Evaluation of Virtual Reality Snowplow Simulator Training: A Literature Review

Final Report
September 2006

Sponsored by
the Iowa Department of Transportation
(CTRE Project 06-245)

ctre
Center for Transportation Research and Education

IOWA STATE UNIVERSITY

Iowa State University's Center for Transportation Research and Education is the umbrella organization for the following centers and programs: Bridge Engineering Center • Center for Weather Impacts on Mobility and Safety • Construction Management & Technology • Iowa Local Technical Assistance Program • Iowa Traffic Safety Data Service • Midwest Transportation Consortium • National Concrete Pavement Technology Center • Partnership for Geotechnical Advancement • Roadway Infrastructure Management and Operations Systems • Statewide Urban Design and Specifications • Traffic Safety and Operations

About CTRE/ISU

The mission of the Center for Transportation Research and Education (CTRE) at Iowa State University is to develop and implement innovative methods, materials, and technologies for improving transportation efficiency, safety, and reliability while improving the learning environment of students, faculty, and staff in transportation-related fields.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.	
CTRE Project 06-245			
4. Title and Subtitle Evaluation of Virtual Reality Snowplow Simulator Training: A Literature Review		**5. Report Date** September 2006	
		6. Performing Organization Code	
7. Author(s) Christopher Masciocchi, Veronica Dark, Derrick Parkhurst		**8. Performing Organization Report No.**	
9. Performing Organization Name and Address Center for Transportation Research and Education Iowa State University 2711 South Loop Drive, Suite 4700 Ames, IA 50010-8664		**10. Work Unit No. (TRAIS)**	
		11. Contract or Grant No.	
12. Sponsoring Organization Name and Address Iowa Department of Transportation 800 Lincoln Way Ames, IA 50010		**13. Type of Report and Period Covered** Final Report	
		14. Sponsoring Agency Code	
15. Supplementary Notes Visit www.ctre.iastate.edu for color PDF files of this and other research reports.			

16. Abstract

Each winter, the Iowa Department of Transportation (Iowa DOT) maintenance operators are primarily responsible for plowing snow off federal and state roads. 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 sprayers. However, operators have few opportunities during the year to practice and refine their skills. An ideal training program would provide operators with the opportunity to practice these skills under realistic yet safe conditions, as well as provide basic training to novice or less-experienced operators.

Recent technological advancements have made driving simulators a desirable training and research tool. This literature review discusses much of the recent research establishing simulator fidelity and espousing its applicability. Additionally, this report provides a summary of behavioral and eye tracking research involving driving simulators. Other research topics include comparisons between novice and expert drivers' behavioral patterns, methods for avoiding cybersickness in virtual environments, and a synopsis of current personality measures with respect to job performance and driving performance.

This literature review coincides with a study designed to examine the effectiveness of virtual reality snowplow simulator training for current maintenance operators, using the TranSim VS III truck and snowplow simulator recently purchased by the Iowa DOT.

17. Key Words		18. Distribution Statement	
driving simulators—eye movement—job performance—snowplow training—virtual reality		No restrictions.	
19. Security Classification (of this report)	**20. Security Classification (of this page)**	**21. No. of Pages**	**22. Price**
Unclassified.	Unclassified.	52	NA

Form DOT F 1700.7 (8-72) **Reproduction of completed page authorized**

EVALUATION OF VIRTUAL REALITY SNOWPLOW SIMULATOR TRAINING: A LITERATURE REVIEW

Final Report
September 2006

Principal Investigator
Derrick Parkhurst
Associate Director, Virtual Reality Application Center
Assistant Professor, Department of Psychology
Iowa State University

Co-Principal Investigator
Veronica Dark
Associate Professor, Department of Psychology
Iowa State University

Research Assistant
Christopher Masciocchi

Authors
Christopher Masciocchi, Veronica Dark, Derrick Parkhurst

Preparation of this report was financed in part
through funds provided by the Iowa Department of Transportation
through its research management agreement with the
Center for Transportation Research and Education,
CTRE Project 06-245.

A report from
Center for Transportation Research and Education
Iowa State University
2711 South Loop Drive, Suite 4700
Ames, IA 50010-8664
Phone: 515-294-8103
Fax: 515-294-0467
www.ctre.iastate.edu

TABLE OF CONTENTS

ACKNOWLEDGMENTS IX

INTRODUCTION 1

DRIVING SIMULATORS 2

- Simulator Fidelity 2
- Applications and Uses of Driving Simulators 7
- Driving Simulators and Training 12
- Eye and Head Movements 15
- Novice Versus Expert Drivers 17
- Utah DOT Snowplow Study 21
- Cybersickness 22

PERSONALITY 26

- The Big Five 26
- The Big Five and Job Performance 28
- The Big Five and Driving Performance 30
- Other Methods of Predicting Job Performance 31

OTHER MEASUREMENTS RELAVENT TO SIMULATOR TRAINING 34

- Sensation Seeking Tendencies 34
- Immersion and Presence 34

CONCLUSION 36

REFERENCES 37

LIST OF TABLES

Table 1. Normative means and standard deviations for the NEO-FFI 27
Table 2. Moderate correlations between Big Five and Big Six personality dimensions...............29

ACKNOWLEDGMENTS

The authors would like to thank the Iowa Department of Transportation for sponsoring this research.

INTRODUCTION

Each winter, Iowa Department of Transportation (Iowa DOT) maintenance operators are primarily responsible for plowing snow off federal and state roads. 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 sprayers. However, operators have few opportunities during the year to practice and refine their skills. An ideal training program would provide experienced operators with a chance to practice these skills under realistic yet safe conditions, as well as supply basic training to novice or less-experienced operators. Virtual reality training is often utilized in circumstances when real-world training would be prohibitively high-priced, inappropriate, or hazardous. For instance, flight simulators are basic requirements of all pilot training programs, as they provide a safe yet realistic environment where students are taught how to operate aircrafts.

This literature review coincides with a study designed to examine the effectiveness of virtual reality snowplow simulator training for current maintenance operators, using the TranSim VS III truck and snowplow simulator recently purchased by the Iowa DOT. Similar to other applications which rely on virtual reality training, real-world snowplow training is often difficult to coordinate. 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 unfortunately 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 mitigates many of these problems. 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. Therefore, virtual reality simulator training, if it can be shown to be effective, would be highly beneficial for snowplow operators.

DRIVING SIMULATORS

Driving simulators have recently become a desirable training and research tool due to technological advances in display quality and research espousing their applicability. Most importantly, driving simulators have been shown to be high in relative behavioral validity, as individuals tend to operate driving simulators in virtual environments similar to real vehicles. Providing both a wide field of view and the ability to manipulate simulator performance are important considerations for simulator fidelity. Due to advances in simulator fidelity, driving simulators are often used in lieu of real vehicles in basic and applied settings, especially when replicating the study or training 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.

Also, because driving simulators provide a realistic yet non-hazardous driving experience, several studies have examined trainees' eye movements as they navigate these environments, without the risk of the equipment or task inadvertently causing a serious accident. Eye movements are generally measured for two reasons: (1) to find the typical adaptive scan pattern employed by experienced drivers during normal driving conditions and (2) to determine 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 paper summarizes current literature on driving simulator research and training practices, discusses possible causes and preventions of cybersickness, and it 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 and 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 important 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 computers, as simulators provide encompassing visual feedback in an immersive environment. However, the authors state 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. However, Kemeny and Panerai conclude that simulators with a large field of view, greater than 120 degrees, provide participants with enough information that they can estimate speed relatively accurately. Incidentally, the TranSim VS III offers a field of view of 180 degrees.

Kinesthetic (body movement) and haptic (touch) cues are equally important in driving simulators but are only found in motion-based simulators. Kemeny and Panerai (2003) state that a lack of kinesthetic cues has been shown to increase reaction time to external disturbances (e.g., simulated strong winds) and 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. Proper simulator calibration not only improves realism, but also may increase drivers' feelings of presence within the virtual environment. In addition to realistic visuals, the equipment and peripherals in the simulator—the pedals, steering wheel, and mirrors—must imitate those of a real snowplow to facilitate the transfer of skills to real driving. While it is certainly not necessary that these features perfectly mimic the driver's actual 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, as would be demonstrated by observing driver behavior in the virtual environment similar to driver behavior in a real vehicle. 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 automobile (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 identical. 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, some researchers (Tornros 1998) suggest 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 (MUARC) 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.

Moreover, 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, Triggs, and Fildes (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 in the simulator than in road trials. In other words, 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 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, therefore, was not achieved 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 simulator and real world performance. Overall, there was a 0.85 correlation of 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 almost identical in the simulator and real vehicle conditions. The results approximate those necessary for absolute validity of speed for the driving simulator.

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 following distance validity. They attribute this null 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 subjects to have maximum control over their environment. Twenty subjects 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 they were driving on the left and it was on their left side. Additionally, curvature was believed to affect lateral position, as past research suggested that people have a tendency to follow curves on the inner side (Harms 1994).

The data showed that speed was higher overall in the simulator, which suggests a lack of absolute validity, as participant 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, there was also no interaction between driving in the simulated and real tunnels. Overall, this suggests that 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 suggests 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 in driving simulators.

Another method of assessing validity is to have participants perform a secondary task and compare its effects on various conditions. Presumably, if the secondary task affects performance in the different conditions similarly, the same skills are being employed across conditions. Santos et al. (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. It 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 field conditions reported consistent decreases in subjective performance as secondary task difficulty increased, while laboratory self-report 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 field 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-world 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 identical measurements 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 measurements of performance in the two conditions are not exact.

Researchers have concluded the following information about simulator fidelity:

- A wide field of view provides a sense of optic flow that is important for visual simulator fidelity.
- There are two types of validity in driving simulator research: physical validity and behavioral validity.
- Driving simulators that can be manipulated to match the performance of real vehicles are high in physical validity.
- Driving simulators are high in relative behavioral validity for some performance criteria; people tend to perform similarly in driving simulators (regarding factors such as speed and lane position) as they do in real world driving conditions across a variety of tasks.

Applications and Uses of Driving Simulators

Driving simulators in particular are adept at 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 and Kiesler 2005), conditions that lead to better in vehicle performance (Bullough and Rea 2001), and devices to help mitigate accidents with in-vehicle warning systems (Enriquez et al. 2001; Lee, Hoffman, and Hayes 2004). Researchers have also utilized driving simulators to investigate fundamental processes of human attention and visual perception while driving (Readinger et al. 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 are akin to those of holding a conversation with a passenger. Talking 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 suggest that subjects' 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 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 the 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, Harbluk, and Engstrom 2005) suggest that participants had different eye movement patterns while performing the auditory task than 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 subjects demonstrated reduced scanning practices immediately after completing the arrow task in the periphery. Thus, the authors were able to find 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 et al. 1998) or do not recall seeing (Drory and 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 (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.

Overall, 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 detection while performing the cell phone distraction tasks complements other studies that show the cognitive 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, one indication that they are considered realistic and appropriate for research and training is that they are used as a surrogate for real world driving studies. 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, subjects' 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 demonstrates 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, on 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 suggest 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 position was also 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 it 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, suggesting that judgment of the lead car's speed was impaired after prolonged driving (O'Hanlon and Kelley 1977). The authors hypothesized that drivers may use a compensatory strategy for dealing with increased fatigue, such as leaving a wider gap between their vehicle and the next vehicle.

Additionally, Hulst, Meijman, and Rothengatter (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. There were two groups of subjects: an experimental group which 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.

Overall, lane position variation was larger during the last ten minutes of each ride, for the first and second times through the course, than the first ten minutes. This suggests 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, suggesting 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 are adept at promoting 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 extremely 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 comprised of dialing and answering normal and hands-free cell phones, retrieving voice mail, tuning the radio, and adjusting the temperature controls. Subjects 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 half 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.

Overall, a variety of research has been conducted using driving simulators, particularly with manipulations that would be difficult to replicate in real life:

- Driving simulators are used in a wide number of research projects and applied settings.
- Driving simulators are often used in studies where real-life manipulations would be dangerous or difficult to control.
- The fact that simulators are often used as a substitute for real vehicles suggests that researchers are confident that performance in driving simulators is a valid indicator of performance on actual road conditions.

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. Typically, trainers interested in using driving simulators for this purpose first want to compare the effectiveness of simulator training to traditional training methodologies. In order to test this, a minimum of two groups of students are required: one group receives the normal training while the other group receives simulator training. Critically, to ensure that differences in performance are not due to an uncontrolled or unknown variable, it is necessary to randomize students into these two groups. Theoretically, any inherit differences between people should be spread approximately equally between the two groups. Moreover, if the trainer is concerned that a particular variable might have a drastic effect on performance, such as age or experience, it can be controlled for by assigning an equal number of individuals from each category (e.g., “young” and “old”) into each group.

A study conducted for the Federal Motor Carrier Safety Administration (Emery et al. 1999) demonstrates this strategy of using a control and experimental group when making a decision on training effectiveness. The goal of this study was to compare the effectiveness of utilizing driving simulators for training, testing, and licensing to that of typical real-vehicle training. A group of students were to be randomly divided into two groups. One group would receive conventional truck training with additional simulator training, while the other group would receive only truck training. After receiving training, the groups would be tested on basic vehicle operations and safe operating practices. Since participants were randomized into the two separate groups, differences in performance could be attributed to the type of training that they received. Leitao et al. (1999) used a similar methodology. In their design, students were to complete a preliminary drive in a driving simulator and then be randomly assigned to one of four groups, which involved varying amounts of simulated and conventional driving training. At the end of training, all students would complete the same driving assessment regardless of which group they were in, and their performance on this assessment could be used to determine which training methodology was most effective.

When conducting research on or providing training for a large number of individuals, it is imperative for sessions to last as short as possible while still being 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 et al. (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. The authors state 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. Research suggests that this leads to better generalization from training to performance in the field (Ivancic and 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 subjects' motivation (Ivancic 1997).

Training and testing were conducted in a driving simulator (Ivancic and Hesketh 2000). Experiment 1 compared error training to errorless training, while Experiment 2 compared guided error training to errorless training. Each subject completed two training scenarios. The first training scenario was designed to familiarize subjects with the simulator controls and provide feedback, in the error conditions on their performance. The second scenario was the test scenario and 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 (2000) 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 received the same 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 subjects 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 appear to be more effective than not providing feedback or requiring trainees to watch other drivers make mistakes.

Studies on using simulators for driver training have yielded the following conclusions:

- Randomization into groups is critical when comparing two different training regimens. Otherwise, differences in performance between the two groups may be attributed to some uncontrolled variable.
- Hands-on training where students are actively involved is preferable to having trainees passively watch videos or listen to an instructor.

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 research: (1) the effects of distractions and (2) 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, Harbluk, and Engstrom (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 the impact of distraction tasks. Previous research suggests that drivers 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 attention 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 eye movement impact 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, Johansson, and Ostlund (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 suggests 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 and Gaillard 1996; Wilson 1993; Goldstein, Bauer, and Stern 1992). The present study 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. Subjects 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. Moreover, 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 duration and frequency based on the type of critical event. For instance, after the first lap, participants did not show a reduction 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 reduce 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.

Studies of drivers' eye movements have provided researchers with the following information:

- Eye movements are measured for two reasons: to find the typical, adaptive scan pattern for drivers during normal vehicle operation, and to determine the effects of certain distractions on people's scanning behavior.

- Typically, drivers use a time-share method, in which they focus primarily on the road ahead of them except for short periods of time (typically up to 1.6 seconds) during which they glance at their mirrors or speedometer.
- People's eye movements become less adaptive and more sporadic as they become fatigued.

Novice Versus Expert Drivers

Recently, Underwood and his colleagues (Crundall and Underwood 1998; Crundall, Underwood, and Chapman 1999; Underwood, Crundall, and Chapman 2002; Crundall, Underwood, and Chapman 2002; Underwood, Chapman, Bowden, and Crundall 2002; Chapman, Underwood, and Roberts 2002; Underwood, Chapman, Berger, and Crundall 2003; Underwood, Chapman, Brocklehurst, Underwood, and Crundall 2003; Crundall, Shenton, and 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 novice 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 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 and 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 are unable to handle the cognitive load of driving a vehicle.

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.

Moreover, Crundall and Underwood (1998) found that novice drivers' mean eye fixation durations were longest on the most demanding road (urban). The authors point 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, subjects 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' adapting 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, and Crundall 2002). Again, participants were not required to operate the vehicle and only needed to direct their eye movements to the most relevant locations in the scene. Thus, if there were differences between fixation locations of novice and expert drivers, it would almost assuredly be due to novice drivers having a lesser understanding of proper scanning techniques. 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, Bowden, and Crundall (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 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 is 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, the authors report 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.

The previous studies reveal the following information regarding the performance of novice and expert drivers:

- Novice and expert drivers show different scanning patterns while driving. Specifically, novice drivers scan a narrower portion of the road.
- These differences are likely due to an inadequate understanding of proper scanning techniques.
- Novice drivers showed some improvement after receiving training designed to increase their spread of search.
- Research using driving simulators shows that police officers employ their specialized training. Thus, differences between novice and expert drivers are likely to manifest themselves in driving simulators as well as real-life driving situations.

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, and 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 suggest 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 truly 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 significant improvement in fuel efficiency. Overall, Strayer, Drews, and Burns (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.

The following discoveries were made from the UDOT snowplow simulator study:

- Utah DOT operators rated the simulator and training very highly.
- There was some evidence to suggest that training may have improved operators' performance, as those individuals who received training had fewer accidents and improved fuel efficiency compared to a comparable control group that did not receive training.

Cybersickness

One of the major drawbacks of virtual reality environments is their tendency to cause discomfort in individuals, a condition known as cybersickness. Studies have reported (Stanney and Salvendy 1998) that up to 95% of participants experience some form of cybersickness, and approximately 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 (involving balance, movement, and orientation system 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: 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 other 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 as environment, the more intense the person's discomfort will become. While this explanation has similar faults to the previous two theories, it does accurately predict 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 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, and Furness 2004). A rest frame is any object that an individual perceives to be stationary (LaViola 2000) and that 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 and others (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. Thus, 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 et al. (1993). The SSQ consists of 16 questions that comprise three subscales: nausea, oculomotor discomfort, and disorientation. Subjects respond from “none” to “severe” (on a scale of 0 to 3) for each question, and the total score is calculated by multiplying the sum of these responses by 3.74. 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 report 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., 1) for every question. Similarly, Arms and Cerney (2005) report 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 subjects 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.

The following conclusions about cybersickness were determined in previous research projects:

- It is normal for the majority of individuals to report some feelings of cybersickness. Typically, though, only around 20% to 30% of individuals experience cybersickness to the extent that they cannot continue to operate the simulator.
- The typical explanation for cybersickness is that there is a conflict between inputs from the vestibular system (which indicates that the individual is stationary) and the visual system (which indicates that the individual is moving).
- Modern high-fidelity simulators tend to cause less simulator sickness than older models due to advances in screen refresh rates.
- Noting stationary objects within the simulator and reducing the amount of steering and braking during the adaptation/introduction phase of training may reduce the amount of cybersickness.

PERSONALITY

Within the past 100 years, personality researchers have set out to develop accurate, reliable, and frugal methods of classifying individuals based on certain 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 and 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 and 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" (160). More recent research, on the other hand has offered cautious support for the use of personality measurements for employment decisions (Barrick and Mount 1991; Tett, Jackson, and 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 and Costa 1985; Costa and 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 there are as many as 18,000 (Allport and 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 and Costa 2003; Barrick and 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 factor. 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.

Table 1. Normative means and standard deviations for the NEO-FFI

Authors	Mean (and Standard Deviation) by Domain				
	Neuroticism	**Extraversion**	**Openness**	**Agreeableness**	**Conscientiousness**
Rolland, Parker, & Stumpf (1998)	17.60 (7.46)	27.22 (5.85)	27.09 (5.82)	31.93 (5.03)	34.10 (5.95)
Saucier (1998)	18.02	26.71	23.82	33.75	33.94

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 et al. 2000; Costa and McCrae 1988; McCrae and 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) report on several studies which concluded that trying to correct for these types of responses does not improve the validity of subjects' scores and may actually impair it (Dicken 1963; McCrae and Costa 1983; Piedmont et al. 2000). Thus, they recommend accepting individuals' self-report scores as long as they have no motivation to fake them.

In summary,

- The Big Five, or NEO Five Factor Inventory, measures peoples' underlying personality traits. The test is divided into five domains: neuroticism, extraversion, openness, agreeableness and conscientiousness.
- Each domain represents a continuum between two extremes (e.g., introversion and extraversion). Scores on the NEO-FFI can range between 12 and 60, and normative scores for each domain are reported in Table 1.

The Big Five and Job Performance

As mentioned above, there has been some reluctance in the past 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 which 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. As mentioned above, although minor differences in names of subscales or questions are sometimes found in the literature, the five dimensions are widely accepted and have been verified across many studies.

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 modest, 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 and 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, Jackson, and Rothstein (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, Jackson, and Rothstein (1991) 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, Jackson, and Rothstein (1991) and Barrick and Mount (1991) 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 tend to agree, 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 vocation 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 1997). 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 subjects) from a wide variety of college majors and professions. Overall, they found several moderately high correlations between Big Five personality domains and interests on Holland's test (see table 2).

Table 2. Moderate correlations between Big Five and Big Six personality dimensions*

Big Five Personality Domain	Holland RIASEC Big Six Interest
Openness	Artistic
Extraversion	Enterprising
Extraversion	Social
Emotional Stability	Enterprising
Agreeableness	Social
Conscientiousness	Conventional

*See De Fruyt and Mervielde (1997, 94) for a list of 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, Rottinghaus, and Borgen 2002, 223). However, neither De Fruyt and Mervielde (1997) nor Larson, Rottinghaus, and Borgen 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-PI 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.

Research indicates the following conclusions regarding the use of personality tests as indicators of job performance:

- The NEO-PI does a reasonable job at predicting job performance. Correlations between job performance measurements for the conscientiousness and agreeableness domains typically ranging between 0.20 and 0.35. Overall, conscientiousness appears to be the best predictor of job performance.
- Another indication that the NEO-PI can be used as an indicator of job compatibility is that several of the domains correlate with Holland's Big Six Interests vocational model, which is designed to match an individual's interests with a 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. Since conscientiousness is associated with traits such as compliance and responsiveness to social norms, 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. Subjects 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 conformed: 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 which require large amounts of driving, conscientiousness may therefore be an even greater predictor of job performance than suggested by previous studies.

To summarize, in addition to being the best predictor of job performance, conscientiousness appears to be the best predictor, out of the domains measured by the NEO-PI, of driving performance.

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 and Viswesvaran 2001a). Examples include integrity tests, stress tolerance, and violence scales. Thus, unlike traditional personality tests, the COPS are explicitly designed to predict job performance measures and differences in employees' work behavior.

Ones and Viswesvaran (2001a) discussed the validity of several of these measures as they relate both to the criteria that they are designed to measure and to overall job performance. Integrity tests, for instance, were shown to be highly correlated with counterproductive behavior at work, supervisors' ratings of employee job performance, and on-the-job accidents (Ones, Viswesvaran, and 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 and Viswesvaran, 2001a, 35 for a list of correlations.) These scales have also been shown to moderately correlate with the conscientiousness domain (Ones and Viswesvaran, 2001b).

Evidence that many of these COPS are highly correlated with well-established personality dimensions leads 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. (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 suggest 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 extol 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.

With regards to traditional personality assessments, Robertson and Smith (2001) agree 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 et al. 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.

When it comes to predicting job performance, assessments other than NEO-FFI have specific benefits to offer:

- Criterion-focused Occupational Personality Scales (COPS) are designed to measure specific aspects of job performance, such as integrity and stress tolerance. Depending on the intentions of the employer, they can be a more informative indicator of job performance than the NEO-FFI.

- Other job performance assessments, such as structured interviews, are also highly correlated with job performance ratings.

OTHER MEASUREMENTS RELAVENT 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.

The version of Zuckerman's Sensation Seeking Scale employed in this study, Form V, is comprised of 40 total questions from four subscales: thrill and adventure seeking (TAS), experience seeking (ES), disinhibition (Dis), and boredom susceptibility (BS). Note that in the present study, some questions were removed due to the circumstances in which the questionnaire was administered. For instance, questions concerning drug and alcohol use were omitted. Some common findings include an overall negative correlation with age as well as some minor cultural differences (Zuckerman, Eysenck, and Eysenck 1978). Positive correlations were discovered with drug use (Forsyth and 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" (227). Presence was defined as "the subjective experience of being in one place or environment, even when one is physically situated in another" (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 tasks in general. 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 subjects' 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. Similar to cybersickness levels, average presence scores are again 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 et al. 2000; Ooms 2004). Both of these studies used head-mounted virtual reality displays, which Ooms (2004, 1) states, "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 subject. 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 and in real vehicles with regards to 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. Moreover, an upcoming study by these authors will seek to address, among other things, 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.

REFERENCES

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

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

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., and W. G. Graziano. 1996. The five-factor model, conscientiousness and driving accident involvement. *Journal of Personality* 64:593–618.

Barrick, M. R., and M. K. Mount. 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 instrumented 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., and M. S. Rea. 2001. Forward vehicular lighting and inclement weather conditions. PAL 2001 Symposium, Darmstadt University of Technology.

Campagne, A., T. Pebayle, and A. Muzet. 2005. Oculomotor changes due to road events during prolonged monotonous simulated driving. *Biological Psychology* 68:353–368.

Chapman, P., G. Underwood, and K. Roberts. 2002. Visual search patterns in trained and untrained novice drivers. *Transportation Research Part F: Traffic Psychology and Behaviour* 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., D. L. Warren, H. Bissell, and S. Taori. 1998. Are the criteria for setting advisory speeds on curves still relevant? *ITE Journal* (Institution of Transportation Engineers) 68:32–45.

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

Costa, P. T., and R. R. McCrae. 1985. *The NEO Personality Inventory Manual*. Odessa, FL: Psychological Assessment Resources.

———. 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., C. Shenton, and G. Underwood. 2004. Eye movements during intentional car following. *Perception* 33:975–986.

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

Crundall, D., G. Underwood, and P. Chapman. 1999. Driving experience and the functional field of view. *Perception* 28:1075–1087.

———. 2002. Attending to the peripheral world while driving. *Applied Cognitive Psychology* 16:459–475.

De Fruyt, F., and I. Mervielde. 1997. The five-factor model of personality and Holland's RIASEC interest types. *Personality and Individual Differences* 23:87–103.

Dicken, C. 1963. Good impression, social desirability, and acquiescence as suppressor variables. *Education and Psychological Measurement* 23:699–720.

Dorn, L., and D. Barker. 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., and D. Shinar. 1982. The effects of roadway environments and fatigue on sign perception. *Journal of Safety Research* 13:25–32.

Duh, H. B. L., D. E. Parker, and T. A. Furness. 2004. An independent visual background reduced simulator sickness in a driving simulator. *Presence* 13:578–588.

Emery, C., J. Robin, R., Knipling, R. Finn, and S. Fleger. 1999. Research design: Validation of simulator technology in the training, testing and licensing of tractor-trailer drivers. FHWA Final Report FHWA-MC-99-060.

Engstrom, J., E. Johansson, and J. Ostlund. 2005. Effects of visual and cognitive load in real and simulated motorway driving. *Transportation Research Part F: Traffic Psychology and Behaviour* 8:97–120.

Enriquez, M., O. Afonin, B. Yager, and K. Maclean. 2001. A pneumatic tactile alerting system for the driving environment. In *Proceedings of the 2001 workshop on perceptive user interfaces*, 1–7.

Fischer, J. 1992. Testing the effect of road traffic signs' informational value on driver behavior. *Human Factors* 34: 231-237.

Forsyth, G., and J. D. Hundleby. 1987. Personality and situation as determinants of desire to drink in young adults. International Journal of the Addictions 22:653–669.

Godley, S. 1999. A driving simulator investigation of perceptual countermeasures to speeding. PhD disseration, Monash University.

Godley, S. T., T. J. Triggs, and B. N. Fildes. 2002. Driving simulator validation for speed research. Accident Analysis and Prevention 24:589–600.

———. 2004. Perceptual lane width, wide perceptual road centre markings and driving speeds. Ergonomics 47:237–256.

Goldstein, R., L. O. Bauer, and J. A. Stern. 1992. Effect of task difficulty and interstimulus interval on blink parameters. International Journal of Psychophysiology 13:111–117.

Greenberg, J., L. Tijerina, R. Curry, B. Artz, L. Cathey, P. Grant, D. Kochhar, K. Kozak, and M. Blommer. 2003. Evaluation of driver distraction using an event detection paradigm. Journal of the Transportation Research Board 1843:1–9.

Guion, R. M., and R. F. Gottier. 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. 1994. Driving performance on a real road and in a driving simulator: results of a validation study. In Vision in Vehicles, North Holland; Elsevier.

Hogan, R. 1986. Manual for the Hogan Personality Inventory. Minneapolis: National Computer Systems.

Holland, J. L. 1997. Making Vocational Choices. Englewood Cliffs, NJ: Prentice Hall.

Holland, J. L. 1985. Self-Directed Search-1985 edition. Odessa, FL: Psychological Assessment Resources.

Horberry, T., J. Anderson, and M. A. Regan. 2005. The possible safety benefits of enhanced road markings: A driving simulator evaluation. *Transportation Research Part F: Traffic Psychology and Behaviour* 9:77–87.
Hulst, M., T. Meijman, T. Rothengatter. 2001. Maintaining task set under fatigue: A study of time-on-task effects in simulated driving. *Transportation Research Part F: Traffic Psychology and Behaviour* 4:103–118.
Ivancic, K. 1997. Errors as a means of promoting transfer of training. PhD dissertation, University of New South Wales.
Ivancic, K., and B. Hesketh. 1995. Making the best of errors during training. *Training Research Journal* 1:103–125.
———. 2000. Learning from errors in a driving simulation: Effects on driving skill and self-confidence. *Ergonomics* 43:1966–1984.
Kemeny, A., and F. Panerai. 2003. Evaluating perception in driving simulation experiments. *Trends in Cognitive Sciences* 7:31–37.
Kennedy, R. S., N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. 1993. A simulator sickness questionnaire (SSQ): A new method for quantifying simulator sickness. *International Journal of Aviation Psychology* 3:203–220.
Larson, L. M., P. J. Rottinghaus, and F. H. Borgen. 2002. Meta-analyses of big six interests and big five personality factors. *Journal of Vocational Behavior* 61:217–239.
LaViola, J. J. 2000. A discussion of cybersickness in virtual environments. *SIGCHI Bulletin* 32:47–56.
Lee, J. D., J. D. Hoffman, and E. Hayes. 2004. Collision warning design to mitigate driver distraction. *CHI 2004*, 65–72.
Leitao, J. M., A. Moreira, J. A. Santos, A. A. Sousa, and F. N. Ferreira. 1999 Evaluation of Driving Education Methods in a Driving Simulator. In *Proceedings of GVE 1999*, 159–164.
McCrae, R. R., and P. T. Costa. 1983. Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology* 51: 882-888.
———. 1985. Openness to experience. In *Perspectives in Personality*, Greenwich, CT: JAI Press.
———. 2003. *Personality in Adulthood.* New York: The Guildford Press.
McGehee, D. V., J. D. Lee, M. Rizzo, J. Dawson, and K. Bateman. 2004. Quantitative analysis of steering adaptation on a high performance fixed-base driving simulator. *Transportation Research Part F: Traffic Psychology and Behaviour* 7:181–196.
Miller, T. R. 1992. Benefit-cost analysis of lane marking. *Transportation Research Record* 1334:38–45.
O’Hanlon, J. F., and G. R. Kelley. 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. PhD dissertation, University of Iowa.
Ones, D. S., C. Viswesvaran. 2001a. Integrity tests and other criterion-focused occupational personality scales (COPS) used in personnel selection. *International Journal of Selection and Assessment* 9:31–39.
———. 2001b. Personality at work: Criterion-focused occupational personality scales (COPS) used in personnel selection. In *Applied Personality Psychology*. Washington DC: American Psychological Association.

Ones, D. S., C. Viswesvaran, and F. L. Schmidt. 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. Paper presented at Le Twente Student Conference on IT, Enschede.

Panerai, F., J. Droulez, J. M. Kelada, A. Kemeny, E. Balligand, and B. Favre. 2001. Speed and safety distance control in truck driving: Comparison of simulation and real-world environment. *DSC2001–Sophia Antipolis.*

Peli, E., A. R. Bowers, A. J. Mandel, K. Higgins, R. Goldstein, and L. Bobrow. 2005. The design of driving simulator performance evaluations for driving with vision impairments and visual aid. Paper presented at TRB Annual Meeting.

Piedmont, R. L., R. R. McCrae, R. Riemann, and A. Angleitner. 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., A. Chatziastros, W. Cunningham, H. H. Bulthoff, and J. E. Cutting. 2002. Gaze-eccentricity effects on road position and steering. *Journal of Experimental Psychology: Applied* 8:247–258.

Redelmeier, D., and R. Tibshirani. 1997. Association between cellular-telephone calls and motor vehicle collisions. *The New England Journal of Medicine* 336:7.

Rizzo, M., R. A. Sheffield, L. Stierman, and J. Dawson. 2003. Demographic and driving performance factors in simulator adaptation syndrome. Paper presented at the 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.

Robertson, I. T., H. Baron, P. Gibbons, R. MacIver, and G. Nyfield. 2000. Concientiousness and managerial performance. *Journal of Occupational and Organizational Psychology* 73:171–180.

Robertson, I. T., and M. Smith. 2001. Personnel selection. *Journal of Occupational and Organizational Psychology* 74:441–472.

Rolland, J. P., W. D. Parker, and H. Stumpf. 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 *International Review of Industrial and Organizational Psychology*. New York: Wiley.

Santos, J., N. Merat, S. Mouta, K. Brookhuis, and D. Waard. 2005. The interaction between driving and in-vehicle information systems: Comparison of results from laboratory, simulator and real-world studies. *Transportation Research Part F: Traffic Psychology and Behaviour* 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., and S. Kiesler. 2005. Calling while driving: Effects of providing remote traffic context. *CHI: 2005*, 561–569.

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

Sodhi, M., B. Reimer, J. L. Cohen, E. Vastenburg, R. Kaars, and S. Kirschenbaum. 2002. On-road driver eye movement tracking using head-mounted devices. Paper presented at the Eye Tracking Research and Applications Symposium, New Orleans, LA.

Stanney, K. M., and G. Salvendy. 1998. Aftereffects and sense of presence in virtual environments: Formulation of a research and development agenda. *International Journal of Human-Computer Interaction* 10:135–187.
Strayer, D. L., and F. A. Drews. 2003. Simulated training improves driver efficiency: Transfer from the simulator to the real world. In *Driving Assessment 2003: International Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*.
Strayer, D. L., F. A. Drews, and S. Burns. 2004. The development and evaluation of a high-fidelity simulator training program for snowplow operators. Utah Department of Transportation Report UT-04.17.
Tett, R. P., D. N. Jackson, and M. Rothstein. 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. *Accident Anaysis and Prevention* 30:497–503.
Underwood, G., P. Chapman, K. Bowden, and D. Crundall. 2002. Visual search while driving: Skill and awareness during inspection of the scene. *Transportation Research Part F: Traffic Psychology and Behaviour* 5: 87–97.
Underwood, G., P. Chapman, N. Brocklehurst, J. Underwood, and D. Crundall. 2003. Visual attention while driving: Sequences of eye fixations made by experienced and novice drivers. *Ergonomics* 46:629–646.
Underwood, G., P. Chapman, Z. Berger, and D. Crundall. 2003. Driving experience, attentional focusing, and the recall of recently inspected events. *Transportation Research Part F: Traffic Psychology and Behaviour* 6:289–304.
Underwood, G, D. Crundall, and P. Chapman. 2002. Selective searching while driving: The role of experience in hazard detection and general surveillance. *Ergonomics* 45:1–12.
Usoh, M., E. Catena, S. Arman, and M. Slater. 2000. Using presence questionnaires in reality. *Presence* 9:497–503.
Veltman, J. A., and A. W. K. Gaillard. 1996. Physiological indices of workload in a simulated flight task. *Biological Psychology* 42:323–342.
Victor, T. W., J. L. Harbluk, and J. A. Engstrom. 2005. Sensitivity of eye-movement measures to in-vehicle task difficulty. *Transportation Research Part F: Traffic Psychology and Behaviour* 8:167–190.
Wierwille, W. 1993. Visual and manual demands of in-car controls and displays. In *Automotive Ergonomics*. New York: Taylor and Francis.
Wilson, G. R. 1993. Air-to-ground training missions: A psychological workload analysis. *Ergonomics* 36:1071–1087.
Witmer, B. G., and M. J. Singer. 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., S. Eysenck, and H. J. Eysenck. 1978. Sensation seeking in England and America: Cross-cultural, age and sex comparisons. *Journal of Consulting and Clinical Psychology* 46:139–149.