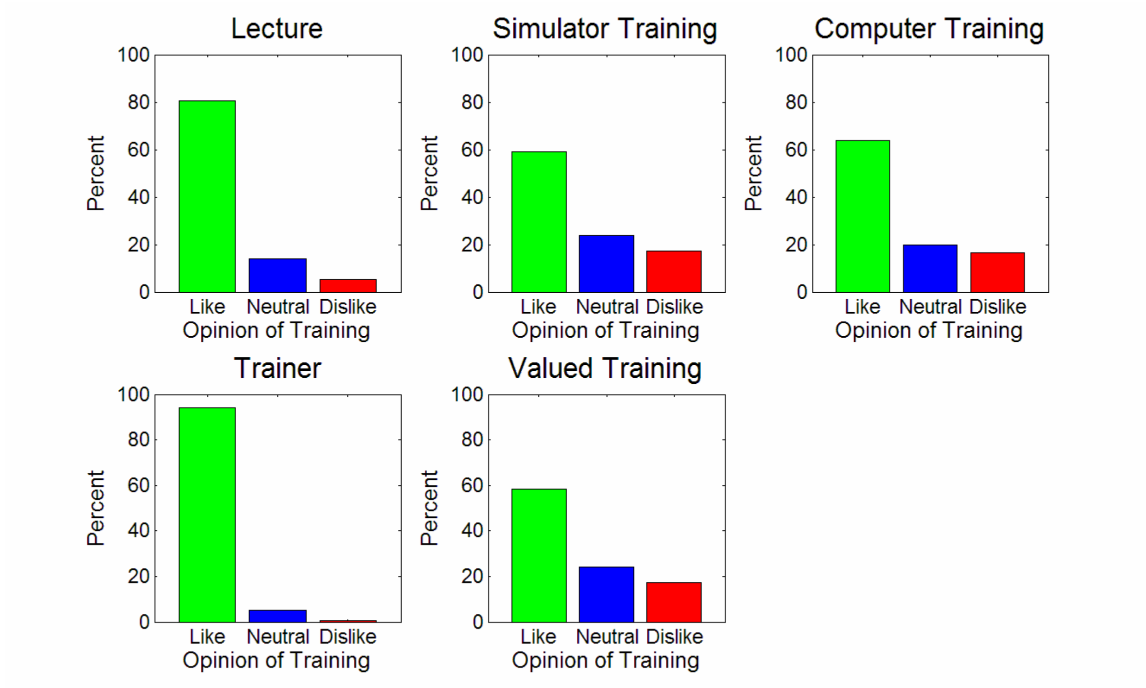


Figure 12. Percent of participants who liked, were neutral, or disliked, each of the five aspects of training.

Qualitative Training Questions

The six qualitative questions and a selection of the more common responses are shown in Table 4. The open-ended responses corroborated several aspects of the quantitative data. As implied by the lower score on the "driving a real snowplow" realism question, many operators commented that the driving simulator lacked some of the equipment that is normally present on snowplows, such as plow and wing controls. Others commented that the calibration of engine sound and speed seemed off, and sounds such as the scraping of the plow on the roadway, that are typically present during snowplowing, were absent in the simulator. Finally, in spite of the apparent discrepancies between the simulator and actual snowplow driving, many operators stated that they thought the driving simulator was the most useful part of training. Thus, although not every operator enjoyed the simulator training, many rated it as the most enjoyable or useful part of training.

Table 4. Selected responses for the qualitative questions portion of training questionnaire

Question	Selected Operators' Responses
1) Is there anything that can be done to make the scenery better?	- Difficult to judge distance of oncoming traffic More snow and slush on windshield during driving Majority thought it was fine or had no comments
2) Is there anything that can be done to make the sound better?	- Better calibration between engine sound and speed Add sounds from plow equipment (e.g., scraping on highway) Louder engine noise
3) Is there anything that can be done to make training better?	- Add controls for plow and wing More time driving the simulator Night time snowplowing
4) What was the most useful part of training?	- All: 15% Simulator/Driving: 40% Computer: 20% Lecture/Powerpoint: 30% Miscellaneous: 5%
5) What was the least Useful part of training?	- Simulator/Driving: 25% Miscellaneous: 15% Lecture: 5% Computer: 20% None (all useful): 35%
6) Do you have any additional comments?	- Simulator accelerates and brakes too fast Steering is too sensitive Training was useful, especially for new operators

Summary of Simulator Training and Realism Results

Similar to the immersion scores, operators from all experience levels reported similar amounts of presence in the simulator, suggesting that they should all benefit from simulator training to the same degree. Operators reported in the quantitative and qualitative sections that they thought the simulator was realistic, although they made some suggestions on aspects that could be improved. Responses to training were also quite positive, and participants reported that they particularly enjoyed the simulator and lecture portions. Finally, simulator sickness scores were overall quite low. Mean simulator sickness scores fell below the level of 'slight' discomfort, and less than 3% of participants in this study elected to end training early due to excessive cybersickness.

Performance Analysis

In this section, improvements in simulator performance that could be found as a result of participants experiencing this training regimen were examined. It is important to keep in mind that training only lasted between 1.5 and 2 hours, and during this time, each participant only drove the simulator for approximately 10 to 20 minutes. Subsequent simulator training provided by IDOT, when additional time does not need to be allocated to administering questionnaires or conducing untrained experimental drives, would almost certainly involve more training time, particularly for driving the simulator. Any differences found as a function of this abridged training regimen, therefore, would be notable. Performance data for participants' first and second drives in the simulator, which were recorded by the driving simulator software, were successfully collected from 124 participants. Note that this number reflects those operators who participated after the simulator upgrade. Three measures of performance were examined: number of collisions, average speed, and fuel efficiency.

Critical Failures – Collisions

Critical failures in this study refer to the number of collisions that participants were involved in while driving the simulator. In the tests reported by the simulator software, critical failures also included failure to wear a seatbelt. However, these violations were removed from this examination, and thus the new critical failures score was limited to accidents. Also, the collisions analysis was conducted on 154 participants, including 30 operators who participated prior to the simulator upgrade. Presumably, minor changes in the visual quality of the simulator should not have lead to any differences on such a broad measure of performance as number of collisions, and thus including these operators was deemed to be appropriate.

A 3 (experience level: low, medium, high) x 2 (group: experimental, control) x 2 (time: pretraining drive, postraining drive) ANOVA was conducted on the collision data. Experience level and time were within-subjects variables and group was a between-subjects variable. Note that this type of analysis will be repeated for subsequent performance, as well as fixation, comparisons. The analysis showed a main effect of group, $F(1, 148) = 4.50$, $MSE = .180$, $p<.05$, in that participants in the experimental group made more collisions than those in the control group. However, as can be seen in Figure 13, the difference between the groups was apparent on the pretraining drive. The fact that the group effect did not interact with time means that the difference cannot be attributed to training. For an unknown reason, the groups differed prior to the experimental manipulation and that difference remained after the manipulation, since the group effect did not interact with time.

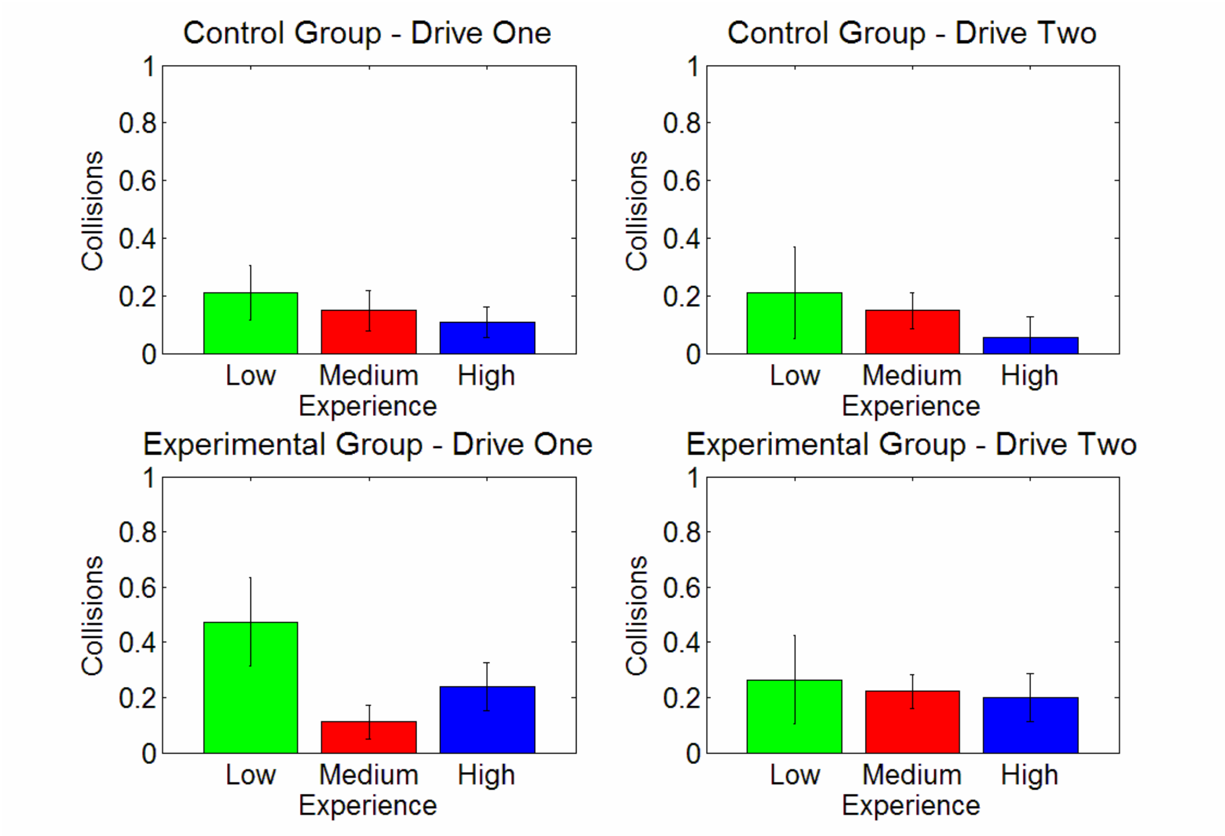


Figure 13. Mean number of Collisions for the three experience groups during the first and second drive.

The ANOVA also showed a trend towards a main effect of experience level, $F(2, 148) = 2.93$, $MSE = .180$, $p = .057$. LSD comparisons among the experience groups showed that the low experience participants were involved in more accidents than the medium or high experience groups, who did not differ from each other. Thus, experienced drivers, who presumably had a longer time to refine their skills in the field prior to receiving simulator training, made fewer collisions in the driving simulator than less experienced operators who had not had as much snowplowing experience.

We had expected that training might be more beneficial for low experience drivers than for more experienced drivers, but the Experience level X Group X Time interaction was not reliable (F<1). Low experience participants in the experimental group showed more improvement from the first to second drive than the low experience participants in the control group, but the apparent benefit was not statistically significant. One problematic aspect of the data deserves comment: just what was counted as an accident by the simulator software is unclear. Each of the drives lasted only 10 minutes, but the number of critical accidents ranged from 0 to 3. It may be that variations in the scenario not under participant control contributed to differences in performance. For instance, an "accident" when no other vehicles were programmed to approach within the next few seconds may have yielded a count of one critical error while an "accident" when another vehicle was programmed to approach in the following few seconds may have yielded a count of more than one critical error (i.e., if it too collided with the operator's vehicle). The lack of training effects may simply reflect the fact that there was not sufficient change in the two hour training period to be detected in an immediately following 10 minute drive, but aspects of the simulation programming as just described may also have introduced variability in the data that reduced the power to detect real differences. In support of the latter is the observation that if those participants who have three "accidents" during one of their drives are excluded from the analyses (2 cases), the predicted Experience level X Group X Time interaction is reliable ($F(1, 151) = 4.16$, MSE = .176, p<.05), with low experience drivers showing a reliable benefit for training.

Speed and Fuel Measures

Speed and fuel comparisons were made between operator's first and second experimental drives in the simulator. Figure 14 shows the mean average speed for the three experience levels across their first and second drive, organized by whether they were in the experimental or control group. A pair of 3 (experience) X 2 (group) X 2 (time) ANOVAs were run to determine whether training or experience level had any affect on operators' average speed or average fuel consumption for the first or second drives. First, a main effect of drive was found for average speed ($F(1, 118) = 32.22$, $MSE = 10.56$, $p<.001$), with operators tending to drive faster on their second drive compared to their first. This was qualified, however, by a reliable interaction with experimental group ($F(1, 118) = 10.70$, $MSE = 10.56$, $p<.01$). The experimental group drove significantly faster on drive two compared to drive one (mean change = 3.86, SD = 4.77, $t(57) = 6.17$, $SE = 0.63$ $p<.001$), while the increase in the control group was smaller (mean change = 0.99, SD = 4.34, $t(65) = 1.84$, $SE = 0.53$ $p=.07$).

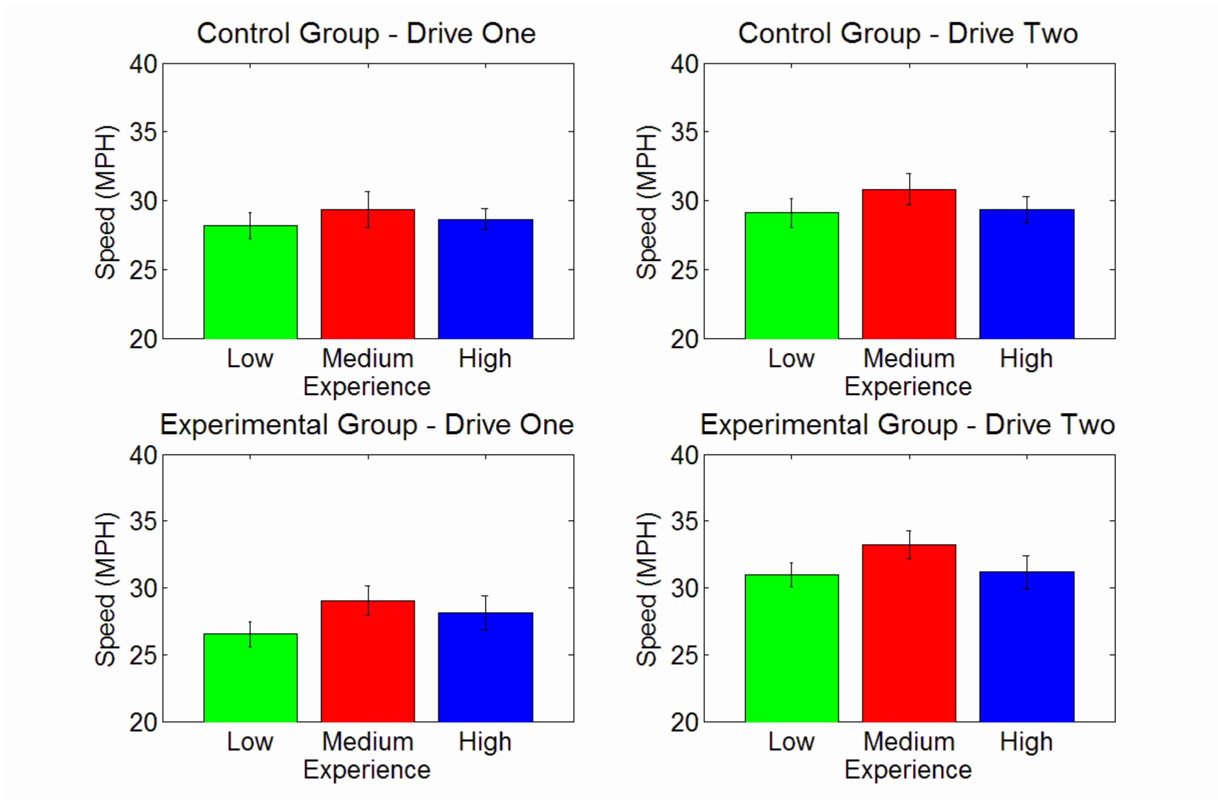


Figure 14. Mean average speed for the three experience groups during the first and second drive.

When operators' average fuel consumption data were examined (see Figure 15), a statistically significant interaction between experimental condition and drive number was found ($F(1, 118) = 6.83$, $MSE = .013$, $p=.01$). This showed that participants in the experimental group showed a trend for slightly lower fuel efficiency in drive two compared to drive one (mean change = 0.04, SD = 0.17, $t(57) = 1.75$, $SE = 0.022$ $p=.085$), while operators in the control group showed a slight improvement (mean change = 0.05, SD = 0.15, $t(65) = 2.43$, $SE = 0.018$ $p<.05$). However, this absolute difference in fuel consumption between the two groups is relatively minor. While it is difficult to determine what a 0.08 average difference between the two groups in fuel consumption as measured by the simulator would translate to under real life estimates, it appears as if operators in the experimental condition tended to drive faster than operators in the control group during their second drive, after they received training, while showing a negligible increase in fuel. Thus, training may have lead to a positive effect on operators' driving performance in the simulator. There was also a main effect of experience ($F(1, 118) = 3.62$, $MSE = .050$, $p<.05$), with an LSD comparison revealing that low experience operators had higher fuel efficiency than medium experienced or high experience operators. The absolute difference was quite small, though, but it may reflect a tendency for low experienced operators to drive more conservatively.

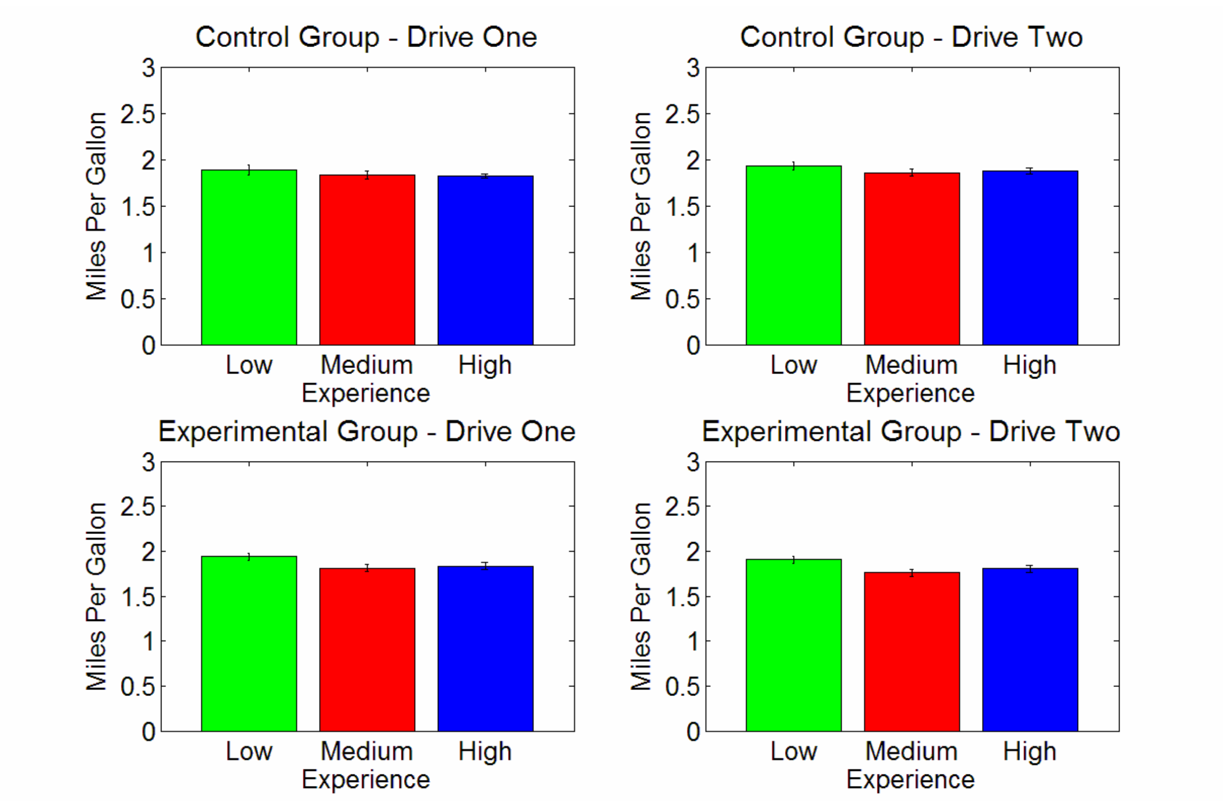


Figure 15. Mean fuel efficiency for the three experience groups during the first and second drive.

Summary of Performance Results

There was some indication that operators who received training performed better on their second drive compared to operators who did not receive training. Specifically, these operators drove faster on their second drive while showing a negligible increase in fuel consumption. Additionally, low experience operators who received training may have shown a tendency to get into fewer collisions than those who did not, but there were difficulties interpreting this analysis. Finally, low experienced operators were involved in more accidents than medium experienced or high experienced operators.

Tracking

Participants' fixation behaviors were examined on their first and second experimental drives in the snowplow simulator. Head tracking data were successfully collected for 119 participants, and eye tracking data were collected for 72 participants. Note that some of these operators participated prior to the simulator upgrade. Three categories of tracking analysis were examined: the percentage of fixation time on and number of fixations made to certain critical objects (center screen, right panel, left panel, right spot mirror, left spot mirror, right wing mirror, left wing mirror and speedometer), horizontal and vertical spread of search, and eye closure.

Object Fixation

The percentage of time that each participant fixated on the center panel was examined as a function of experience. A 3 (experience) X 2 (group) X 2 (time) ANOVA examining the amount of time that operators fixated on the center panel revealed no differences as a function of experimental group, experience or time (all ps $> .2$). Thus, participants in all groups and levels of experience spent approximately equal amounts of time fixating on the center screen in their first and second drive.

Next, the fixation pattern of operators in each experience group was examined to determine whether training influenced the relative amount of time that operators spent looking at objects other than the center panel. In order to take into account differences among operators in the absolute amount of time spent on the center panel, the amount of time *not* looking at the center panel was determined for each operator. The proportion of this time that was spent on each object was calculated. For instance, if an operator spent 20% of the time fixating on objects other than the center panel, the proportion of that amount of time that he or she looked at each other object was determined. Multiple 2 (group) X 2 (time) ANOVAs were conducted for each of the critical objects described above. No differences were found for operators in the high experience group as a function of experimental versus control group or drive number, and some differences were found for operators in the medium experience group. Specifically, these operators spent less time looking at the right panel ($F(1, 43) = 4.82$, $MSE = 353.2$, $p<.05$) and tended to spend less time looking at the right wing mirror ($F(1, 43) = 3.19$, $MSE = 55.4$, $p=.08$). However, a consistent pattern emerged for operators in the low experience group: these operators spent less time looking at objects on the right side than on the left side of the simulator on their second drive compared to their first. Specifically, low experience operators spent less time looking at the right panel ($F(1, 30) = 3.08$, $MSE = 303.0$, $p=.09$) and the right spot mirror ($F(1, 30) = 4.65$, $MSE = 102.9$, $p<.05$), and more time looking at the left panel ($F(1, 30) = 2.95$, $MSE = 77.9$, $p=.09$) and the left wing mirror ($F(1, 30) = 4.42$, $MSE = 64.6$, $p<.05$) (see Figure 16).

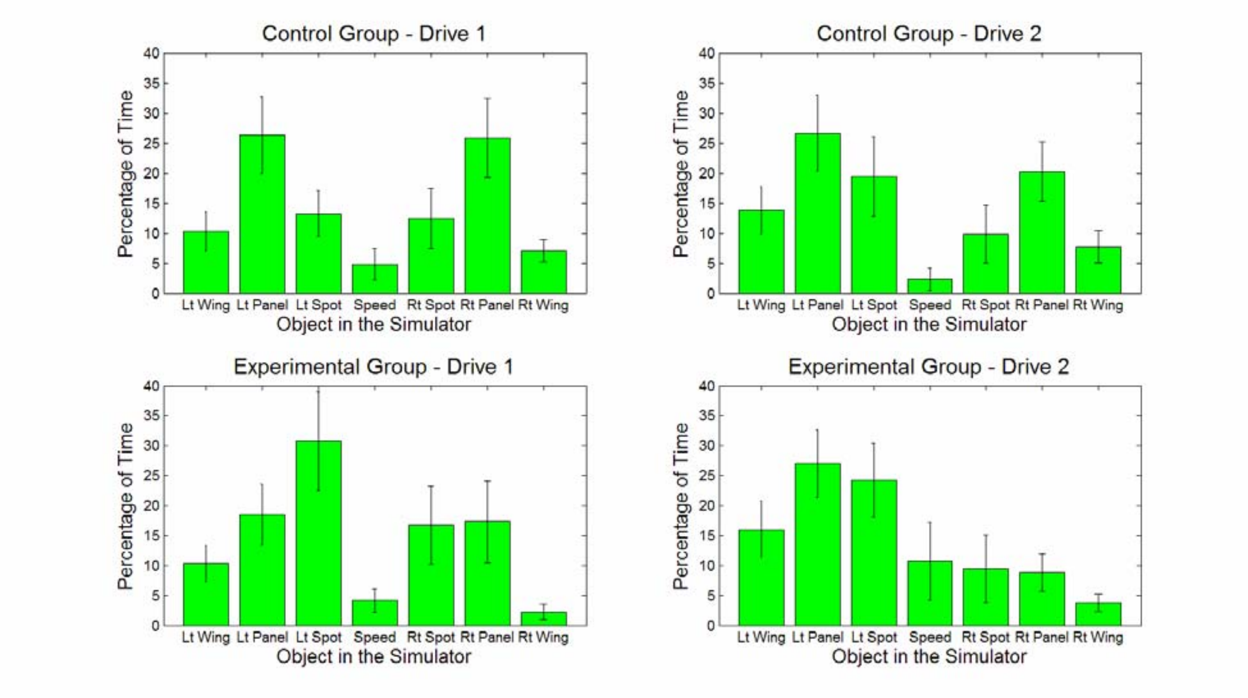


Figure 16. Mean percentage of fixation time outside of the center panel on objects for Low experienced drivers in the control and experimental group on the first and second drives.

This pattern was confirmed by examining the percentage of time spent looking outside of the center panel collapsed across objects on the left versus right side (see Figure 17). In other words, the amount of time operators in the low experience group spent looking at the left compared to the right side of the center panel on their first and second drives was determined using a 2 (left versus right) X 2 (time) ANOVA. This analysis showed a main effect for location ($F(1, 30) = 5.93$, $MSE = 645.4$, $p<.05$), suggesting that operators spent more looking at the left versus the right side of the snowplow. This was qualified by a reliable location by time interaction ($F(1, 30) = 6.12$, $MSE = 645.4$, $p<.05$), which showed that operators decreased the amount of time spent looking at the right side on their second drive compared to the first, and fixated more often on the left side during their second drive compared to the first.

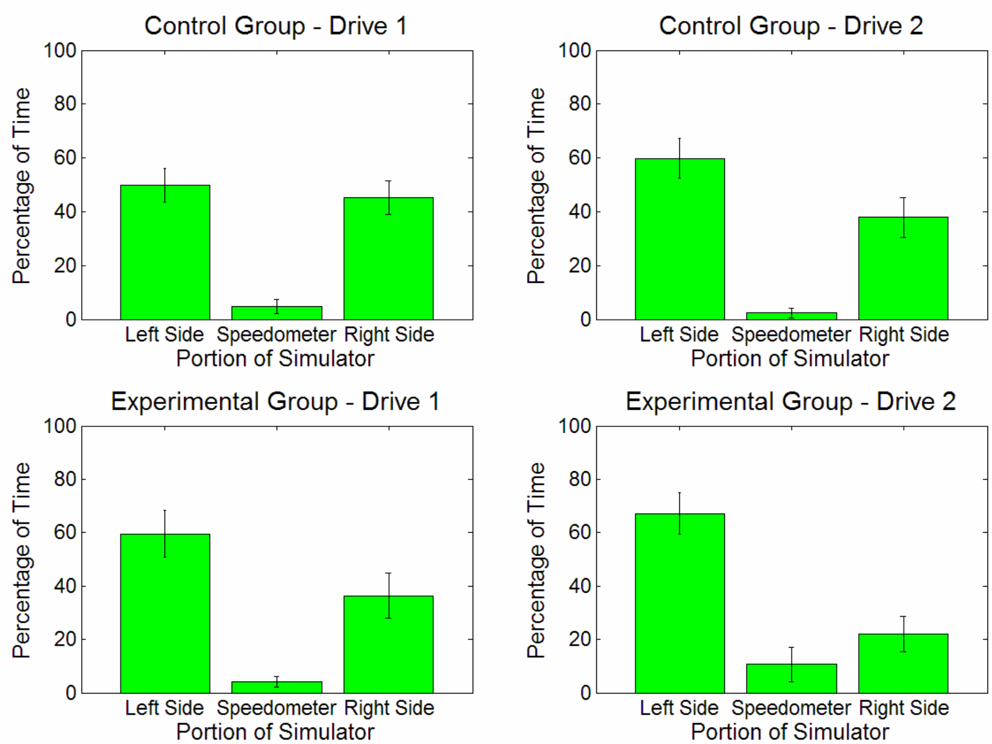


Figure 17. Mean percentage of fixation time outside of the center panel on left and ride side for Low experienced drivers in the control and experimental group on the first and second drives.

Finally, an analysis of object fixation changes was conducted in an attempt to provide converging evidence for these findings. An object fixation change in this study is defined by a fixation change from one object to another. For example, a participant looking from the left spot mirror to the center panel would constitute a fixation change away from the left spot mirror. Multiple 2 (group) X 2 (time) ANOVAs were conducted to examine the number of fixations made from each critical object in the simulator (see Figure 18). Overall, the general pattern that was found for time spent fixating on the objects outside of the center panel was replicated in this analysis. Specifically, low experience operators made fewer fixations to the right spot mirror ($F(1, 31) = 6.15$, $MSE = .002$, $p<.05$) and more fixations to the left panel ($F(1, 31) = 7.48$, $MSE = .003$, $p=.01$) and the left wing mirror ($F(1, 31) = 8.48$, $MSE = .002$, $p<.01$) on their second drive compared to their first drive. A main effect of time was also found for the middle panel ($F(1, 31) = 4.61$, $MSE = .003$, $p<.05$), demonstrating that operators made fewer fixations to the center panel on their second drive. Additionally, the right spot mirror reduction was qualified by a significant group interaction ($F(1, 31) = 6.51$, $MSE = .002$, $p<.05$), which showed that operators in the experimental group looked at this location less on their second drive than operators in the control group. These analyses of fixation changes replicate the general pattern of head fixation duration on the critical objects, demonstrating that low experienced operators spent less time fixating on the right side and more time on the left side of the simulator on their second drive compared to their first.

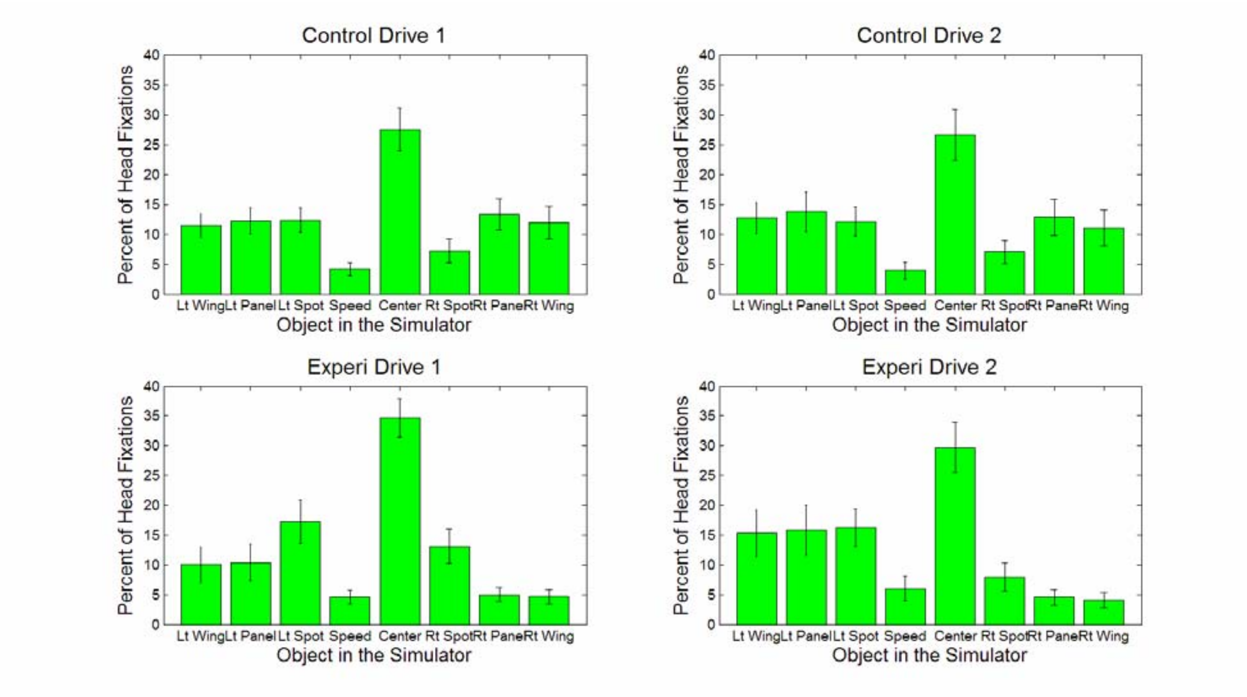


Figure 18. Mean percent of fixations towards each object for Low experienced drivers in the control and experimental group on the first and second drives.

To clarify, these results show that low experienced operators in both the control and experimental groups showed a tendency to look less to the right side of the simulator and more to the left side on their second experimental drive compared to their first. Since this pattern occurred both for the control and experimental groups (i.e., there were generally no group interactions), these findings cannot be attributed to training. Instead, the most likely explanation is that low experience operators in both groups learned during their first drive that there was an advantage for looking out the left side of the vehicle, and thus they altered their scanning behaviors for their second drive. Specifically, other simulated vehicles, that were most likely to cause a collision with the driver, frequently passed on the left side of the vehicle.

Spread of Search

Spread of search is defined as the variation in fixation location across the horizontal and vertical axes. For instance, an operator who spends more time checking his or her mirrors would have a larger spread of horizontal search, as he or she is fixating on objects that are farther away from the center, than an operator who fixates mostly on the center screen, as he or she would focus primarily on a small, restricted area of the screen. Two 3 (experience) X 2 (group) X 2 (time) ANOVAs examining operators horizontal and vertical spread of search both revealed a main effect of time ($F(1, 113) = 6.34$, $MSE < .001$, $p<.05$, and $F(1, 113) = 10.91$, $MSE = .001$, $p=.001$, respectively). Regardless of experience or group, operators showed a reduced spread of search in their second drive compared to their first drive (see Figure 19 and Figure 20). It is important to reiterate that these graphs represent the standard deviations of participants' scanning behaviors, not mean locations. In other words, a value of 5 degrees on spread of horizontal search would indicate that approximately 95% of all fixations occurred within 20 degrees of the horizontal mean. Thus, the differences represented in Figures 19 and 20 between groups of operators are larger than they may appear. Additionally, the larger values seen for vertical spread of search compared to horizontal search may reflect that operators primarily kept their gaze level horizontally, but frequently checked their speed on the speedometer. This would lead to the observed pattern of results, specifically, lower amounts of horizontal than vertical spread of search.

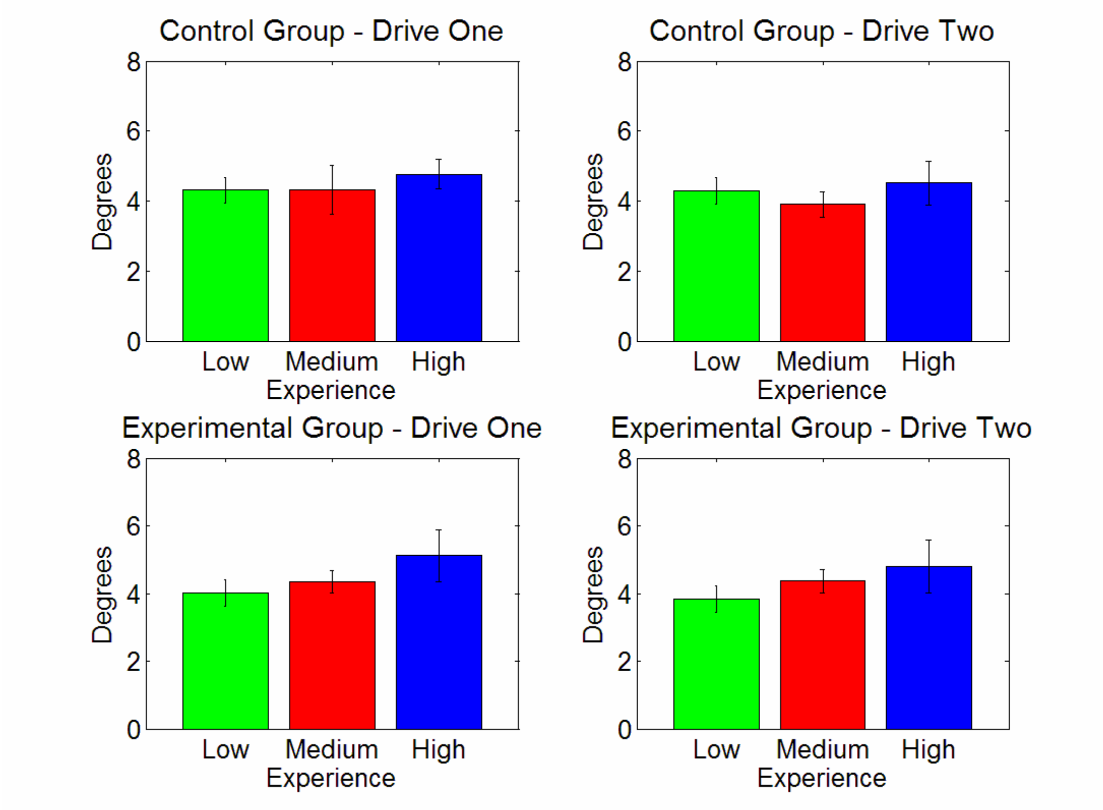


Figure 19. Mean standard deviation of horizontal search for the three experience groups in the control versus experimental conditions during the first and second drive.

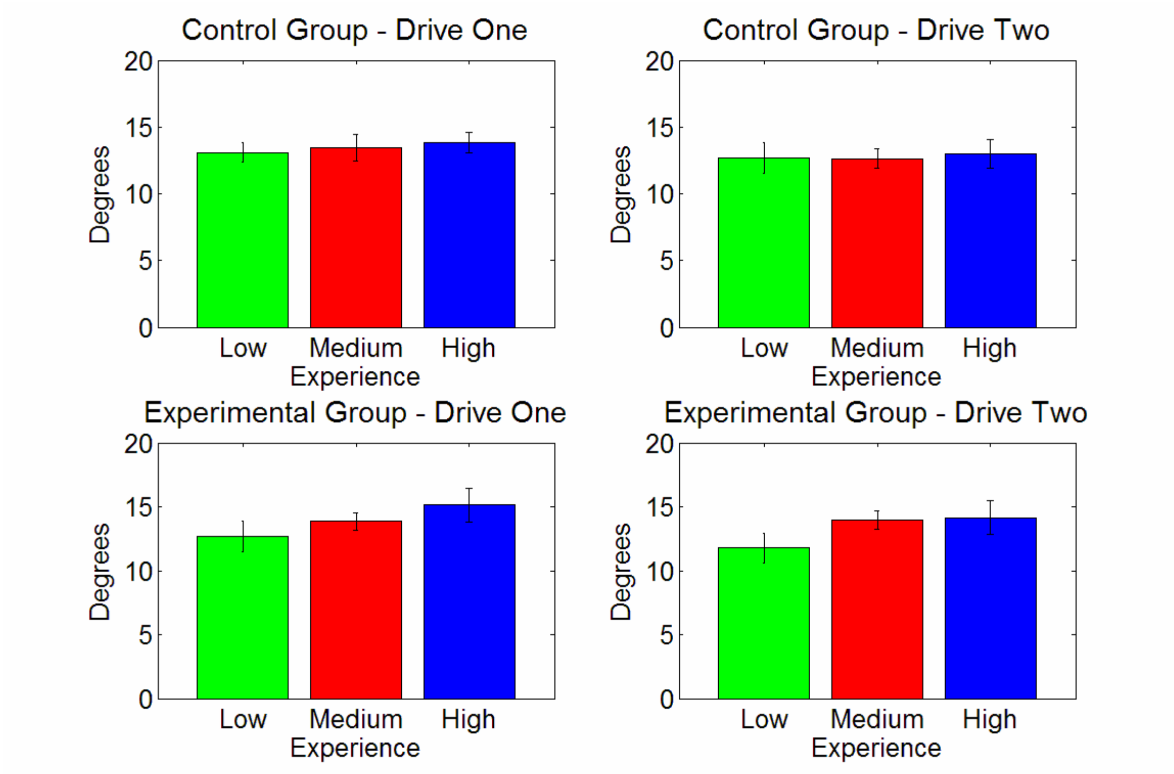


Figure 20. Mean standard deviation of vertical search for the three experience groups in the control versus experimental conditions during the first and second drive.

Because this pattern was found in both the control and experimental groups, it is unclear as to what led to these differences. One possible explanation is that operators adapted to the simulator by their second drive, and thus they may have been behaving similarly to how they would in a real snowplow. Another explanation is that both groups may have learned that there was a benefit for spending more time looking at the center of the road during the first drive in that scenario, for instance, if that's where the majority of hazards appeared from. Thus, they chose to concentrate more on the center of the roadway during their second drive.

Eye Closure

Research on eye movements suggests that more blinking and a greater amount of eye closure is typically associated with boredom or low effort tasks. Specifically, more blinking and numerically greater eye closures is taken as evidence that participants are comfortable or exerting less effort on the task. Eye closure, as measured by the eye tracker, refers to the distance between the eye lids, with a smaller distance representing more eye closure. A pair of 3 (experience) X 2 (group) X 2 (time) ANOVAs were conducted on participants' blink frequency, as well as their overall percentage of eye closure. The analysis of blink frequency revealed a main effect of time ($F(1, 63) = 4.97$, $MSE = .005$ $p < .05$), suggesting that drivers tended to blink more during their second drive. Additionally, while the main effect of experience did not reach statistical significance ($p = .12$), an LSD comparison showed that high experienced operators tended to blink more often than low experienced operators (see Figure 21).

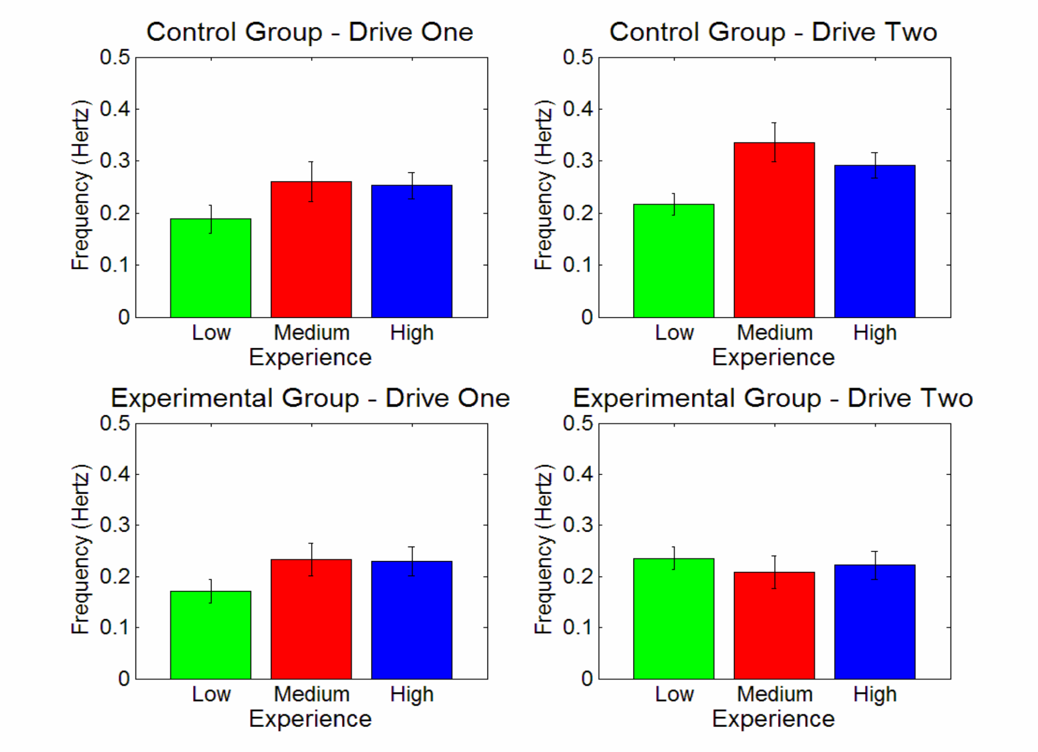


Figure 21. Mean blink frequency for the three experience groups in the control versus experimental conditions during the first and second drive.

The results from the analysis of eye closure converged with these findings. The ANOVA revealed a trend for a main effect of time ($F(1, 63) = 3.56$, $MSE<.001$, $p=.064$), with drivers tending to have their eyes closed more during their second drive than their first. Moreover, although the main effect of experience once again failed to reach statistical significance ($p=.135$), an LSD comparison showed that high experienced participants had their eyes closed more than medium experienced operators (see Figure 22). Taken together, these results suggest that operators may have found their second drive less difficult than their first, and that the simulator driving may have been less demanding for highly experienced operators.

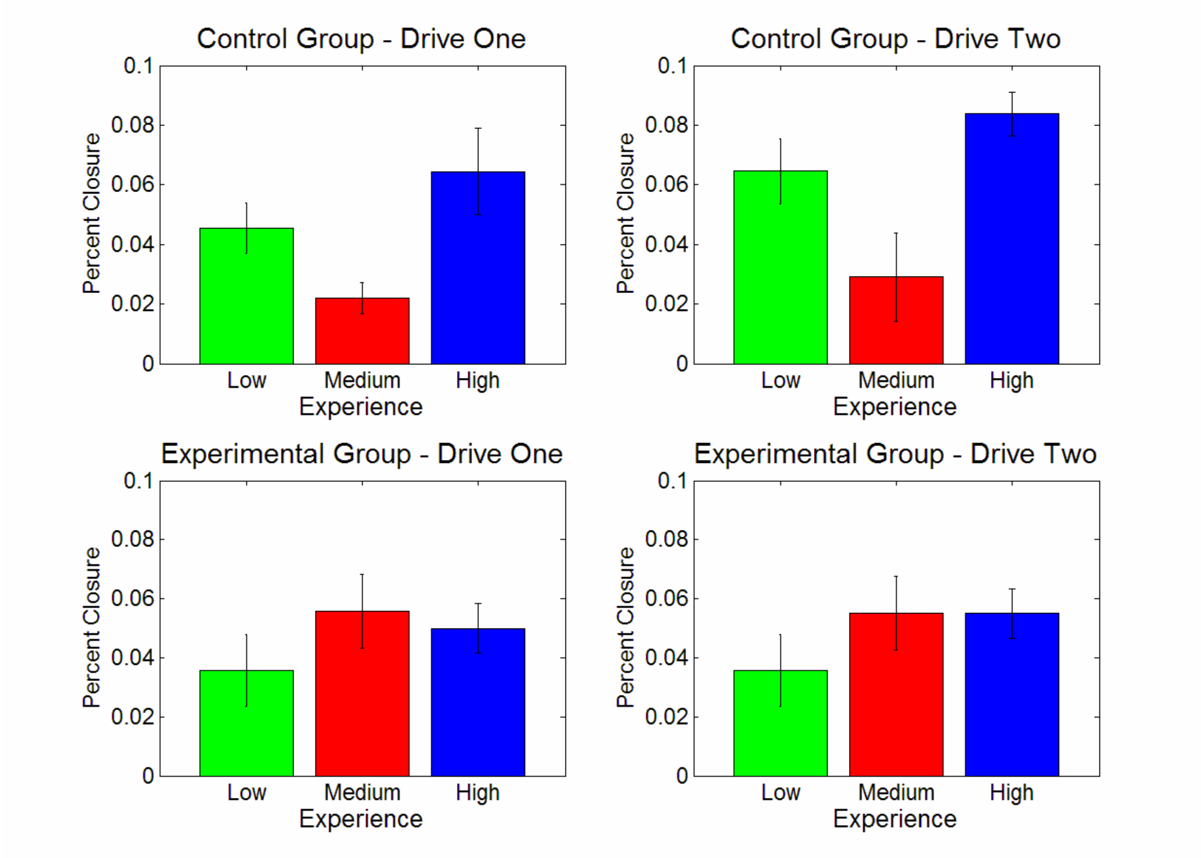


Figure 22. Mean percent eye closure for the three experience groups in the control versus experimental conditions groups during the first and second drive.

Summary of Tracking Results

The analyses of object fixation time and fixation changes showed that operators in the Low experience group demonstrated a pattern of looking more towards the left side of the simulator on their second drive and less time spent looking towards the right side. Since this trend was found for operators in the experimental and control groups, the most likely explanation is that these drivers learned that there was a benefit for looking towards the left side of the simulator during their first experimental drive. The spread of search analyses showed that operators tended to have wider horizontal and vertical scan paths during their first drive compared to their second drive. However, this difference was quite small, and may have been due to operators adopting a more narrow searching strategy due to already driving the scenario and thus knowing where many of the critical events were. Finally, high experience operators showed a tendency to have their eyes closed more than other drivers, implying that the task of driving in the simulator was not as strenuous for them.

Summary of Results

Overall, there were few personality or dispositional differences between less experienced and more experienced snowplow operators. What differences that were found typically disappeared once age was taken into account by treating it as a covariant. Thus, personality variables, measured by the NEO-FFI, Zuckerman's (1979) sensation seeking questionnaire, or an immersion questionnaire (Whitmer & Singer, 1994), did not appear to differ as a function of experience. There did appear to be some differences between snowplow operators' responses to the NEO-FFI questionnaire and the sensation seeking questionnaire compared to studies that report scores for a normative American male population. However, it is premature to make any decisions based on these data without first comparing the responses of snowplow operators to a more appropriate sample, such as non-snowplow operating IDOT employees.

Overall levels of reported simulator sickness were modest. Based on the scale used in the questionnaire, the mean responses for the categories measured were below the level of response of 'slight' discomfort for all symptoms. Features of the simulator were rated by operators as somewhat realistic, as seven out of the eight features investigated by the questionnaire were above a realism rating of 'neutral.' Finally, operators' reactions to training were quite positive. Many operators reported in their qualitative comments that the driving simulator training was the most useful aspect of training, and the lecture and computer based training were rated highly, also.

Measures of operators' performance in the simulator during the two experimental drives indicated that less experienced drivers tended to be involved in more collisions than the other two groups of more experienced drivers. The most likely explanation for this finding is that experienced drivers managed to apply the skills they learned in the field to operating the simulator. Moreover, low experience drivers may have shown more improvement as a result of training, but difficulties interpreting the results of the performance output files make any conclusions tenuous. Drivers from all experience levels in the experimental group drove faster during their second drive while consuming a comparable amount of fuel to the control group operators. Thus, the driving simulator may be a practical tool for teaching drivers better fuel management techniques.

Finally, there were some fixation differences for operators in the low experience group between their first and second drives in the simulator. Specifically, they reduced the amount of time looking towards the right side, and increased the amount of time they spent looking towards the left side of the simulator. Also, there did appear to be some evidence that drivers adopted a more narrow spread of search for their second drive compared to their first regardless of whether or not they received training. One interpretation of this finding is that by the second drive the majority of participants had adapted to the simulator and thus their fixation behaviors may have been mimicking how they would scan while driving a real snowplow, although since this difference was so small it may have simply occurred by chance. Additionally, more experienced drivers tended to blink more often than low experienced drivers, which in the driving research literature is taken as an indication that they were not concentrating as hard or that the task was not as strenuous for them. Thus, converging with the evidence on collision frequency, these results suggest that more experienced operators had an easier time operating the simulator than low experience operators.

CHAPTER 5. GENERAL DISCUSSION

The results of this study suggest that Iowa DOT operators enjoyed and seemed to benefit from virtual reality snowplow simulator training. Operators from all age groups and levels of experience reported having similar immersive tendencies in their everyday life and experienced a similar amount of presence within the simulator. One interpretation is that operators of all ages and levels of experience have the potential to benefit from training equally well. The responses to the training questionnaire tend to support this explanation: operators from all three levels of experience rated all aspects of training similarly. Additionally, although there was a lot of variation in reported simulator sickness scores, mean ratings were relatively low. There was some evidence that amounts of simulator sickness increased with age, although the mean ratings for all experience groups were less than what would constitute a response of 'slight' for the listed symptoms. These ratings will likely be even lower when training is given at subsequent times, because this training regimen required additional time for the experimental protocol that could be used to acclimate operators with the simulator.

The most positively rated aspect of training was the lecture portion, and operators also rated the quality of the trainers, who gave the lectures, very highly. The other two components of training, computer training and simulator training, were rated almost equally and were generally well received. Approximately 80% of participants were not opposed to simulator or computer training, and many reported on the free response section that the simulator portion was the most useful aspect of training. Moreover, many of the responses to the free response questions elucidated what operators felt was lacking from simulator training. These comments will be discussed shortly.

The performance data suggest that there were some noticeable benefits of training. Specifically, low experience operators who received training prior to their second experimental drive may have shown an improvement in the number of collisions they accrued from their first drive than low experience operators who did not receive training before their second experimental drive. Difficulties in interpreting the performance files from the simulator software, however, make analyzing these results problematic. This finding was actually somewhat unexpected, considering that training only lasted for around 1.5 to 2 hours. Training programs for less experienced drivers that are able to devote more time to training, particularly within the driving simulator, would have a better chance of demonstrating performance improvements.

Operators in the experimental group also showed a significant increase in their average speed in their second drive compared to their first drive, as well as compared to the second drive for operators in the control group. Importantly, this increase in speed came with a negligible increase in fuel consumption. Thus, this finding also shows that drivers who received training tended to perform better in the simulator than drivers who did not receive training before their second experimental drive.

Low experience operators seemed to intuitively adapt to driving the simulator. Specifically, operators in the control and experimental groups spent more time looking at the left panel and left wing mirror and less time looking at the right panel and right spot mirror on their second drive compared to their first. Possibly, they may have concluded during the first drive that there was a benefit of looking at the left side of the simulator as opposed to the right, or they may have become more comfortable with the virtual reality environment and started behaving as they normally do when driving a snowplow.

Finally, there were some eye closure differences between operators as a function of experience. More experienced operators demonstrated tendencies associated with being more comfortable or not needing to exert as much effort on driving. Specifically, they tended to show a higher blink frequency and more eye closure than low experience operators. Both of these behaviors have been interpreted to mean that the driver is not concentrating as hard on the task.

Thus, there are several indications that drivers were behaving in the simulator as one would expect based on their amount of snowplowing experience. Low experienced operators showed a tendency to be involved in more accidents than more experienced operators, while experienced operators seemed not to concentrate as hard on driving in the simulator. Training also seemed to benefit novice operators in particular, and there is some evidence to show that all drivers who received training used better driving habits on their second experimental drive.

The major caveat of this study, though, is that these results only directly apply to behaviors within the simulator. Although drivers who received training appeared to perform better on their second drive than those who had not had training at that point, it is unclear how these benefits would transfer to real world snowplowing behaviors. Previous research has shown that individuals tend to behave similarly in driving simulators as they do in real world settings (Godley, Triggs, & Fildes, 2002; Panerai et al., 2001; Tornros, 1998), although this has never directly been tested for snowplowing. Another potential concern is that the benefits for the experimental drivers may simply reflect the fact that these operators had additional time to drive the simulator during training. This explanation is actually unlikely as drivers only drove two 5 to 10-minute scenarios during training, which involved truck driving rather than snowplow operating. Thus, the benefits seen as a result of training were not likely due solely to the fact that participants in the experimental group spent more time in the simulator. Information learned during the lecture and computer based portions of training were undoubtedly influential on improving operators' performance as well.

There are also some other minor concerns with the experimental paradigm. First, several different trainers were used throughout this study. Although all trainers used the same training scenarios and followed a predetermined script for the lecture, operators may have responded to some trainers more positively than others. Potentially, this may have increased the variance in participants' data, thus reducing the likelihood of finding statistically significant results. Additionally, the simulator upgrade at the end of the third week of training excluded approximately 35 participants from being included in many of the performance analyses. Several of the thresholds for causing certain types of failure were altered during the upgrade, and thus it would have been misleading to include these operators in the statistical analyses for these performance measures. These participants were, however, included in the questionnaire, fixation and collision measures, as the minor improvements in simulator quality made during the upgrade were not likely to affect any of these measures.

It is important to keep in mind, also, that each session lasted for approximately four hours, and around half of that time was allocated to tasks involving this experiment. Training itself only lasted on average from 1.5 to 2 hours, and as mentioned above, operators were only scheduled for two trained drives in the simulator, which even involved non-snowplow scenarios. Although the purpose of training was to disseminate better driving practices rather than teaching operators how to plow snow per se, allocating more time for training would increase the chances that skills learned during training would generalize to real world performance.

There are certainly several methods to judge the effectiveness or generalizability of simulator training in real world situations. One possibility would be to compare the performance, during one or more winter seasons, of a group of operators who receive training versus a matched control group who did not receive training. However, this would require a large portion of drivers to forgo what may be highly valuable training. Thus, a possible alternative would be to train as many drivers as possible, even bringing operators back for multiple sessions if possible, and compare overall collision or fuel consumption rates for the next few years to an equal number of previous years. Although this method lacks complete experimental control, it allows for a greater number of operators to receive training and should provide a critical examination of the effectiveness of this particular training regimen in real world situations. Conversely, short term tests are also an option. For instance, a group of trained operators, and a matched control group, could be tested on a series of courses designed to approximate the demands of operating a snowplow. Differences between the groups' performance could then be attributed to whether or not they received training. This option may not be ideal, though, as there would still be concern over the generalizability of those results to actual snowplow operation in the field.

Another area of interest may be to improve the simulator itself based on participants' recommendations. Generally, the simulator reality ratings were consistent with participants' free response comments. For instance, one of the lower realism ratings was for the responsiveness of the controls. Accordingly, many operators noted that: a) the relationship between engine sound and speed was incongruous, b) steering was too sensitive, and c) the vehicle tended to accelerate and slow down unrealistically fast. Several operators also commented that driving conditions during storms were not realistically captured by the simulator. For instance, they noted that visibility is typically poorer, and there is usually more snow and slush on the windshield. All of these suggestions, are easily addressable. Many operators also suggested that the simulator would be a much more effective training tool if the driver had control over the wing and plow, and if features such as a salt spreader were included. Many of these potential upgrades reported currently being added or have already been implemented in the snowplow simulator.

Another interesting finding, which may warrant some additional research, was that participants reported some differences across the five NEO-FFI personality sub scales compared to a population of normative American males. Additionally, operators also reported less sensation seeking tendencies than the general population, although this may be an artifact of only examining some of the subscales or an overall higher mean age in this sample compared to the normative group. To make any strong inferences about personality differences between IDOT operators that may be characteristic of becoming a successful snowplow operator, it would be necessary to identify a more suitable control group, for instance, IDOT employees who do not operate snowplows. This would control for important factors such as age, race and socioeconomic class that may have been responsible for the differences.

While these findings suggest that trained participants gained some information on more effective driving techniques from receiving training, it may be useful to speculate on the type of knowledge that training provided. When considering skills that have a motor component, a distinction can be made between declarative knowledge, which deals with facts or knowing 'what,' and procedural knowledge, which is associated with actual performance or knowing 'how' (Fitts & Posner, 1967). Declarative knowledge is considered to be explicit and under conscious control, while procedural knowledge is believed to be automatic or unconscious (Shimamura & Squire, 1984). For example, being able to describe which gear to shift into, given a certain speed and driving conditions, is an example of declarative knowledge, but automatically shifting into the correct gear without thinking about it, as may be done by an experienced driver with a lot of practice is an example of procedural knowledge. Fitts and Posner also identified an intermediate stage between declarative and procedural knowledge, the associative stage, where people begin to implement their declarative knowledge, but do so with a lot of effort and conscious control. Thus, a question of interest is whether operators gained declarative or procedural knowledge or both from training.

The different aspects of training (lecture, computer, and simulator) emphasized different types of knowledge. By definition, because they only dealt with disseminating new facts or information, lecture and computer training conveyed declarative knowledge only. For instance, during these sessions, participants were given instructions concerning better scanning techniques and better speed and space management recommendations. Because participants did not have an opportunity to apply or practice this information, they would have only gained declarative knowledge.

Simulator training, in contrast, likely allowed both declarative and procedural knowledge, or at least aided the transition from declarative to procedural knowledge (i.e., the associative stage). Typically, procedural knowledge is described as automatically and rapidly performing an action without having to consciously initiate each step involved in the task.

For instance, athletes and musicians do not explicitly think about the process needed to make contact with a ball or to play a sheet of music. Instead, the processes appear to occur outside of conscious awareness, and the necessary behaviors are apparently initiated automatically. Understandably, this requires a lot of practice. In terms of procedural knowledge with regards to driving, this may be demonstrated by automatically performing a correct maneuver under dangerous circumstances when there is not time to plan the correct actions. For example, snowplow operators may know what to do when a tire blows (i.e., they may have declarative knowledge), but without being able to practice and implement the correct response (i.e., without the opportunity to develop procedural knowledge), they still might not react in time to avoid an accident.

The approximately 20 minutes that participants drove the simulator are insufficient for the information learned during training to become completely procedural in nature, however, simulator training did provide participants an opportunity to practice the material that they had just learned. Thus, the most likely explanation for the impact of snowplow simulator training on participants' driving behaviors is that it affected learning in the associative stage. Moreover, if longer training sessions are conducted during the summer, operators will have the opportunity to gain more practice at using the declarative knowledge learned during the other phases of training, and hence might advance further through the associative stage. Hence, simulator training may be particularly useful for new operators if enough time can be allocated to allow them to practice the good driving techniques that they have learned in a declarative fashion. Simulator training in which the driver must actually react to hazardous situations like a blown tire may better enable the driver to react appropriately in a real world situation. Therefore, while the simulator may not provide procedural knowledge per se, it gives operators a chance to demonstrate and use declarative knowledge gained in other aspects of training, and may aid in eventually automatizing the skills learned during training, particularly if participants have to wait several months before otherwise being able to use the material that they learned during training.

Overall, this study found that: 1) On the basis of respondents' immersion and presence scores, operators of all levels of experience should be able to benefit from virtual reality snowplow simulator training; 2) Simulator sickness ratings were relatively low, implying that the simulator is appropriate for training a wide range of IDOT operators; 3) IDOT operators tended to enjoy training, and many reported that simulator training was the most useful aspect of training for them; 4) There is some evidence that operators who received training prior to completing their second drive in the simulator performed better than operators in the control condition; 5) Snowplow experience was associated with performance, as more experienced operators tending to perform better in the simulator; 6) Low experience operators showed a tendency to alter their fixation behaviors from the first to the second drive.

REFERENCES

Allport, G. W., & Odbert, H. S. (1936). Trait names: A psycho-lexical study. *Psychological Monographs, 47*.

Arms, L. L., & Cerney, M. M. (2005). The relationship between age and incidence of cybersickness among immersive environment users. *Proceedings of the IEEE Virtual Reality 2005,* 267-268.

Arthur, W. (1991). Individual differences in the prediction and training of complex perceptual-motor skill tasks: The development and validation of the Computer-Administered Visual Attention Test. College Station: Texas A&M University.

Arthur, W., & Graziano, W. G. (1996). The five-factor model, conscientiousness and driving accident involvement. *Journal of Personality, 64,* 593-618.

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44,* 1-27.

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

Borgatta, E. F. (1964). The structure of personality characteristics. *Behavioral Science, 12,* 8-17.

Bullough, J. D., & Rea, M. S. (2001). Forward vehicular lighting and inclement weather conditions. *PAL 2001 Symposium; Darmstadt University of Technology,* 79-89.

Campagne, A., Pebayle, T., & Muzet, A. (2005). Oculomotor changes due to road events during prolonged monotonous simulated driving. *Biological Psychology, 68,* 353-368.

Chapman, P., Underwood, G., & Roberts, K. Visual search patterns in trained and untrained novice drivers. *Transportation Research Part F, 5,* 157-167.

Charlton, S. G. (2004). Perceptual and attentional effects on drivers' speed selection at curves. *Accident Analysis and Prevention, 36,* 877-884.

Chowdury, M. A., Warren, D. L., Bissell, H., & Taori, S. (1998). Are the criteria for setting advisory speeds on curves still relevant? *ITE J. (Institution of Transportation Engineers), 68,* 32-45.

Costa, P. T., Herbst, J. H., McCrae, R. R., & Siegler, I. C. (2000). Personality at midlife: Stability, intrinsic maturation, and response to life events. *Assessment, 7,* 365-378.

Costra, P. T., & McCrae, R. R. (1985). *The NEO Personality Inventory manual.* Odessa, FL: Psychological Assessment Resources.

Costa, P. T., & McCrae, R. R. (1988). Personality in adulthood: A six-year longitudinal study of self-reports and spouse ratings on the NEO Personality Inventory. *Journal of Personality and Social Psychology, 54,* 853-863.

Crundall, D., Shenton, C., & Underwood, G. (2004). Eye movements during intentional car following. *Perception, 33,* 975-986.

Crundall, D. E., & Underwood, G. (1998). Effects of experience and processing demands on visual information acquisition in drivers. *Ergonomics, 41,* 448-458.

Crundall, D., Underwood, G., & Chapman, P. (1999). Driving experience and the functional field of view. *Perception, 28,* 1075-1087.

Crundall, D., Underwood, G., & Chapman, P. (2002). Attending to the peripheral world while driving. *Applied Cognitive Psychology*, 16, 459-475.

De Fruyt, F., & Mervielde, I. (1996). The five-factor model of personality and Holland's

RIASEC interest types. *Person. Individ. Diff., 23,* 87-103.
Dicken, C. (1963). Good impression, social desirability, and acquiescence as suppressor variables. *Education and Psychological Measurement, 23,* 699-720.
Dorn, L., & Barker, D. (2005). The effects of driver training on simulated driving performance. *Accident Analysis and Prevention*, *37*, 63-69.
Drew, G. C. (1951). Variations in reflex blink-rate during visual-motor task. *Quarterly Journal of Experimental Psychology, 3* 73-88.
Drory, A., & Shinar, D. (1982). The effects of roadway environments and fatigue on sign perception. *Journal of Safety Research, 13,* 25-32.
Duh, H. B. L., Parker, D. E., & Furness, T. A. (2004). An independent visual background reduced simulator sickness in a driving simulator. *Presence, 13,* 578-588.
Emery, C., Robin, J., Knipling, R., Finn, R., & Fleger, S. (1999). Research design: Validation of simulator technology in the training, testing and licensing of tractor-trailor drivers. FHWA Final Report FHWA-MC-99-060.
Engstrom, J., Johansson, E., & Ostlund, J. (2005). Effects of visual and cognitive load in real and simulated motorway driving. *Transportation Research Part F, 8,* 97-120.
Enriquez, M, Afonin, O., Yager, B., & Maclean, K. (2001). A pneumatic tactile alerting system for the driving environment. In Proceedings of the 2001 workshop on perceptive user interfaces, 1-7.
Fisher, J. (1992). Testing the effect of road traffic signs' informational value on driver behavior. Human Factors, 34, 231-237.
Fitts, P.M., & Posner, M.I. (1967). *Learning and skilled performance in human performance*. Belmont CA: Brock-Cole.
Forsyth, G., & Hundleby, J. D. (1987). Personality and situation as determinants of desire to drink in young adults. International Journal of the Addictions, 22, 653–669.
Godley, S. T. (1999). A driving simulator investigation of perceptual countermeasures to speeding. Unpublished doctoral disseration, Monash University.
Godley, S. T., Triggs, T. J., & Fildes, B. N. (2002). Driving simulator validation for speed research. *Accident Analysis and Prevention*, *24,* 589-600.
Godley, S. T., Triggs, T. J., & Fildes, B. N. (2004). Perceptual lane width, wide perceptual road centre markings and driving speeds. *Ergonomics, 47,* 237-256.
Goldstein, R., Bauer, L. O., & Stern, J. A. (1992). Effect of task difficulty and interstimulus interval on blink parameters. *International Journal of Psychophysiology, 13*, 111-117.
Greenberg, J., Tijerina, L., Curry, R., Artz, B., Cathey, L., Grant, P., et al. (2003). Evaluation of driver distraction using an event detection paradigm. *Journal of the Transportation Research Board*, *1843*,1-9.
Guion, R. M., & Gottier, R. F. (1965). Validity of personality measures in personnel selection. *Personnel Psychology, 18,* 135-164.
Hakel, M. D. (1974). Normative personality factors recovered from ratings of personality descriptors: The beholder’s eye. *Personnel Psychology*, *27,* 409-421.
Harms, L. (1996). Driving Performance on a Real Road and in a Driving Simulator. In Gale, A.G., Brown, I.D., Haslegrave, C.M., Moorhead, I. & Taylor, S.P. (Eds.), *Vision in Vehicles V*. Elsevier Science: Amsterdam.
Hogan, R. (1986). *Manual for the Hogan Personality Inventory.* Minneapolis: National Computer Systems.

Holland J.L. (1985*). Making vocational choices: A theory of vocational personalities and work environments*. New Jersey: prentice-Hall.

Holland, J. L. (1979). *The Self-Directed Search.* Palo Alto, CA: Consulting Psychology Press.

Horberry, T., Anderson, J., & Regan, M. A. (2006). The possible safety benefits of enhanced road markings: A driving simulator evaluation. *Transportation Research Part F, 9,* 77-87.

Hulst, M., Meijman, T., Rothengatter, T. (2001). Maintaining task set under fatigue: A study of time-on-task effects in simulated driving. *Transportation Research Part F, 4,* 103-118.

Ivancic, K. (1997). Errors as a means of promoting transfer of training. Unpublished doctoral dissertation, University of New South Wales, Sydney, Australia.

Ivancic, K., & Hesketh, B. (1995). Making the best of errors during training. *Training Research Journal, 1*, 103-125.

Ivancic, K., & Hesketh, B. (2000). Learning from errors in a driving simulation: Effects on driving skill and self-confidence. *Ergonomics, 43,* 1966-1984.

Kemeny, A., & Panerai, F. (2003) Evaluating perception in driving simulation experiments. *Trends in Cognitive Sciences, 7,* 31-37.

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

Larson, L. M., Rottinghaus, P. J., & Borgen, F. H. (2002). Meta-analyses of big six interests and big five personality factors. *Journal of Vocational Behavior*, *61*, 217-239.

LaViola, J. J. A discussion of cybersickness in virtual environments. *SIGCHI Bulletin, 32,* 47-56.

Lee, J. D., Hoffman, J. D., & Hayes, E. (2004). Collision warning design to mitigate driver distraction. *CHI* 2004, 65-72.

Leitao, J. M., Moreira, A., Santos, J. A., Sousa, A. A., & Ferreira, F. N. (1999) Evaluation of Driving Education Methods in a Driving Simulator. In J. C. Teixeira, W. Hansmann, & M. B. McGrath (Eds) *Proceedings of GVE'99*, 159–164.

McCrae, R. R., & Costa, P. T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology, 51,* 882-888.

McCrae, R. R., & Costa, P. T. (1985). Openness to experience. In R. Hogan & W. H. Jones (Eds.), *Perspectives in personality,* Greenwich, CT: JAI Press, 145-172.

McCrae, R. R., & Costa, P. T. (2003). *Personality in adulthood.* New York: The Guildford Press.

McGehee, D. V., Lee, J. D., Rizzo, M., Dawson, J., Bateman, K. (2004). Quantitative analysis of steering adaptation on a high performance fixed-base driving simulator. *Transportation Research Part F, 7,* 181-196.

Miller, T. R. (1992). Benefit-cost analysis of lane marking. *Transportation Research Record, 1334,* 38-45.

O'Hanlon, J. F., & Kelley, G. R. (1977). Controlled and automatic human information processing: II. Perceptual learning automatic attending and a general theory. *Psychological Review, 84*, 127-190.

Ones, D. S. (1993). The construct validity of integrity tests. Unpublished doctoral

dissertation, University of Iowa City, Iowa.
Ones, D. S., & Viswesvaran, C. (2001a). Integrity tests and other criterion-focused occupational personality scales (COPS) used in personnel selection. *International Journal of Selection and Assessment, 9,* 31-39.
Ones, D. S., & Viswesvaran, C. (2001b). Personality at work: Criterion-focused occupational personality scales (COPS) used in personnel selection. In B. Roberts, & R. T. Hogan (Eds.), *Applied Personality Psychology.* Washington, DC: American Psychological Association.
Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology, 78,* 679-703.
Ooms, M. (2004). Measuring presence and awareness of external events. *Le Twente Student Conference on IT, Enschede*.
Panerai, F., Droulez, J., Kelada, J. M., Kemeny, A., Balligand, E., & Favre, B. (2001). Speed and safety distance control in truck driving: Comparison of simulation and real-world environment. *DSC2001-Sophia Antpolis.*
Peli, E., Bowers, A. R., Mandel, A. J., Higgins, K., Goldstein, R., & Bobrow, L. (2005). The design of driving simulator performance evaluations for driving with vision impairments and visual aid. *TRB 2005 Annual Meeting.*
Piedmont, R. L., McCrae, R. R., Riemann, R., & Angleitner, A. (2000). On the invalidity of validity scales: Evidence from self-reports and observer ratings in volunteer samples. *Journal of Personality and Social Psychology, 78,* 582-593.
Readinger, W. O., Chatziastros, A., Cunningham, W., Bulthoff, H. H., & Cutting, J. E. (2002). Gaze-eccentricity effects on road position and steering. *Journal of Experimental Psychology: Applied, 8,* 247-258.
Redelmeier, D., & Tibshirani, R. (1997). Association between cellular-telephone calls and motor vehicle collisions. *The New England Journal of Medicine, 336,* 453-458.
Rizzo, M., Sheffield, R. A., Stierman, L., & Dawson, J. (2003). Demographic and driving performance factors in simulator adaptation syndrome. *In Proceedings of 2nd International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, The Grand Summit Resort Hotel and Conference Center, Park City, Utah,* 201-208.
Robertson, I. T., Baron, H., Gibbons, P., MacIver, R., & Nyfield, G. (2000). Concientiousness and managerial performance. *Journal of Occupational and Organizational Psychology, 73,* 171-180.
Robertson, I. T., & Smith, M. (2001). Personnel selection. *Journal of Occupational and Organizational Psychology*, *74*, 441-472.
Rolland, J. P., Parker, W. D., & Stumpf, H. (1998). A psychometric examination of the French translations of the NEO-PI-R and NEO-FFI. *Journal of Personality Assessment, 71,* 269-291.
Salgado, J. F. (1999). Personnel selection methods. In C. L. Cooper, & I. T. Roberston (Eds.), *International Review of Industrial & Organizational Psychology*. New York: Wiley.
Santos, J., Merat, N., Mouta, S., Brookhuis, K., & De Waard, D. (2005). The interaction

between driving and in-vehicle information systems: Comparison of results from laboratory, simulator and real-world studies. *Transportation Research Part F, 8,* 135-146.

Saucier, G. (1998). Replicable item-cluster subcomponents in the NEO Five-Factor Inventory. *Journal of Personality Assessment, 70,* 263-276.

Schneider, M., & Kiesler, S. (2005). Calling while driving: Effects of providing remote traffic context. *CHI: 2005*, 561-569.

Shimamura, A. P., & Squire, L. R. (1984). Paired-associate learning and priming effects in Amnesia: A neuropsychological study. *Journal of Experimental Psychology: General, 133,* 556-570.

So, R. H. Y., Lo, W. T., Ho, & A. T. K. (2001). Effects of navigation speed on motion sickness caused by an immersive virtual environment. *Human Factors, 43,* 452-461.

Sodhi, M., Reimer, B., Cohen, J. L., Vastenburg, E., Kaars, R., & Kirschenbaum, S. (2002). On-road driver eye movement tracking using head-mounted devices. In Proceedings of the Eye Tracking Research and Applications Symposium. New Orleans, LA., 61-68.

Stanney, K. M., Mourant, R. R., & Kennedy, R. S. (1998). Human factors issues in virtual environments: A review of the literature. *Presence, 7,* 327-351.

Strayer, D. L., & Drews, F. A. (2003). Simulated training improves driver efficiency: Transfer from the simulator to the real world. In D. V. McGehee, J. D. Lee, & M. Rizzo (Eds.) *Driving Assessment 2003: International Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design,* 190-193.

Strayer, D. L., Drews, F. A., & Burns, S. (2004). The development and evaluation of a high-fidelity simulator training program for snowplow operators. *Utah Department of Transportation, Report No. UT-04.17*.

Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology, 44,* 703-742.

Tornros, J. (1998). Driving behavior in a real and a simulated road tunnel-A validation study. *Accid. Anal. And Prev., 30,* 497-503.

Underwood, G., Chapman, P., Bowden, K., & Crundall, D. (2002). Visual search while driving: Skill and awareness during inspection of the scene. *Transportation Research Part F, 5,* 87-97.

Underwood, G., Chapman, P., Brocklehurst, N., Underwood, J., & Crundall, D. (2003). Visual attention while driving: sequences of eye fixations made by experienced and novice drivers. *Ergonomics, 46,* 629-646.

Underwood, G., Chapman, P., Berger, Z, & Crundall, D. (2003). Driving experience, attentional focusing, and the recall of recently inspected events. *Transportation Research Part F, 6,* 289-304.

Underwood, G, Crundall, D., & Chapman, P. (2002). Selective searching while driving: The role of experience in hazard detection and general surveillance. *Ergonomics, 45,* 1-12.

Usoh, M., Catena, E., Arman, S., & Slater, M. (2000). Using presence questionnaires in reality. *Presence, 9,* 497-503.

Veltman, J. A., & Gaillard, A. W. K. (1996). Physiological indices of workload in a simulated flight task. *Biological Psychology, 42*, 323-342.

Victor, T. W., Harbluk, J. L., Engstrom, J. A. (2005). Sensitivity of eye-movement measures to in-vehicle task difficulty. *Transportation Research Part F, 8*, 167-190.

Wierwille, W. (1993). Visual and manual demands of in-car controls and displays. In B. Peacock, & W. Karwowski, (Eds.) *Automotive Ergonomics*, 299-320.

Wilson, G. R. (1993). Air-to-ground training missions: A psychological workload analysis. *Ergonomics*, *36*, 1071-1087.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence, 7*, 225-240.

Zuckerman, M. (1979). *Sensation seeking: Beyond the optimal level of arousal.* Hillsdale, NJ: Erlbaum.

Zuckerman, M. (1994). *Behavioral expression and biosocial bases of sensation seeking*. New York: Cambridge University Press.

Zuckerman, M., Eysenck, S., & Eysenck, H. J. (1978). Sensation seeking in England and America: Cross-cultural, age and sex comparisons. *Journal of Consulting and Clinical Psychology*, *46,* 139-149.

APPENDIX A. LECTURE POWER POINT

MPRI

101

Circles of Influence

MPRI

Learning Objectives

- List the Circles of Influence that affect driving performance and decision-making.
- Define the steps in the SIPDE Method to hazard perception and apply each to a driving situation.
- Demonstrate the ability to make successful decisions in simulation using applicable aspects of the Circles of Influence methodology.

2

MPRI

What Is the Leading Cause of Collisions?

- Poor vehicle-handling skills
- Poor decision-making skills

80%

3

MPRI

The Decision-Making Process

- Preparation
- Practice

4

MPRI

Driver Safety

- Most driver-safety programs target the 15% of drivers who get in collisions.
- Management strategies should be designed to help the 85% of drivers who drive safely.

5

MPRI

The Circles of Influence

- You
- Health and family
- Job and equipment knowledge
- Environment
- Hazard perception
- Successful decision-making

6

MPRI

You

- Age, gender, personality, and region you're from
- Driving experience and familiarity with equipment
- Traffic violations and collisions

7

MPRI

Health and Family

- Breaks, rest, relaxation, and exercise
- Family crises and unresolved issues
- Fatigue
- Medications

8

MPRI

Job and Equipment Knowledge

- DOT policies
- Pre-trip inspections
- Vehicle and equipment
- Special gear or paperwork

9

MPRI

Environmental Conditions

- Weather
- Traffic
- Road conditions and structures

10

Hazard Perception—The SIPDE Method

- **S**can and **S**earch
- **I**dentify
- **P**redict
- **D**ecide
- **E**xecute

The SIPDE Method

***S**can and Search:*

- Search for hazards
- Avoid tunnel vision
- Be alert to sounds, smells, and vibrations

MPRI

The SIPDE Method

Identify:

- Interpret sensory input
- Sort information
- Identify hazards
- Weigh dangers

13

MPRI

The SIPDE Method

Predict:

- Imagine result of potential threat
- Decide if you can change the result
- Anticipate what other drivers might do

14

MPRI

The SIPDE Method

Decide:

- Critical moments often require fast decisions
- Practice decision-making by imagining conflicts and outcomes

15

MPRI

The SIPDE Method

Execute:

- Quickly follow decisions with action
- Maintain strong vehicle-handling skills

16

MPRI

Benefits of Successful Decision-Making

- Job satisfaction and self-esteem
- Confidence and trust
- More good decisions
- Wise driving habits

MPRI

Things to think about !

MPRI

Maintaining Focus

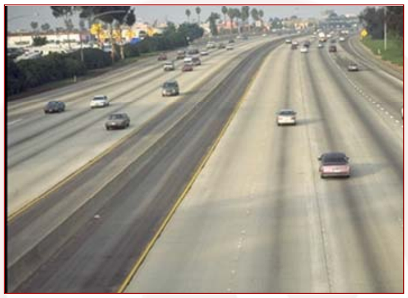
- Talk out loud.
- Analyze what other drivers are doing.

19

MPRI

Driving Consistently

- Hold your lane, speed, and position.
- Your consistency helps others make safe decisions.



20

MPRI

On to the Simulator!

Advanced Simulation Exercises

TTSCN_04

MPRI

402

Space Management

MPRI

Learning Objectives

- Calculate a safe following distance under a variety of circumstances by applying the following-distance formula.
- List appropriate space-management techniques used to manage the space around a vehicle when encountering a variety of hazards.
- Demonstrate ability to manage space cushion around a vehicle when encountering typical driving hazards in simulation.

23

MPRI

CBT Lab
Space Management Part 1

24

MPRI

5 Minute Break

25

MPRI

On to the CBT Lab!
Space Management Part 2

26

MPRI

SIMULATOR

First Driver: TTR 402C

Second Driver: TTSCN_8

27

MPRI

SUMMARY

28

MPRI

Key Points: Circles of Influence

- You
- Health and family
- Knowledge
- Environment
- Hazard perception
- Successful decision-making

29

MPRI

Key Points: SIPDE Method ?

- **S**can and **S**earch
- **I**dentify
- **P**redict
- **D**ecide
- **E**xecute

30

MPRI

Key Points: Space, Stopping Time

- The space cushion: space ahead, behind, above, below, and to the sides.
- The timed-interval method indicates the following distance in seconds.
- Stopping time includes time to see, think, react, and brake.
- Calculate stopping distance according to vehicle length. Adjust for adverse conditions.

32

MPRI

Key Points: Best Practices

- Drivers develop their own best practices through experience.
- Drivers benefit from one another's experiences when they share best practices.

31

L3 MPRI

Key Points: Maneuverability, Right-of-Way

- Maneuverability depends on vehicle size, weight, load, speed, center of gravity, and condition.
- Managing space during turns, curves, passes, merges, and lane changes is critical to safe driving.
- Respect laws for yielding right-of-way.

APPENDIX B. NEO-FFI QUESTIONNAIRE

Please rate the extent to which you agree with the following statements.
Use the scale below for your responses.

Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Agree
1	2	3	4	5

______ 1. I am not a worrier.
______ 2. I like to have a lot of people around me.
______ 3. I don't like to waste my time daydreaming.
______ 4. I try to be courteous to everyone I meet.
______ 5. I keep my belongings neat and clean.
______ 6. I often feel inferior to others.
______ 7. I laugh easily.
______ 8. Once I find the right way to do something, I stick to it.
______ 9. I often get into arguments with my family and co-workers.
______ 10. I'm pretty good about pacing myself so as to get things done on time.
______ 11. When I'm under a great deal of stress, sometimes I feel like I'm going to pieces.
______ 12. I don't consider myself especially "light-hearted."
______ 13. I am intrigued by the patterns I find in art and nature.
______ 14. Some people think I'm selfish and egotistical.
______ 15. I am not a very methodical person.
______ 16. In dealing with other people, I always dread making a social blunder.
______ 17. I really enjoy talking to people.
______ 18. I believe letting students hear controversial speakers can only confuse and mislead them.
______ 19. I would rather cooperate with others than compete with them.
______ 20. I try to perform all the tasks assigned to me conscientiously.
______ 21. I often feel tense and jittery.
______ 22. I like to be where the action is.
______ 23. Poetry has little or no effect on me.
______ 24. I tend to be cynical and skeptical of others' intentions.
______ 25. I have a clear set of goals and work toward them in an orderly fashion.
______ 26. Sometimes I feel completely worthless.
______ 27. I usually prefer to do things alone.
______ 28. I often try new and foreign foods.
______ 29. I believe that most people will take advantage of you if you let them.
______ 30. I waste a lot of time before settling down to work.
______ 31. I rarely feel fearful or anxious.

______ 32. I often feel as if I'm bursting with energy.
______ 33. I seldom notice the moods or feelings that different environments produce.
______ 34. Most people I know like me.
______ 35. I work hard to accomplish my goals.
______ 36. I often get angry at the way people treat me.
______ 37. I am a cheerful, high-spirited person.
______ 38. I believe we should look to our religious authorities for decisions on moral issues.
______ 39. Some people think of me as cold and calculating.
______ 40. When I make a commitment, I can always be counted on to follow through.
______ 41. Too often, when things go wrong, I get discouraged and feel like giving up.
______ 42. I am not a cheerful optimist.
______ 43. Sometimes when I am reading poetry or looking at a work of art, I feel a chill or wave of excitement.
______ 44. I'm hardheaded and tough-minded in my attitudes.
______ 45. Sometimes I'm not as dependable or reliable as I should be.
______ 46. I am seldom sad or depressed.
______ 47. My life is fast-paced.
______ 48. I have little interest in speculating on the nature of the universe or the human condition.
______ 49. I generally try to be thoughtful and considerate.
______ 50. I am a productive person who always gets the job done.
______ 51. I often feel helpless and want someone else to solve my problems.
______ 52. I am a very active person.
______ 53. I have a lot of intellectual curiosity.
______ 54. If I don't like people, I let them know it.
______ 55. I never seem to be able to get organized.
______ 56. At times I have been so ashamed I just wanted to hide.
______ 57. I would rather go my own way than be a leader of others.
______ 58. I often enjoy playing with theories or abstract ideas.
______ 59. If necessary, I am willing to manipulate people to get what I want.
______ 60. I strive for excellence in everything I do.

APPENDIX C. SENSATION SEEKING QUESTIONNAIRE

Interest and Preference Test

Each of the items below contains two choices, A and B. Please indicate which of the choices most describes your likes or the way you feel. In some cases you may find items in which both choices describe your likes or feelings. Please choose the one which better describes your likes or feelings. In some cases you may find items in which you do not like either choice. In these cases mark the choice you dislike least. Do not leave any items blank. It is important you respond to all items with only one choice, A or B. We are interested only in your likes or feelings, not in how others feel about these things or how one is supposed to feel. There are no right or wrong answers as in other kinds of tests. Be frank and give your honest appraisal of yourself.

1. A. I like uninhibited parties.
   B. I prefer quiet parties with good conversation.

2. A. There are some movies I enjoy seeing a second or even a third time.
   B. I can't stand watching a movie that I've seen before.

3. A. I often wish I could be a mountain climber.
   B. I can't understand people who risk their necks climbing mountains.

4. A. I dislike all body odors.
   B. I like some of the earthly body smells.

5. A. I get bored seeing the same old faces.
   B. I like the comfortable familiarity of everyday friends.

6. A. I like to explore a strange city or section of town by myself, even if it means getting lost.
   B. I prefer a guide when I am in a place I don't know well.

7. A. I dislike people who do or say things just to shock or upset people.
   B. When you can predict almost everything a person will do and say he or she must be a bore.

8. A. I usually don't enjoy a movie or play where I can predict what will happen in advance.
   B. I don't mind watching a movie or play where I can predict what will happen in advance.

9 A. A sensible person avoids activities that are dangerous.
   B. I sometimes like to do things that are a little frightening.

10. A. I like to try new foods that I have never tasted before.
B. I order the dishes with which I am familiar so as to avoid disappointment and unpleasantness.

11. A. I enjoy looking at home movies, videos, or travel slides.
B. Looking at someone's home movies, videos or travel slides bores me tremendously.

12. A. I would like to take up the sport of water skiing.
B. I would not like to take up water skiing.

13. A. I would like to try surfing.
B. I would not like to try surfing.

14. A. I would like to take off on a trip with no preplanned or definite routes, or timetable.
B. When I go on a trip I like to plan my route and timetable fairly carefully.

15. A. I prefer the "down to earth" kinds of people as friends.
B. I would like to make unusual friends.

16. A. I would not like to learn to fly an airplane.
B. I would like to learn to fly an airplane.

17. A. I prefer the surface of the water to the depths.
B. I would like to go scuba diving.

18. A. I would like to try parachute jumping.
B. I would never want to try jumping out of a plane, with or without a parachute.

19. A. I prefer friends who are excitingly unpredictable.
B. I prefer friends who are reliable and predictable.

20. A. I am not interested in experience for its own sake.
B. I like to have new and exciting experiences and sensations even if they are a little frightening or unconventional.

21. A. The essence of good art is in its clarity, symmetry of form, and harmony of colors.
B. I often find beauty in the "clashing" colors and irregular forms of modern painting.

22. A. I enjoy spending time in the familiar surroundings of home.
B. I get very restless if I have to stay around home for any length of time.

23. A. I like to dive off the high board.
B. I don't like the feeling I get standing on the high board (or I don't go near it at all).

24. A. I like to date persons who are physically exciting.
B. I like to date persons who share my values.

25. A. The worst social sin is to be rude.
B. The worst social sin is to be a bore.

26. A. I like people who are sharp and witty even if they do sometimes insult each other.
B. I dislike people who have their fun at the expense of hurting the feelings of others.

27. A. People should dress according to some standard of taste, neatness and style.
B. People should dress in individual ways even if the effects are sometimes strange.

28. A. Sailing long distances in small sailing crafts is foolhardy.
B. I would like to sail a long distance in a small but seaworthy sailing craft.

29. A. I have no patience with dull or boring people.
B. I find something interesting in almost every person I talk to.

30. A. Skiing down a high mountain slope is a good way to end up on crutches.
B. I think I would enjoy the sensations of skiing very fast down a high mountain slope.

APPENDIX D: IMMERSION QUESTIONNAIRE

Pre experiment questionnaire
Please read the questions carefully and circle the appropriate number from 1 to 7 .

1. Do you ever get extremely involved in projects that are assigned to you by your boss or your instructor, to the exclusion of other tasks?

1	2	3	4	5	6	7
NOT OFTEN			MODERATELY OFTEN			VERY OFTEN

2. How easily can you switch your attention from the task in which you are currently involved to a new task?

1	2	3	4	5	6	7
NOT EASILY			MODERATELY EASILY			VERY EASILY

3. How frequently do you get emotionally involved (angry, sad, or happy) in the news stories that you read or hear?

1	2	3	4	5	6	7
NOT FREQUENTLY			MODERATELY FREQUENTLY			VERY FREQUENTLY

4. How well do you feel today?

1	2	3	4	5	6	7
NOT WELL			MODERATELY WELL			VERY WELL

5. Do you easily become deeply involved in movies or TV dramas?

1	2	3	4	5	6	7
NOT EASILY			MODERATELY EASILY			VERY EASILY

6. Do you ever become so involved in a television program or book that people have problems getting your attention?

1	2	3	4	5	6	7
NOT OFTEN			MODERATELY OFTEN			VERY OFTEN

7. How mentally alert do you feel at the present time?

1	2	3	4	5	6	7
NOT ALERT			MODERATELY ALERT			VERY ALERT

8. Do you ever become so involved in a movie that you are not aware of things happening around you?

1	2	3	4	5	6	7
NOT OFTEN			MODERATELY OFTEN			VERY OFTEN

9. How frequently do you find yourself closely identifying with the characters in a story line?

1	2	3	4	5	6	7
NOT OFTEN			MODERATELY OFTEN			VERY OFTEN

10. Do you ever become so involved in a video game that it is as if you are inside the game rather than moving a joystick and watching the screen?

1	2	3	4	5	6	7
NOT OFTEN			MODERATELY OFTEN			VERY OFTEN

11. On average, how many books do you read for enjoyment in a month?

12. What kind of books do you read most frequently?

(CIRCLE ONE ITEM ONLY!)

Spy novels   Fantasies   Science fiction   Adventure
Romance novels   Historical novels   Westerns   Mysteries   Other fiction
Biographies   Autobiographies   Other non-fiction

13. How physically fit do you feel today?

1	2	3	4	5	6	7
NOT FIT			MODERATELY FIT			VERY FIT

14. How good are you at blocking out external distractions when you are involved in something?

1	2	3	4	5	6	7
NOT GOOD			MODERATELY			VERY GOOD

15. When watching sports, do you ever become so involved in the game that you react as if you were one of the players?

1	2	3	4	5	6	7
NOT OFTEN			MODERATELY OFTEN			VERY OFTEN

16. Do you ever become so involved in a daydream that you are not aware of things happening around you?

1	2	3	4	5	6	7
NOT OFTEN			MODERATELY OFTEN			VERY OFTEN

17. Do you ever have dreams that are so real that you feel disoriented when you awake?

1	2	3	4	5	6	7
NOT OFTEN			MODERATELY OFTEN			VERY OFTEN

18. When playing sports, do you become so involved in the game that you lose track of time?

1	2	3	4	5	6	7
NOT OFTEN			MODERATELY OFTEN			VERY OFTEN

19. Are you easily disturbed when working on a task?

1	2	3	4	5	6	7
NOT EASILY			MODERATELY EASILY			VERY EASILY

20. How well do you concentrate on enjoyable activities?

1	2	3	4	5	6	7
NOT WELL			MODERATELY WELL			VERY WELL

21. How often do you play arcade or video games? (OFTEN should be taken to mean every day or every two days, on average.)

1	2	3	4	5	6	7
NOT OFTEN			MODERATELY OFTEN			VERY OFTEN

22. How well do you concentrate on disagreeable tasks?

1	2	3	4	5	6	7
NOT WELL			MODERATELY WELL			VERY WELL

23. Have you ever gotten excited during a chase or fight scene on TV or in the movies?

1	2	3	4	5	6	7
NOT OFTEN			MODERATELY OFTEN			VERY OFTEN

24. To what extent have you dwelled on personal problems in the last 48 hours?

1	2	3	4	5	6	7
NOT MUCH			MODERATELY			VERY MUCH

25. Have you ever gotten scared by something happening on a TV show or in a movie?

1	2	3	4	5	6	7
NOT OFTEN			MODERATELY OFTEN			VERY OFTEN

26. Have you ever remained apprehensive or fearful long after watching a scary movie?

1	2	3	4	5	6	7
NOT OFTEN			MODERATELY OFTEN			VERY OFTEN

27. Do you ever avoid carnival or fairground rides because they are too scary?

1	2	3	4	5	6	7
NOT OFTEN			MODERATELY OFTEN			VERY OFTEN

28. How frequently do you watch TV soap operas or docu-dramas?

1	2	3	4	5	6	7
NOT OFTEN			MODERATELY OFTEN			VERY OFTEN

29. Do you ever become so involved in doing something that you lose all track of time?

1	2	3	4	5	6	7
NOT OFTEN			MODERATELY OFTEN			VERY OFTEN

APPENDIX E. PRESENCE QUESTIONNAIRE

Post experiment questionnaire
Please read the questions carefully and circle the appropriate number from 1 to 7 .

1. How much were you able to control events?

1	2	3	4	5	6	7
NOT WELL			MODERATELY WELL			VERY WELL

2. How responsive was the environment to actions that you initiated (or performed)?

1	2	3	4	5	6	7
NOT RESPONSIVE			MODERATELY RESPONSIVE			VERY RESPONSIVE

3. How natural did your interactions with the environment seem?

1	2	3	4	5	6	7
NOT NATURAL			MODERATELY NATURAL			VERY NATURAL

4. How completely were all of your senses engaged?

1	2	3	4	5	6	7
NOT ENGAGED			MODERATELY ENGAGED			VERY ENGAGED

5. How much did the visual aspects of the environment involve you?

1	2	3	4	5	6	7
NOT MUCH			MODERATELY			VERY MUCH

6. How much did the auditory aspects of the environment involve you?

1	2	3	4	5	6	7
NOT MUCH			MODERATELY			VERY MUCH

7. How natural was the mechanism which controlled movement through the environment?

1	2	3	4	5	6	7
NOT NATURAL			MODERATELY NATURAL			VERY NATURAL

8. How aware were you of events occurring in the real world around you?

1	2	3	4	5	6	7
NOT AWARE			MODERATELY AWARE			VERY AWARE

9. How aware were you of your display and control devices?

1	2	3	4	5	6	7
NOT AWARE			MODERATELY AWARE			VERY AWARE

10. How compelling was your sense of objects moving through space?

1	2	3	4	5	6	7
NOT COMPELLING			MODERATELY COMPELLING			VERY COMPELLING

11. How inconsistent or disconnected was the information coming from your various senses?

1	2	3	4	5	6	7
NOT INCONSISTENT			MODERATELY INCONSISTENT			VERY INCONSISTENT

12. How much did your experiences in the virtual environment seem consistent with your real- world experiences?

1	2	3	4	5	6	7
NOT CONSISTENT			MODERATELY CONSISTENT			VERY CONSISTENT

13. Were you able to anticipate what would happen next in response to the actions you performed?

1	2	3	4	5	6	7
NOT WELL			MODERATELY WELL			VERY WELL

14. How completely were you able to actively survey or search the environment using vision?

1	2	3	4	5	6	7
NOT COMPLETELY			MODERATELY			COMPLETELY

15. How well could you identify sounds?

1	2	3	4	5	6	7
NOT WELL			MODERATELY WELL			VERY WELL

16. How well could you localize sounds?

1	2	3	4	5	6	7
NOT WELL			MODERATELY WELL			VERY WELL

17. How well could you actively survey or search the environment using touch?

1	2	3	4	5	6	7
NOT WELL			MODERATELY WELL			VERY WELL

18. How compelling was your sense of moving around inside the virtual environment?

1	2	3	4	5	6	7
NOT COMPELLING			MODERATELY COMPELLING			VERY COMPELLING

19. How closely were you able to examine objects?

1	2	3	4	5	6	7
NOT CLOSELY			MODERATELY CLOSELY			VERY CLOSELY

20. How well could you examine objects from multiple viewpoints?

1	2	3	4	5	6	7
NOT WELL			MODERATELY WELL			VERY WELL

21. How well could you move or manipulate objects in the virtual environment?

1	2	3	4	5	6	7
NOT WELL			MODERATELY WELL			VERY WELL

22. To what degree did you feel confused or disoriented at the beginning of breaks or at the end of the experimental session?

1	2	3	4	5	6	7
NOT DISORIENTED			MODERATELY DISORIENTED			VERY DISORIENTED

23. How involved were you in the virtual environment experience?

1	2	3	4	5	6	7
NOT INVOLVED			MODERATELY INVOLVED			VERY INVOLVED

24. How distracting was the control mechanism?

1	2	3	4	5	6	7
NOT DISTRACTING			MODERATELY DISTRACTING			VERY DISTRACTING

25. How much delay did you experience between your actions and expected outcomes?

1	2	3	4	5	6	7
NOT MUCH			MODERATELY			VERY MUCH

26. How quickly did you adjust to the virtual environment experience?

1	2	3	4	5	6	7
NOT QUICKLY			MODERATELY QUICKLY			VERY QUICKLY

27. How proficient in moving and interacting with the virtual environment did you feel at the end of the experience?

1	2	3	4	5	6	7
NOT PROFICIENT			MODERATELY PROFICIENT			VERY PROFICIENT

28. How much did the the visual display quality interfere or distract you from performing assigned tasks or required activities?

1	2	3	4	5	6	7
NOT MUCH			MODERATELY			VERY MUCH

29. How much did the control devices interfere with performance of assigned tasks or other activities?

1	2	3	4	5	6	7
NOT MUCH			MODERATELY			VERY MUCH

30. How well could you concentrate on the assigned tasks or required activities rather than on the mechanisms used to perform those tasks or activities?

1	2	3	4	5	6	7
NOT WELL			MODERATELY WELL			VERY WELL

31. Did you learn new techniques that enabled you to improve your performance?

1	2	3	4	5	6	7
NOT MUCH			MODERATELY			VERY MUCH

32. Were you involved in the experimental task to the extent that you lost track of time?

1	2	3	4	5	6	7
NOT MUCH			MODERATELY			VERY MUCH

APPENDIX F. SIMULATOR SICKNESS QUESTIONNAIRE

Simulator Sickness Questionnaire

Please report the degree to which you experience each of the below symptoms as one of ``None", ``Slight", ``Moderate" and ``Severe". Using the scale from "0" (none) to "3" (severe).

	None	Slight	Moderate	Severe
General discomfort	0	1	2	3
Fatigue	0	1	2	3
Headache	0	1	2	3
Eyestrain	0	1	2	3
Difficulty focusing	0	1	2	3
Increased salivation	0	1	2	3
Sweating	0	1	2	3
Nausea	0	1	2	3
Difficulty concentrating	0	1	2	3
Fullness of head	0	1	2	3
Blurred vision	0	1	2	3
Dizzy (eyes open)	0	1	2	3
Dizzy (eyes closed)	0	1	2	3
Vertigo	0	1	2	3
Stomach awareness	0	1	2	3
Burping	0	1	2	3

Scoring in this Report: Participants report 0, 1, 2 or 3 for each of these above symptoms, corresponding to rating of "none," "slight," "moderate," and "severe." To calculate the total simulator sickness score, multiply each rating by 3.74, and then sum up all 16 values for each participant.

APPENDIX G. TRAINING AND REALISM QUESTIONNAIRE

Please indicate how much you agree or disagree with the following statements.

1) The snowplow training package was very useful

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

2) The classroom/lecture portion of the training was very useful.

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

3) The simulator training was very useful.

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

4) This training should be part of IDOT training for all snowplow operators.

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

5) The training helped prepare me for dealing with non-routine situations.

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

6) The training helped prepare me for situations involving passing cars.

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

7) The training helped prepare me for situations involving vehicles or pedestrians along the side of the road.

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

8) This training explained why speed management is important for safe plowing.

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

9) This training explained why space management is important for safe plowing.

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

10) This training explained why good communication is important for safe plowing.

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

11) I would recommend this training for other snowplow drivers.

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

12) The course objectives satisfied my needs.

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

13) The driving simulations were realistic for the course objectives.

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

14) I practiced skills during the driving simulation part of the course that will be very useful on the road.

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

15) The skills I practiced during the SIPDE part of the course will be very useful on the road.

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

16) The time spent in the lecture portion of the course was appropriate.

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

17) The time spent in the lecture portion of the course was appropriate.

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

18) The time spent in the driving simulation portion of the course was appropriate.

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

19) The trainer had a good understanding of the course material.

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

20) The trainer worked will with the drivers.

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

21) The trainer understood your needs and issues.

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

22) The trainer gave very useful feedback.

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

23) The scenery (trees, sky roads) in the simulator looks realistic.

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

24) The vehicles (cars, trucks) in the scenarios look realistic.

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

25) The mirrors were where they would be in a real snowplow.

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

26) The mirrors worked as well as mirrors in a real snowplow.

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

27) The behavior of other vehicles in the scenarios was realistic.

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

28) The controls (steering wheel, pedals) in the simulator were where they would be in a real snowplow.

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

29) The controls (steering wheel, pedals) in the simulator were as responsive as normal snowplow controls.

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

30) I felt like I was driving a real snowplow while operating the simulator.

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

ACKNOWLEDGEMENTS

Many people deserve thanks and recognition for their aid in this project. First, I would like to thank Derrick Parkhurst, Veronica Dark and William Robinson for helping with data analysis, interpretation and editing various portions and drafts of this manuscript. I would also like to acknowledge the help of Jeremiah Still and Holly Brown with data collection. Finally, I would like to thank the Iowa Department of Transportation for the use their equipment, buildings and operators to conduct this project.