COMPARISON OF VIRTUAL DRIVING TEST PERFORMANCE AND ON-ROAD EXAMINATION FOR LICENSURE PERFORMANCE: A REPLICATION STUDY

Elizabeth A. Walshe[1 2], Natalie Oppenheimer[2], Venk Kandadai[3], Flaura K. Winston[2]
[1] Annenberg Public Policy Center, University of Pennsylvania, Philadelphia, PA, USA
[2] Center for Injury Research and Prevention, The Children's Hospital of Philadelphia, PA, USA
[3] Diagnostic Driving Inc., Philadelphia, PA, USA
Email: walshee@email.chop.edu

Summary: For novice drivers, passing the on-road examination (ORE) for licensure marks the transition from supervised to unsupervised driving. However, the first months post-licensure pose the highest lifetime risk of crashing. In partnership with the Ohio Bureau of Motor Vehicles (OBMV), we have developed a virtual driving test (VDT) to enhance new driver skills testing. Through simulation, license applicants were exposed to common serious crash scenarios too dangerous for inclusion in the ORE. In a previous study of an initial sample of 2,143 driver applicants in Ohio, the acceptability, feasibility and construct validity for the VDT was demonstrated: VDT performance variables (simulated traffic collisions and failing to stop at red lights and stop signs) were associated with failing the ORE (all $p < 0.001$). In this study, we aimed to replicate these results with a second sample of 2,500 novice drivers. The findings were in line with the previous study: VDT performance variables and driving errors differentiated those who went on to pass and fail the ORE. Future work will build and validate a more comprehensive and robust set of performance metrics and examine the predictive ability of the VDT, both for licensing results and future crashes.

INTRODUCTION

A driver's highest lifetime risk of crashes is in the months immediately following licensure (Curry et al, 2015; Gershon et al, 2018). Multiple analyses have revealed that inadequate skills and insufficient experience explain much of the early licensure crash risk (Curry et al, 2011; McDonald et al, 2014; McKnight & McKnight, 2003). However, the on-road examination for licensure (ORE)—which remains the "gate" between the learner period and licensed driving—is limited in the hazards to which it exposes applicants in part due to the risk that it would pose to the evaluators and applicants. Research has shown the value of extending the ORE in order to better identify at-risk drivers (Horswill et al, 2015; Wells et al, 2008).

Laboratory-based simulated assessments have demonstrated their utility in differentiating drivers according to skill and experience, and reveal skill deficits that are associated with real-world driving (e.g. hazard anticipation, speed management and attention maintenance) (McDonald et al, 2015; Chan et al, 2010). However, specialized hardware and software have prohibited the wide-scale field deployment of tools developed in the laboratory.

In 2017 the Ohio Bureau for Motor Vehicles (OBMV) began a pilot to assess the feasibility and utility of a virtual driving test (VDT) as a safety screener to ensure adequate preparedness and safety skills of new driver's license applicants prior to proceeding to the ORE. As described previously (Walshe et al, 2018), the VDT builds on previously validated laboratory-based simulated assessments, but is delivered as portable, self-guided virtual test that runs on standard computer hardware and can be completed in less than 15 minutes. The VDT utilizes the hierarchical model of driving behavior proposed by Berg (2006) and others (Hatakka et al, 2002) as its theoretical foundation, while also exposing drivers to the most common, serious crash scenarios in a safe and replicable way. Each drive includes situations that tax skills at the operational level for vehicle maneuverability and control, as well as those at the tactical level for knowledge of traffic rules, managing hazards and interacting with other road users (McDonald et al, 2012). A previous study demonstrated initial construct validity of the VDT by showing association between VDT metrics (simulated traffic collision, errors at red lights and stop signs, and went off-road/route) and failure on the ORE (Walshe et al., 2018). The current study aimed to replicate these findings in a new sample of data that has subsequently been collected.

METHOD

Dataset

Diagnostic Driving, Inc., provided a de-identified dataset with VDT and ORE performance data for 2,500 driver license applicants over a three-month period (January 2018-March 2018). As described previously (Walshe et al., 2018), this was part of a pilot phase and no additional data were collected during this pilot phase (demographics). All applicants took the VDT only once and fewer than 2% did not complete the VDT (largely due to being called for their ORE before completing).

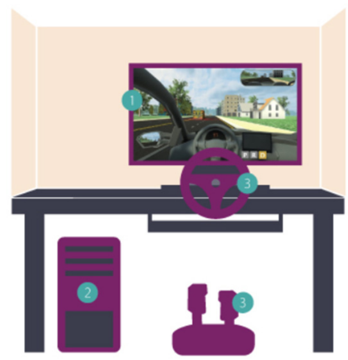

Figure 1. Workstation Setup: 1) standard monitor, 2) standard computer, 3) off-the-shelf USB steering wheel & pedals.

Procedures

Apparatus and Materials. The VDT (*Ready-Assess*™: Diagnostic Driving Inc.) was developed in collaboration between OBMV subject matter experts and the Children's Hospital of Philadelphia. The VDT is a completely self-guided, cloud-based and portable software assessment that is delivered on ubiquitous hardware (see Figure 1). See Walshe et al. (2018) for more description.

Workflow. OBMV staff invited driver license applicants arriving at test centers to volunteer to complete the VDT immediately before their pre-scheduled ORE. The VDT workflow lasted 15 minutes, and included an orientation video, acclimation drive, a brief instructions test, and finally the assessment drive (see Walshe et al., 2018). Applicants completed one of ten randomly assigned assessment course variants, all of which incorporated common, serious crash scenarios (McDonald et al, 2012; USDOT, 2008). These scenarios included: rear-end events (lead car brakes suddenly); intersections (between 8-13 with instructions to turn left, right, or continue straight at three- and four-way stop signs and traffic lights); curved roads; merges; and hazard zones (construction zones, crosswalks, etc.). In order to best reflect the local driving environment, these driving scenarios also included varied setting (urban and suburban) and on-road elements (school buses, pedestrians, etc.). These courses, designed with Ohio subject matter experts, are proprietary to Diagnostic Driving Inc. and the State and are currently an active part of the licensing workflow: thus we cannot describe the drives in greater detail. After the assessment, applicants completed a three-item feedback survey using a 5 point Likert response scale (see Walshe et al., 2018).

Analytical Procedures

Variables of Interest. Variables used included: i) categorical workflow metrics exploring applicants' ability to successfully complete various steps in the workflow, and ii) driving performance variables, examining simulated driving skills (Table 1).

Table 1. Key VDT performance variables, definitions & metrics based on hierarchical model (Berg, 2006)

Variable	Definition	Metric
Operational Level: lower level skills for vehicle maneuverability		
Off-Road/Route	Driving off-road or off-route: e.g. inability to keep vehicle on-roadway.	At least 1 unique occurrence of driving off-road or off-route.
Tactical Level: skills for traffic and hazard management		
Speed	Median of driver mean velocity in miles per hour.	n/a.
Speed Ratio	Median ratio of driver velocity to posted speed limit.	n/a.
Red Light Error	Driving through red traffic lights without coming to a complete stop (velocity >0)	At least 1 occurrence of red light running
Stop Sign Error	Driving through stop signs without coming to a complete stop (velocity >0)	At least 1 occurrence of rolling stop
Traffic Collision	Any overlap between the driver's vehicle body with either the body of another vehicle or a pedestrian.	At least 1 occurrence of a collision.
Environmental Collision	Any overlap between the driver's vehicle body with a stationary environmental object: e.g., sign pole, curb, etc.	At least 1 occurrence of a collision.

Statistical Tests. The researchers received a de-identified dataset of 2,500 cases. Of these, 2,368 applicants (94.7%) completed the entire VDT workflow. With this sample, the analytical plan was replicated from the previous study. (Walshe et al., 2018) Descriptive analyses were conducted for ORE outcomes and VDT variables of interest. Chi square tests and Wilcoxon rank-sum tests, respectively, compared frequency distributions and continuous variables with ORE pass/fail. In order to avoid a Type 1 error, a Bonferroni-corrected alpha value was used (**α**=.004). All analyses were conducted using R (https://www.r-project.org/).

RESULTS

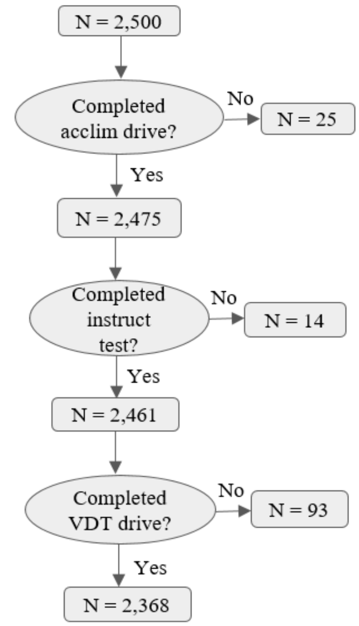

Figure 3. Derivation of sample.

Of the 2,500 total cases, 679 (27.2%) failed the ORE, which is similar to the OBMV's overall ORE fail rate. Nearly all (94.7%) applicants presenting at the OBMV completed the VDT immediately prior to the ORE (see Figure 3). User feedback was positive—"I understood the directions provided in the simulator", "I felt comfortable with the driving controls", and "the simulated scenarios were a reasonable representation of what I see on the road" were all rated $\geq$3.4 (of a possible 5). OBMV staff reported that less than 2% of applicants who took the VDT reported symptoms of simulator sickness.

Tables 2 and 3 illustrate the workflow metrics and performance variables frequencies, respectively, in relation to those who passed and those who failed the ORE. Overall, failure to complete each step of the VDT workflow was significantly associated with failing the ORE. Furthermore, while the overall minority of drivers committed performance errors on the VDT, a greater frequency of driving errors were committed among those who failed the ORE compared to those who passed the ORE (all p<0.001). Specifically, the proportion of applicants with traffic collisions, off-road/route errors, stop sign errors, and red light errors was significantly greater among those who failed the ORE than those who passed (all p<0.001, see Table 3). Environmental collisions did not differentiate drivers regarding ORE pass/fail (p=0.38). Speed metrics differentiate drivers (p<0.001), but the numerical difference was small.

Table 2. Workflow metrics with p-values for comparisons between those who passed and failed the ORE for the replication and original validation samples

	Replication Sample				Original Sample			
Workflow Metric	**All Applicants** (n=2500)	**Passed ORE** (n=1821)	**Failed** (n=679)	**P-value**	**All Applicants** (n=2143)	**Passed ORE** (n=1576)	**Failed ORE** (n=567)	**P-value**
Did not complete acclimation drive	25 (1.0%)	7 (0.4%)	18 (2.7%)	**<0.0001**	33 (1.5%)	19 (1.2%)	14 (2.5%)	0.03
Did not pass instructions test▪	98 (3.9%)	31 (1.7%)	67 (9.9.%)	**<0.0001**	109 (5.2%)	42 (2.7%)	67 (12.3%)	**<0.0001**
Did not complete entire VDT workflow▪	93 (3.7%)	54 (3.0%)	39 (5.7%)	**0.0009**	162 (7.6%)	87 (5.5%)	75 (13.2%)	**<0.0001**

▪ *Calculated on those who completed acclimation drive (n=2461 for replication; n=2093 for original); α= 0.004.*
Note: p-values in bold are statistically significant. Subject matter experts provided information on why some applicants could not complete either the introductory drive or the assessment drive: driver had a language barrier; driver did not understand the instructions; driver was frustrated with the VDT software; driver was called for ORE earlier than expected; driver walked away from the workstation; driver experienced symptoms of simulator sickness (<2%) and ceased..

Table 3. VDT performance variables with p-values for comparisons between those who passed and failed the ORE for the replication and original validation samples

	Replication Sample				Original Sample			
Performance Variable	**All Applicants** (n=2368)	**Passed ORE** (n=1755)	**Failed ORE** (n=613)	**P-value**	**All Applicants** (n=1981)	**Passed ORE** (n=1489)	**Failed ORE** (n=492)	**P-value**
Off-Road/Route⁺	269 (11.4%)	146 (8.3%)	123 (20.1%)	**<0.0001**	315 (15.9%)	189 (12.7%)	126 (25.6%)	**<0.0001**
Speed⁺	18.6 mph	18.7 mph	18.1 mph	**<0.0001**	18.3 mph	18.4 mph	17.9 mph	**0.0005**
Speed Ratio⁺	0.52	0.53	0.51	**<0.0001**	0.51	0.51	0.50	**0.0004**
Red Light Error⁺	405 (17.1%)	254 (14.5%)	151 (24.6%)	**<0.0001**	430 (21.7%)	293 (19.7%)	137 (27.8%)	**<0.001**
Stop Sign Error⁺	347 (14.7%)	229 (13.0%)	118 (19.2%)	**0.0002**	338 (17.1%)	218 (14.6%)	120 (24.4%)	**<0.0001**
Traffic Collision⁺	807 (34.1%)	567 (32.3%)	240 (39.2%)	**0.002**	689 (34.8%)	480 (32.2%)	209 (42.5%)	**<0.0001**
Environmental Collision⁺	655 (27.7%)	471 (26.8%)	184 (30.0%)	0.14	614 (31.0%)	452 (30.4%)	162 (32.9%)	0.38

⁺ *Calculated on those who completed VDT assessment drive (n=2368 for replication; 1981 for original);* α*= 0.004.*
Note: p-values in bold are statistically significant at the corrected alpha level.

CONCLUSION

This study replicated previous validation study results of the VDT (Walshe et al., 2018) in a new sample of 2,500 applicants. Consistent with the previous study, the current sample demonstrated positive user acceptability ratings, and almost all completed the assessment drive. In the replication sample, a smaller number of applicants were unable to complete the VDT and inability to complete the VDT remained strongly associated with failing the ORE. Further, the VDT performance variables continued to differentiate applicants regarding ORE result. Drivers who failed the ORE showed errors in i) controlling the simulated vehicle (go off-road/off-route), ii) adhering to traffic light intersections (stop sign/red light errors) and iii) avoiding traffic collisions (with vehicles or pedestrians). As in the previous study, environmental collisions did not differentiate drivers, and although there was a significant difference in the speed metrics, the difference is too small to interpret meaningfully. This suggests a need for more sensitive metrics.

Replication is a key tool to verify facts for empirical studies (Schmidt, 2009). Especially in situations in which a new tool will be used to make important decisions (as in this case, approval to proceed to an ORE), it is essential to confirm that the results are correct and reliable, remaining consistent when the same methods are applied to a different sample. This study took another step towards VDT validation by repeating previous methods to validate the VDT within a new sample. Thus far, these results confirm that the VDT is reliable in differentiating drivers according to ORE results. It is also consistent with research on novice drivers that a large percentage of drivers experienced errors, including simulated collisions: compared to experienced drivers, novice drivers at the time of licensure display deficits in skills needed for safe driving (McKnight & McKnight, 2003; Gershon et al, 2018). Further replication studies for the VDT will be needed in new sites (rural, as well as urban/suburban) to ensure the generalizability of these results.

The study was limited by two key factors (as in the previous study). First, this study lacked demographic information due to restrictions placed by OBMV. Age and sex are known risk factors for crashes among novice drivers (Bingham & Ehsani, 2012), and may explain some variability in performances among new drivers seeking licensure. Future work will link VDT and ORE outcomes to licensing data records to add demographic variables to the dataset. Second, this analysis used a small number of validated variables that do not comprehensively capture all safety-critical driving skills previously identified in the literature. The VDT collects time series data which can be used to create additional variables (e.g., headway time, lane deviation, etc.); however, these variables involve post-processing and were not validated at the time of this analysis. The addition of these variables will allow us to go beyond comparisons of frequency distributions and assess comprehensive regression models to identify which driving performance skills are predictive of ORE outcomes. Future planned analyses will examine the predictive ability of the VDT not only for ORE results but also for crashes in their first months of licensed driving (Curry et al, 2015; Gershon et al, 2018). For this we are developing Machine Learning modeling methods (Grethlein et al, under review) which includes exploring clusters of performance variables/errors.

ACKNOWLEDGMENTS

The authors would like to thank the Ohio Department of Public Safety and OBMV; the drivers who participated; and Dr. John Bolte for his independent review of this manuscript. Flaura K. Winston and Venkatesh Kandadai disclose their intellectual property and financial interest in Diagnostic Driving, Inc. and The Children's Hospital of Philadelphia's institutional interest in Diagnostic Driving, Inc. Dr. Winston, serves as the CSA and Mr. Kandadai serves as the CEO of Diagnostic Driving, Inc. Dr. Winston operates under a Conflict Management Plan from The Children's Hospital of Philadelphia and the University of Pennsylvania whereby she has no interaction with applicants (all study procedures were carried out by Ohio BMV personnel) and all analyses were reviewed and approved by an outside consultant with no intellectual or financial interest (John Bolte).

REFERENCES

Bingham, C. R., & Ehsani, J. P. (2012). The relative odds of involvement in seven crash configurations by driver age and sex. Journal of Adolescent Health, 51(5), 484-490.

Berg, H. Y. (2006). Reducing crashes and injuries among young drivers: what kind of prevention should we be focusing on?. Injury Prevention, 12(suppl 1), i15-i18.

Chan, E., Pradhan, A. K., Pollatsek, A., Knodler, M. A., & Fisher, D. L. (2010). Are driving simulators effective tools for evaluating novice drivers' hazard anticipation, speed management, and attention maintenance skills?. Transportation research part F: traffic psychology and behaviour, 13(5), 343-353.

Curry, A. E., et al (2011). Prevalence of teen driver errors leading to serious motor vehicle crashes. Accident Analysis & Prevention, 43(4), 1285-1290.

Curry, A. E., et al (2015). Young driver crash rates by licensing age, driving experience, and license phase. Accident Analysis & Prevention, 80, 243-250.

Gershon, P., et al (2018). Crash risk and risky driving behavior among adolescents during learner and independent driving periods. Journal of Adolescent Health, 63(5), 568-574.

Grethlein, D., Winston, F. K., Walshe, E., Tanner, S., Kandadai, V., & Ontañón, S. (In Press). Novel Machine Learning Method for Prediction Using Times Series Data: Initial Application to Prediction of On Road Exam Outcomes from Virtual Driving Test Data. Journal of Medical Internet Research.

Hatakka, M., Keskinen, E., Gregersen, N. P., Glad, A., & Hernetkoski, K. (2002). From control of the vehicle to personal self-control; broadening the perspectives to driver education. Transportation Research Part F: Traffic Psychology and Behaviour, 5(3), 201-215.

Horswill, M., Hill, A., & Wetton, M. (2015). Can a video-based hazard perception test used for driver licensing predict crash involvement?. Accident Analysis & Prevention, 82, 213-219.

Insurance Institute for Highway Safety Highway Loss Data Institute. *GDL requirements by state*. www.iihs.org/iihs/topics/laws/GDLstatelaws?topicName=teenagers. Accessed July 26, 2018.

McDonald, C. C., Curry, A. E., Kandadai, V., Sommers, M. S., & Winston, F. K. (2014). Comparison of teen and adult driver crash scenarios in a nationally representative sample of serious crashes. Accident Analysis & Prevention, 72, 302-308.

McDonald, C. C., Kandadai, V., Loeb, H., Seacrist, T. S., Lee, Y. C., Winston, Z., & Winston, F. K. (2015). Simulated Driving Assessment (SDA) for teen drivers: results from a validation study. Injury prevention, 21(3), 145-152.

McDonald, C., et al. (2012). Using crash data to develop simulator scenarios for assessing novice driver performance. Transportation Research Record 2321, 73-78.

McKnight, A. J., & McKnight, A. S. (2003). Young novice drivers: careless or clueless?. Accident Analysis & Prevention, 35(6), 921-925.

National Motor Vehicle Crash Causation Survey: Report to Congress. (2008) *Publication DOT HS 811 059*. NHTSA, U.S. Department of Transportation.

Ohio Department of Public Safety. *Ohio Highway Safety Plan 2018*. A Safer Ohio. www.nhtsa.gov/sites/nhtsa.dot.gov/documents/ohio_fy2018_hsp.pdf. Accessed July 26, 2018.

Ready-Assess™, Diagnostic Driving Inc. http://www.diagnosticdriving.com/

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, 13(2), 90.

Wells, P., Tong, S., Sexton, B., Grayson, G., & Jones, E. (2008). Cohort II: A study of learner and new drivers. Vol. 1, Main Report. Road Safety Research Report No. 81. Department for Transport. UK, London.

Walshe, E.A., N. M. Oppenheimer, V. Kandadai, S. Ontañón, & F. K. Winston. (2018). Virtual Driving Test Performance Skills among New Drivers who Pass and Fail the On-Road Examination for Licensure. T*ransportation Research Board 98th Annual Meeting Compendium of Papers,* Paper No 19-01210; Washington, DC, 2019.