LOW HANGING FRUIT: USE OF VIRTUAL REALITY DRIVING SIMULATION IN DEPARTMENT OF MOTOR VEHICLES TO ASSESS MINIMAL COMPETENCE OF NOVICE DRIVERS.

Daniel J. Cox[1], Matthew Moncrief[1], Matthew Rizzo, Donald Fisher,
Ann Lambert[1], Sarah Thomas[1], Sean Eberhart[1], Rick Moncrief
[1]University of Virginia - Charlottesville, Virginia, U.S.A
[2]University of Nebraska Medical Center
[3]University of Massachusetts – Amherst
djc4f@virginia.edu

Summary: Nationally, Departments of Motor Vehicles (DMV) license novice drivers based in part on on-road assessments. Intuitively it is assumed that such assessments are fair, reliable and valid measures of minimal driving competency. Upon further reflection, this would be difficult, given the subjectivity of a huge range of driving examiners that approach this assessment with different training backgrounds, life distractions and biases from examination to examination, the different road, traffic, lighting and weather conditions from one examination and DMV center to the next, and the minimal driving challenges in such assessments. For example, a typical on-road test involves only a 4 mile road segment with 2 left turns, 4 right turns, 1 lane change, pulling into a turn lane, and 1 speed limit change. It does not include highway driving nor defensive driving maneuvers. Additionally, such on-road assessments are both potentially dangerous and time demanding/expensive. A less expensive, safer, more challenging, objective, reliable, and valid procedure may be the use of Virtual Reality Driving Simulation (VRDS) that administers consistent and more extensive driving challenges to all examinees, which is evaluated in an objective manner based on normative data from current safe drivers. This presentation describes the experience and presents the data from a project where VRDSs were set up in two DMV facilities and a Research facility. The goals of this project were to determine whether VRDS assessments are just as reliable, discriminating and acceptable to the public as on-road assessments, and whether performance on the simulator predicts future on-road driving mishaps.

OBJECTIVE

One of the most extensive investigations of Department of Motor Vehicles (DMV) on-road licensure exam was conducted by the California DMV between 1990 and 1995. They compared the reliability and validity (Hagge, 1994, Romanowicz PA and. Hagge RA 1995) of their standard and a revised Driving Performance Evaluation (***DPE***) on-road assessment. The DPE differed from the standard exam in that it was 11 minutes longer (13.6 minutes vs. 24.8 minutes), included freeway driving, and used more objective or standardized criteria in which defined maneuvers were observed and scored only at preplanned locations and times during the test. The DPE had a .81 inter-rater reliability, a .83 inter-route reliability, and a .78 net reliability. The standard drive test was less reliable, with coefficients of .69, .66, and .60 on the same three reliability measures, respectively. The evaluation also found the DPE to be more difficult than the standard drive test, with fail rates of 45.6% and 26.2% for the two tests, respectively. Validity of the DPE was partially confirmed by an experienced driver group who performed

significantly better on the DPE than did the novice driver group. The fail rates and average scores (number of errors) for these two groups were 5% vs. 48% and 10.6% vs. 15.8%, respectively. Although the difference was not statistically significant, accident-involved subjects had a higher DPE fail rate than did accident-free subjects.

However the DPE also carried several limitations. Scores on the DPE did not correlate with future driving mishaps. The pass/fail rate for the DPE was only slightly higher than the current test, although scores on the current test "were believed to be inflated by examiner's knowledge that they were involved in a study and the fact that they were given refresher training in current drive test scoring criteria prior to the study" (Hagge, 1995, ii). In addition to these reliability and validity concerns there are practical limitations of such procedures:

- Need for specially trained staff to perform these exams,
- Staff time/costs to provide these exams,
- Customers' time waiting for their exams,
- Delays of the exam due to factors such as bad weather and traffic jams,
- Risk of injury and the cost to cover any mishaps that occur while performing such exams,
- Public relations issue in that whenever a driver fails the exam, they can accuse the DMV of subjectivity and prejudice. e.g. "He failed me because I was old, female, black, Hispanic…"

An alternative approach is to use Virtual Reality Driving Simulation (VRDS) assessments that can:

- Deliver standardized scenarios, incorporating standardized challenges, employing standardized objective and validated criteria,
- Deliver assessments in a safe environment,
- Allow quick transitions between Urban, Rural, and Highway driving when on-road exams are bound to the roads surrounding their Service Center,
- Automate assessments in various languages, not requiring specialized staff,
- Operate regardless of ambient weather and lighting conditions, and mimic challenging but rare weather and lighting conditions
- Allow predictable scheduling of assessments at fixed and known time intervals,
- Generate immediate paper and computer reports that could be given to the examinee and stored for later analyses,
- Eliminate the need for a pre-test drive in the DMV parking lot that documents minimal competency/safety before venturing onto actual roads with traffic,

However, there are limitations to VRDS, which include:

- Initial costs of the simulator and subsequent maintenance costs,
- Space to house the simulator,
- Simulation Adaptation Syndrome that may preclude its use with all customers, and
- The general public's relative unfamiliarity with VRDS of assessments and technology.

Hypotheses

1) VRDS would demonstrate reliability (similar to DPE) and additionally not demonstrate a practice effect, i.e. not improve with repeat testing. 2) VRDS would discriminate low-risk drivers (ages 25-75) from novice drivers. As a further measure of validity, VRDS would differentiate low-risk from higher risk drivers (ages 16-25) and neruotypical novice drivers from novice

drivers with Autism Spectrum Disorder (ASD). 3) These healthy drivers would not differ on basic driving abilities, like vision, foot reaction time. 4) VRDS performance would predict future driving mishaps unlike DPE...

METHODS

Overview

With funds from the Virginia state legislature, in collaboration with the Virginia DMV, a project was executed to evaluate the routine use of VRDS as a possible alternative to on-road assessment. VRDS was employed at two DMV sites (Westfield and Charlottesville) collecting normative and reliability data from customers coming to the DMV for non-driver related issues, e.g. registering their vehicle, bringing in a child for testing, etc. 448 DMV customers drove the simulator and 37 came back two weeks later to repeat testing. In addition, we evaluated 38 adolescent novice drivers who had just qualified for an independent driver's license. It was hypothesized that novice drivers would not differ from experienced drivers in regard to basic driving *abilities*, like foot and arm reaction time, peripheral vision, and response inhibition, but would differ from experienced drivers in terms of driving *skills*, like steering, brake and speed control which would predict novice drivers' future driving mishaps. Additionally, we evaluated a second group of novice drivers who had earned a learner's permit with the diagnosis of Autism Spectrum Disorder. They were hypothesized to perform worse than the neuro-typical novice drivers in terms of driving skills.

Facilities

At the DMV service centers, the simulators were housed in a cubical in the waiting area. The novice driver data was collected in the Virginia Driving Safety Laboratory on the campus of the University of Virginia Health Sciences Center.

With SBIR grants from NIA and DARPA, we developed a VRDS system that has an 8ft diameter 210º curved screen, with a realistic driver's seat and controls, side and rear view mirrors and air conditioning (Figure 1). It has two assessment capabilities, operational and tactical tests. Scores on different tests are converted to z-scores, allowing a common metric, and then summed and normalized into a **Driving Quotient**, much like an Intelligence Quotient with a mean of 100 and a SD of 15.
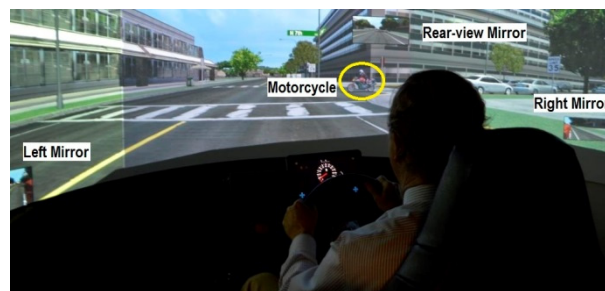

Figure 1. Simulator displaying a road hazard (motorcyclist emerging from behind traffic) requiring a defensive maneuver

Operational Testing. Operational Testing was developed with driving-specific visual, motor, and cognitive tests modeled after traditional neuropsychological tests. These tests use driving-relevant stimuli, responses, and context in order to enhance ecological validity. All of these tests

employ the same environment, thus reducing re-adaptation from one test to another. The examinee drives down the middle lane of a three-lane highway at 35mph, maintaining a constant distance from a lead car. To equate task instructions, all subjects hear the same instructions, delivered at the same time, by the simulator's synthetic voice.

Table 1. Listing of Operational Variables

Operational Tests		Task	Dependent Variable	Reliability
Operational Driving Quotient: Sum of z scores				r= .52**
Motor	*Braking Reaction time*	Lead car's brake lights come on 10 times: 5 for 3 sec. and 5 for 0.5 sec. Driver removes foot from accelerator and presses the brake as quickly as possible	Milliseconds between brake lights on and 5 lbs. of pressure applied to brake pedal.	r= .85**
	Steering Reaction time	Lead car passes over 12 potholes: 6 filled grey and 6 deep black potholes. Driver steers around potholes as quickly as possible while staying in lane	Milliseconds between when pothole appears and initiating steering maneuver.	r= .96**
Executive Function	*Dual Processing*	Lead car's brake lights come on 8 times and passes over 8 potholes. Driver is to brake to all brake lights and steer around all potholes as quickly as possible	Total correct responses: brake to brake lights and steer around potholes.	r= .82**
	Response Inhibition	Same as Dual Processing, but inhibit previous prepotent response, i.e. do not respond to brief brake lights or grey potholes, while continuing to brake to long brake lights and steer around black potholes	Total correct responses: brake to long brake lights and steer to deep potholes, and not respond to brief brake lights and grey potholes.	r= .59*
	Working Memory	Same as Response Inhibition and additionally driver has to remember 1 to 3 road signs recently passed in the order they appeared.	Signs recalled in correct order.	r= .82**
* Correlation is significant at the .05 level.				
** Correlation is significant at the .01 level.				
Underlined values are measures of internal consistency.				

Tactical Testing. This is analogous to an on-road test of driving skills, but performed in a safe, reliable, and challenging virtual world. The tactical test involves driving on a standardized route, which includes five miles of rural, six miles of highway, and four miles of urban roads. Drivers negotiate realistic anticipated and unanticipated road, signal, traffic, and hazard demands. Thirty-one performance variables are monitored, such as swerving, rolling stops, speeding, and collisions. Thirteen of these variables were selected for a composite Driving Quotient because these variables identified drivers with a history of collisions. Table 2 details the thirteen tactical variables.

Table 2. Listing of tactical variables

Tactical Variables	Description	Reliability
Tactical Driving Quotient: Sum of z scores		r= .80**
B *Rolling Stops*	Ratio of incomplete (>0 and <5 mph) to complete (0 mph) stops.	r= -.04

	Deceleration Smoothness	Total magnitude of rapid decelerations; i.e., slamming on brakes.	r= .67**
	Bumps	Number of collisions with another vehicle below 5 mph.	r= .49**
	Collisions	Number of collisions with another vehicle exceeding 5 mph.	r= -.42*
Speed	*Acceleration Smoothness*	Total magnitude of rapid accelerations; i.e., slamming on gas.	r= .94**
	Speeding	Total time spent driving 5-19 mph above the posted speed limit.	r= .68**
	Reckless Driving	Total time spent driving 20 mph or more over the posted speed limit.	r= .80**
	Tailgating	Number of times driver is within 15 feet of lead car during open road segments.	r= .18
Steering	*Swerving*	Lane position variability (standard deviation of lane position; i.e., swerving).	r= .83**
	Midline	Average magnitude active; integrated score of how far across and for how long driver was in oncoming lane of traffic.	r= .58**
	Off Road	Standard deviation time active; variability of length of time driver was off road.	
Judgment	*Lane Changes*	Total number of lane changes made on route.	r= .53**
	Driving Slow	Average time spent 20 mph or more under the posted speed limit.	r= .38*

Variables are recorded for the entirety of the route.
a. Cannot be computed because at least one of the variables is constant.
* Correlation is significant at the .05 level.
** Correlation is significant at the .01 level.

Subjects

DMV subjects consisted of 327 males and 121 females, with a mean age of 40, with an average of 27 years driving experience. Given the relatively flat distribution of driving collisions and citations between the ages of 25-75, this "Normative group" was used to generate means and standard deviations to calculate z scores for an individual's different test performance culminating in an individual's Driving Quotients. Since drivers progressively younger than age 25 have progressively more driving mishaps, we used this Under 25 sample to compare to the Normative group to validate the VRDS. U.Va. was concurrently conducting two studies evaluating: 1) 38 Neuro-typical novice drivers at the time of their independent licensure, and 2) 17 Autism Spectrum Disorder (***ASD***) novice drivers who were trying to earn their independent driver's license. In addition to VRDS assessment, neurotypical novice drivers were contacted every two weeks, during their first six months of independent driving, with text messages inquiring whether they had experienced driving mishaps (collision, near collision, citation) over the previous 2 weeks. Neuro-typical novice drivers had a mean age of 16.2 and were 35% female, while the ASD sample had a mean age of 18.3, and were 0% female.

Procedure

DMV customers were not paid for driving the simulator, but the 37 who came back for test-retest reliability trial were compensated $50. The neuro typical and ASD drivers were both compensated $50 for the assessment.

RESULTS

Hypothesis 1. As illustrated in Table 2, Test-retest reliability for the Tactical Driving Quotient, which parallels on-road testing, was robust (r= +.80, p<.001). This is despite the fact that drivers knew what to expect in the second test, like turning at wrong intersections or vehicles pulling out in front of the driver, therefore allowing the driver to avoid anticipated difficulties during repeat testing. When comparing performance across sites, Tactical and Operational Driving Quotients were similar, demonstrating across sites reliability. As a further demonstration of Tactical reliability, repeat testing did not demonstrate a practice effect, i.e. there was no significant increase in the Tactical Driving Quotient from test 1 to test 2.

Hypothesis 2. Tactical Driving Quotients were progressively worse, and significantly different from normative sample for the under 25 (p<.05) and neurotypical novice drivers (p<.01, with respective Tactical Driving quotients being: 100, 94, and 92. Neurotypical novice drivers performed significantly better than ASD novice drivers (p<.001), with respective Tactical Driving quotients being: 92 and 24.

Hypothesis 3. Operational Driving Quotient, assessing basic driving abilities did not differ between the normative, under 25, neurotypical novice and ASD novice drivers, with respective means scores being 100, 104, 102 and 101, respectively.

Hypothesis 4. Neurotypical novice drivers reported an average of 1.05 driving mishaps over their initial six months of driving (range 0-6). Multiple regression analyses indicated Tactical driving performance predicted future driving mishaps, R^2 =.73 (p=.002).

Simulator Adaptation Syndrome (SAS). The percentage of individuals who discontinued the Operational testing because of SAS for the Normative, Under 25, Neurotypical and ASD samples were: 3%, 0%, 0%, 0%. For the Tactical test, the percent dropouts due to SAS for the four groups were: 17%, 3%, 0%, 0%, respectively.

Acceptance. In the below table are two questions asked of drivers after completing the Tactical tests based on a 0-4scale

Table 3. Questions asked after testing

Reactions to Tactical driving assessment	Mean/Mode
1. How well did these tasks evaluate your overall driving ability?	2.3/3
2. Which procedure would you prefer to evaluate your driving ability: (0) Driving a real car on the road with an examiner beside you evaluating your driving ability, or (4) Driving the simulator alone with a computer evaluating your driving ability?	1.5/0

DISCUSSION

The current results suggest that assessment of driving competency was similar, whether it was done by Tactical VRDS or DPE, demonstrating similar reliability and validity. VRDS further demonstrated no practice effect and ability to differentiate Under 25 from Normative drivers and Neurotypical from ASD novice drivers. As further validation of VRDS, tactical driving

performance predicted future driving mishaps by novice drivers (R2 =.73), as tactical simulator performance predicts future collisions by senior drivers (Cox et al 1999). We are currently analyzing ASD Tactical test performance and their subsequent on-road test performance. Psychometrically, this would suggest VRDS could serve as an alternative to well-structured on-road exams administered by well-trained examiners. While possibly equivalent in terms of assessing driving competency, VRDS has several previously mentioned advantages, including safety, rapidly processing customers in a predictable manner, regardless of outside weather conditions, not requiring highly trained examiners.

However, there are two caveats. First SAS was an issue, albeit small since none of the novice drivers and only 3% of the Under 25 drivers experienced SAS. Second, while the public (DMV customers) thought VRDS did an adequate job assessing their driving competency, there was a preference for on-road assessments. As for the former, VRDS is always improving, with faster frame rates, greater resolution, employment of less provocative maneuvers and subject instruction (e.g. "avoid eating just before driving the simulator) so that SAS could be an ever diminishing issue. As for the latter, a public relations effort may need to be launched to promote acceptance of VRDS, which has typically been done with the introduction of most new technologies.

REFERENCES

Cox DJ, Taylor P & Kovatchev B (1999). Driving simulation performance predicts future accidents among older drivers. J Am Geriatr Soc, 47(3), 381-382. PMID: 10078912

Hagge RA, (1994) THE CALIFORNIA DRIVER PERFORMANCE EVALUATION: AN EVALUATION OF A NEW DRIVER LICENSING ROAD TEST. California Department of Motor Vehicles, RSS-94-150

Romanowicz PA and. Hagge RA (1995), STATE OF CALIFORNIA DEPARTMENT OF MOTOR VEHICLES, An Evaluation of the Validity of California's Driving Performance Evaluation Road Test. California Department of Motor Vehicles RSS-95-154