

PREDICTING OLDER ADULTS' ON-ROAD DRIVING PERFORMANCE

Rachel W. Ross, Charles Scialfa, Scheila Cordazzo & Katherine Bubric
University of Calgary
Calgary, Alberta, Canada
Email: scialfa@ucalgary.ca

Summary: We examined the predictive utility of the Roadwise Review and the Hazard Perception Test on a standardized driving assessment using both conventional and alternative scoring criteria in a sample of health older adults ($N = 57$). Our results indicate that both tests can predict passing or failing the road evaluation. The Hazard Perception Test was more consistent in predicting total points and hazardous errors in on-road performance. Future research should examine the predictive validity of these tests in cognitively impaired drivers.

INTRODUCTION

Older drivers have one of the highest rates of collision per distance driven, despite the fact that they tend to drive less and in lower-risk situations (see Evans, 2004 for a review). Therefore, it is important to accurately predict driver safety in this age group. However, there is some debate over how the fitness of older drivers should be evaluated (Dobbs, Heller, & Schopflocher, 1998), and there have been inconsistencies in the methods and tools used to assess driving fitness in older adults.

Some people view on-road evaluations as a “gold standard” in assessing driving fitness in older adults. However, Dobbs et al. (1998) argue that older adults should not be evaluated on the same criteria as newly licensed drivers. They compared a group of younger drivers and a group of older “normal” drivers to that of a large sample of cognitively impaired older drivers. They found that the cognitively impaired group made more hazardous errors than the healthy older group or the younger group and that the two control groups did not differ from each other. The authors argue that the conventional scoring procedures used for licensure in North America do not differentiate between drivers that pose a hazard to themselves and others, and drivers who commit errors that are common to experienced drivers. Because of this, they found that drivers who were rated as safe by the examiner still failed the on-road evaluation because they accumulated too many points. Consequently, Dobbs and colleagues (1998) argued that new, empirically based scoring schemes are needed to meet the challenges of identifying at-risk drivers.

Hazard Perception Tests (HPTs) are empirically based tests that have recently been developed as screening tools for driving assessment. There are various versions of HPTs (e.g., video based, still image based), but all of them are designed to measure one’s ability to detect and respond to hazards in the driving environment using reaction time. These tests have been found to positively correlate with retrospective crash involvement in various adult samples (Darby, Murray, Raeside, 2009; McKenna & Horswill, 1999) as well as in samples of older adults (Horswill, Anstey, Hatherly, & Wood, 2010; Wells, 2008). They have also been found to successfully distinguish between novice and experienced drivers (Scialfa, Deschênes, Ference, Boone, Horswill, &

Wetton, 2011). More recently, HPTs were found to predict on-road performance in older adults (Wood, Horswill, Lacherez, & Anstey, 2013). However, this study did not use conventional scoring criteria for on-road examinations. Rather, Wood and colleagues (2013) used evaluators' ratings on a scale of 1-10, with a cut-off of 5 to determine whether participants were likely pass a standardized driving exam.

The DrivingHealth[®] Inventory (DHI) is another empirically based measure that was developed and marketed as a comprehensive and standardized approach to assessing driving fitness. The DHI was developed through an extensive study of a battery of different tests, with the goal of identifying those measures that could be used to predict past collision involvement in older adults (Staplin, Gish, & Wagner, 2003). It has since been adopted as a comprehensive and standardized approach to testing driving fitness in older adults (Staplin et al., 2003). Furthermore, the American Automobile Association and the Canadian Automobile Association commercially distribute this assessment tool as the Roadwise Review, which assesses physical, visual, and cognitive abilities that are related to safe driving.

However, the DrivingHealth[®] Inventory has shown inconsistent predictive utility. Edwards et al. (2008) found that older adults performed significantly worse than younger adults on the DHI, and older adults with past collisions performed worse than older adults without involvement in a collision in the previous 2 years. On the other hand, Scialfa and colleagues (Scialfa, Ference, Boone, Tay, & Hudson, 2010) found that the Roadwise Review was unable to predict self-reported driving issues or retrospective collision involvement in older adults. Similarly, Bédard and colleagues (Bédard, Riendeau, Weaver, & Clarkson, 2011) found that the Roadwise Review has limited congruence with on-road evaluations of driving ability.

The goal of Bédard and colleagues' (2011) study was to test the predictive validity of the Roadwise Review in on-road performance using conventional scoring criteria (i.e., accumulated points and pass/fail). However, based on findings that conventional scoring criteria on North American driving assessments do not discriminate between common errors committed by experienced drivers, and hazardous errors that are indicative of declines in driving fitness (Dobbs et al., 1998), it may be more appropriate to determine whether the Roadwise Review can predict outcomes that are more directly related to safe driving, such as hazardous safety errors.

The goal of the current study, therefore, is to assess the predictive utility of both the Roadwise Review and the Hazard Perception Test using conventional scoring methods (i.e., accumulated points and pass/fail) as well as using an alternative scoring method (i.e., number of serious safety errors committed during an on-road driving evaluation, calculated by tabulating a total score for the number of automatic disqualifications the driver received). By comparing the two types of evaluation criteria, we may be better able to identify measures that distinguish between older adults who make errors common to experienced drivers and older adults who make serious safety errors that may put themselves or other road users at risk.

METHODS

Sixty-four individuals who hold a valid operator’s license were recruited from community organizations. In addition to a valid operator’s license, participants were required to have normal (corrected or uncorrected) visual acuity, and a Mini Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975) score of 1 (broadly normal). Of the 64 participants, 6 were excluded from the analysis because they showed impairment on the MMSE, and 1 individual was excluded because he was under the age of 55. Sixty-eight percent of participants drove between 5,000 and 20,000 kilometers per year. All participants received \$40 for their involvement in the study. Summary demographic data are presented in Table 1. The Conjoint Faculties Research Ethics Board (CFREB) of the University of Calgary approved this study (File #7207).

Table 1. Demographic Data

	Mean	SD
Age	75.45	7.81
Education in years (starting with grade 1)	16.21	3.40
Self Reported Health Ratings (Scale of 1 to 5)	4.23	0.65
MMSE	28.56	1.60

The current study was completed in two sessions. The first session took place in the laboratory and lasted roughly two hours. During this time, participants provided demographic information such as age, distance driven, and health concerns.

Next, we tested participants’ vision. Their visual acuity was assessed with a Post- script generated Landolt C chart, contrast sensitivity was assessed with the VISTECH 6500, and color vision was measured with the Farnsworth D-15 Color Test (Farnsworth, 1943). All vision testing was carried out within recommended photopic luminance levels.

Participants then completed the Mini Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975). Although the predictive validity of the MMSE is debatable, it is one of the most frequently used assessment tools to screen for cognitive impairment (e.g., Margolis, Kerani, McGovern, Songer, Cauley, & Ensrud, 2002; Owsley, Sloane, Ball, Roenker, & Bruni, 1991).

After the MMSE, participants completed the commercially available version of the Roadwise Review. It is comprised of eight tests that appear in the following fixed order; leg strength and general mobility, head and neck flexibility, high contrast visual acuity, low contrast visual acuity, visualizing missing information, processing speed (UFOV), visual search, and working memory.

Finally, participants completed the Hazard Perception test (HPT; Scialfa et. al, 2011). The HPT consists of a series of 26 silent driving scenes that last between 16 and 62 seconds. Seventeen of those scenes contained a traffic conflict that required an evasive action such as slowing, stopping, or swerving to avoid a collision with a road user or stationary object. The remaining 9 scenes did not contain a traffic conflict. Theses scenes were included to moderate participants’ criterion for responding, because a test containing only traffic conflicts might bias participants to respond to improbable conflicts. Participants were instructed to identify the presence of a traffic

conflict that would lead to a collision between the “camera” vehicle and another road user if an evasive action was not taken. Performance was assessed using accuracy and reaction time on trials ending in a correctly identified hazard.

On-Road Assessment: During the second session, participants were given an on-road driving evaluation. The route that was designed with the assistance of trained driving staff from the Alberta Motor Association (AMA) in order to ensure that it closely resembled a route that would be used for the provincial driving exam. It took participants through a mixed residential and commercial area in northwest Calgary, was 18 km long, and took between 35 and 45 minutes to drive. Performance was monitored and evaluated by a trained evaluator in a hybrid Toyota Camry with a second brake pedal and extra mirrors. Scoring was initially identical to that used for a first Class 5 license in Alberta. The standard Class 5 Examination Checklist includes evaluation of controls (e.g., knowledge and use of equipment), parking (e.g., failing to set the parking brake), lane driving (e.g., failing to check mirror), intersections (e.g., blocking the crosswalk), traffic lights (e.g., failing to anticipate or observe), right of way (e.g., uncertain, hesitant) speed (e.g., too fast for conditions) and automatic disqualifications (e.g., right of way violation for a vehicle or pedestrian). As an additional outcome measure we also tabulated a total score for the number of automatic disqualifications the driver received. Automatic disqualifications result from serious safety errors such as stop sign violations and obstructing traffic.

RESULTS

Neither age, nor simple spatial reaction time accounted for any unique variance and therefore were not included in the following analyses. Descriptive statistics for the Roadwise Review, HPT, and on-road evaluation are presented in Table 2. Zero-order correlations between the Roadwise Review, the HPT and performance indicators on the driving exam are presented in Table 3.

Table 2. Descriptive statistics for Roadwise Review, HPT, and On Road Evaluation

Test Items	<i>M</i>	<i>SD</i>
Roadwise Review		
Leg Strength & General Mobility	6.16	1.38
Head/Neck Flexibility	1.52	.50
High Contrast Visual Acuity	1.02	.13
Low Contrast Visual Acuity	1.13	.38
Visualizing Missing Information	2.75	2.09
Visualizing Missing Information Processing Speed	187.43	100.54
Visual Search	114.45	44.82
Working Memory	.26	.61
HPT		
Reaction Time	3.39	1.10
On Road Evaluation		
Total Points	75.26	31.80
Total Number of Disqualifications	1.65	2.01

Using a linear regression analysis, the Roadwise Review was not a significant predictor of the number of hazardous errors committed, $F(8, 43) = 1.71, p = .12$, or of accumulated points on the driving evaluation, $F(8, 43) = .84, p = .57$. However, using binary logistic regression, the Roadwise Review significantly improved the classification rate of pass or fail on the driving exam., $\Delta\chi^2(8) = 18.0, p = .02$. Sensitivity was 73.3% and specificity was 77.3%.

Table 3. Zero order correlations between the Roadwise Review, HPT and on-road evaluation

	HPT Reaction Time	On-Road Total Points	Total Hazardous Errors
Roadwise Review			
Leg Strength & General Mobility	.29*	.16	.20
Head & Neck Flexibility	.09	-.06	-.03
High Contrast Visual Acuity	-.01	-.09	-.11
Low Contrast Visual Acuity	.27*	-.12	.02
Visualizing Missing Information	.31*	.14	.26*
Visualizing Missing Information Processing Speed	.34*	.04	.21
Visual Search	.32*	.19	.23
Working Memory	.19	.17	.29*
Hazard Perception Test			
Reaction Time	-----	.25	.42**

* Significant at $< .05$ level, ** Significant at $< .001$ level.

Reaction time (RT) to traffic conflict scenes on the HPT proved to be a significant predictor of the number of hazardous errors committed, $F(1, 54) = 11.78, p = .001, R^2 = .18$, and a marginally significant predictor of the number of points accumulated, $F(1, 54) = 3.48, p = .068, R^2 = .06$. Using logistic regression, the HPT improved classification accuracy, $\Delta\chi^2(1) = 10.09, p = .001, \beta = .92, W(1) = 7.77, p = .005$. For every one unit increase in RT a participant was 2.52 times more likely to be classified as failing the on-road assessment. Sensitivity was 65.6% and specificity was 62.5%.

DISCUSSION

Identifying at-risk older drivers is a goal for researchers and health practitioners, alike. Two of the issues that complicate this goal are identifying the appropriate evaluation criteria, and a lack of empirically based scoring schemes that can accurately predict who is safe to drive and who is not.

Our results contribute to the literature on driving safety in a number of ways. First, we found that the Roadwise Review demonstrated inconsistent congruence with on-road driving performance. Our results replicated those of Bédard and colleagues (2011), in that the Roadwise Review was not a significant predictor of total points or hazardous safety errors. However, in contrast to their work, the Roadwise Review could predict with reasonable accuracy whether a driver would pass or fail the on-road assessment.

The HPT demonstrated modest success at predicting performance on the driving exam using the conventional scoring criteria. These results should nonetheless be encouraging for researchers interested in driving safety because the on-road evaluation is a global assessment of both hazardous safety errors and a broad range of other errors that are less likely to pose a threat to driving safety (e.g., knowledge or use of equipment, grinding gears, hitting the curb, and improper use of the clutch). Therefore, the HPT – as a measure of one’s ability to detect hazards – is not likely going to be associated with many areas of assessment on a standard driving exam. Despite this limitation, the HPT was a significant predictor of driving performance using conventional scoring criteria of pass or fail, and a marginally significant predictor of total points.

Using alternative scoring criteria, the HPT accounted for 18% of the total variance in hazardous safety errors in our sample of healthy older adults. We view this as a particularly significant contribution to driving safety research, as the goal is not to identify individuals who commit errors that are common to experienced drivers, but rather to identify individuals who commit errors that pose a risk to themselves and to other road users.

This study is a notable step toward identifying measures that can be used to assess different aspects of driving performance. To date, no one particular measure has been successful at predicting fitness to drive. However, consistent with previous literature (Wood et. al., 2013) our results suggest that the HPT could be used in combination with other measures to accurately predict who is safe to drive and who is not. Some of the benefits of the HPT are that it is easy to administer and it only takes 15 minutes to complete.

It is important to note that our results are based on a sample of relatively healthy older adults, which may limit the predictive utility of both the HPT and the Roadwise Review. Future research should examine the predictive accuracy of the HPT and the Roadwise Review on driving performance in a sample of cognitively impaired older adults.

ACKNOWLEDGEMENTS

This research was funded by a grant from the Alberta Motor Association. We would like to express our sincere gratitude to the staff of the Alberta Motor Association, whose efforts were indispensable in the data collection phase of this study.

REFERENCES

- Ball, K., Roenker, D., Wadley, V., Edwards, J., Roth, D., McGwin, G., Raleigh, R., Joyce, J., Cissell, G., & Dube, T. (2006). Can high-risk older drivers be identified through performance based measures in a Department of Motor Vehicles setting? *Journal of the American Geriatrics Society*, 54, 77–84.
- Bédard, M., Riendeau, J., Weaver, B., & Clarkson, A. (2011). Roadwise Review has limited congruence with actual driving performance of aging drivers. *Accident Analysis and Prevention*, 43, 2209-2214.
- Darby, P., Murray, W., Raeside, R. (2009). Applying online fleet driver assessment to help identify, target and reduce occupational road safety risks. *Safety Science*, 47, 436- 442.

- Dobbs, A. R., Heller, R. B. & Schopflocher, D. (1998). A comparative approach to identifying unsafe older drivers. *Accident Analysis and Prevention*, 30(3), 363-370.
- Edwards, J.D., Leonard, K.M., Lunsman, M., Dodson, J., Bradley, S., Myers, C.A., & Hubble, B. (2008). Acceptability and validity of older driver screening with the DrivingHealth[®] Inventory. *Accident Analysis and Prevention*, 40, 1157-1163.
- Evans, L. (2004). *Traffic safety*. Bloomfield Hills, MI: Science Serving Society.
- Farnsworth, D., (1943). Farnsworth 100 Hue and Dichotomous tests for color vision. *Journal of the Optical Society of America*, 33(10), 568–578.
- Folstein, M., Folstein, S., & McHugh, P. (1975). Mini-Mental State: A practical method for grading the cognitive status of patients for the clinician. *Journal of Psychiatric Research*, 12, 189–198.
- Horswill, M.S., Anstey, K.J., Hatherly, C.G., Wood, J., (2010). The crash involvement of older drivers is associated with their hazard perception latencies. *Journal of the International Neuropsychological Society*, 16(5), 939-944.
- Margolis, K. L., Kerani, R. P., McGovern, P., Songer, T., Cauley, J. A., Ensrud, K. E. (2002). Risk factors for motor vehicle collisions in older women. *Journal of Gerontology: Medical Sciences*, 57, M186–M191.
- McKenna, F.P., Horswill, M.S., (1999). Hazard perception and its relevance for driver licensing. *Journal of the International Association of Traffic and Safety Sciences*, 23, 26-41.
- Owsley, C., Sloane, M. E., Ball, K., Roenker, D. L., & Bruni, J. R. (1991). Visual – Cognitive correlates of vehicle collisions in older drivers. *Psychology and Aging*, 6, 403–415.
- Reeves, B., Wood, J., Hill, A. (1991). Vistech VCTS 6500 charts – within – and between – session reliability. *Optometry and Vision Science*, 68, 728–737.
- Scialfa, C. T., Deschênes, M. C., FERENCE, J., Boone, J., Horswill, M. S., & Wetton, M. (2011). A hazard perception test for novice drivers. *Accident Analysis and Prevention*, 43, 204-208.
- Scialfa, C., FERENCE, J., Boone, J., Tay, R., & Hudson, C. (2010). Predicting older adults' driving difficulties using the Roadwise Review. *Journal of Gerontology: Psychological Sciences*, 65B(4), 434–437, doi:10.1093/geronb/gbq032.
- Scialfa, C.T., Adams, E., Giovanetto, M. (1991). Reliability of the Vistech Contrast Test System in a life-span sample. *Optometry and Vision Science*, 68, 270–274.
- Staplin, L, Gish, K. W., Wagner, E. K. (2003). MaryPODS revisited: Updated crash analysis and implications for screening program implementation. *Journal of Safety Research*, 34, 389-397.
- Wood, J. M., Horswill, M. S., Lacherez, P. F., Anstey, K. J. (2013). Evaluation of screening tests for predicting older driver performance and safety assessed by an on-road test. *Accident Analysis and Prevention*, 50, 1161-1168.