

# Driving Skills Training for Older Adults: An Assessment of DriveSharp\*

Katherine A. Johnston, David Borkenhagen, and Charles T. Scialfa  
*Department of Psychology, University of Calgary*

## RÉSUMÉ

Les procédures de formation cognitive informatique visent à augmenter la sécurité en améliorant les compétences relatives à la conduite, comme la vitesse-de-traitement et le Useful Field of View. L'étude actuelle a évalué l'efficacité du DriveSharp dans la formation des conducteurs âgés dans un cadre de classe réaliste. Les participants ( $n = 24$ ) ont assisté à 10 heures de cours de DriveSharp pendant 5 semaines. Les séances pré- et post-test ont évalués améliorations sur un essai dynamique de la perception du risque, Trails A et Trails B. Un groupe de contrôle ( $n = 18$ ) a terminé seulement les séances pré- et post-test. En classe, les temps de formation étaient plus bas que prévus. L'amélioration des participants aux jeux ont stabilisée après la première évaluation, et le groupe de DriveSharp n'a pas démontré une amélioration significative des performances sur les tests, par rapport au groupe de contrôle. Parmi plusieurs questions relatives à la facilité d'utilisation, les plus problématiques étaient le malentendus objectifs de la tâche et la différence entre la formation et l'évaluation. Il y a plusieurs implications pour ceux qui utilisent DriveSharp pour améliorer la sécurité des conducteurs âgés.

## ABSTRACT

Computer-based, cognitive training procedures aim to increase safety by improving skills related to driving, such as speed-of-processing and the Useful Field of View. The current study assessed the effectiveness of DriveSharp in training older drivers in a naturalistic class setting. Participants ( $n = 24$ ) attended 10 hours of DriveSharp classes over 5 weeks. Pre- and post-testing sessions assessed improvements on a dynamic hazard perception test, Trails A and Trails B. A control group ( $n = 18$ ) completed only pre- and post-testing sessions. In-class training times were lower than expected. Participants' improvement in the games leveled off after the first assessment and the DriveSharp group did not demonstrate a significant improvement in performance compared to the control group. Among several usability issues, the most problematic were misunderstanding task goals and the difference between training and evaluation. There are several implications for those using DriveSharp to enhance older drivers' safety.

\* This research was supported by a grant from the Alberta Traffic Safety Fund. We would like to thank The Kerby Centre of Calgary, Alberta, and Wendy Rodger in particular for her invaluable assistance in gaining research support and coordinating data collection for this project.

Manuscript received: / manuscrit reçu : 13/08/14

Manuscript accepted: / manuscrit accepté : 10/02/15

**Mots clés :** vieillissement, formation cognitive, DriveSharp, compétences de conduite, perception des risques

**Keywords:** aging, cognitive training, DriveSharp, driving skills, hazard perception

La correspondance et les demandes de tirés-à-part doivent être adressées à: / Correspondence and requests for reprints should be sent to:

Charles T. Scialfa, Ph.D.  
 Department of Psychology  
 University of Calgary  
 Calgary, AB T2N 1N4  
 (scialfa@ucalgary.ca)

Older adults are generally safe drivers, but when compared to experienced, middle-aged adults after controlling for distance driven, they have greater collision risk (see Evans, 2004, for a review). Some have asserted that this is a reflection of the "low-mileage bias" (Hakamies-Blomqvist, Raitanen, & O'Neill, 2002) and inadequate controls for exposure (Chipman, MacGregor,

Smiley, & Lee-Gosselin, 1992). Additionally, because of the increases in age-related medical issues, older drivers are more susceptible to injury and fatalities from collisions (Li, Braver, & Chen, 2003). As the population ages and the presence of older drivers increases in absolute and relative terms, this will become a more important and, admittedly, controversial issue.

Currently, a variety of assessment tools are being evaluated for their effectiveness in determining fitness to drive (Dobbs & Schopflocher, 2010; Wood, Horswill, Lacherez, & Anstey, 2013). These tools have relied on identifying factors associated with increased risk of collisions for older adults, including visual abilities, physical strength, and cognitive skill (Ball et al., 1991; Ball et al., 2006; Meuleners, Harding, Lee, & Legge, 2006; Rubin, Ng, Bandeen-Roche, Keyl, Freeman, & West, 2007; Ross, Cordazzo, & Scialfa, 2014). Tests such as Trails A and B, which assess visual search, memory, and attention, have been found to be moderate predictors of collision risk (Ball et al., 2006; Edwards et al., 2008). As well, brief hazard perception tests (HPTs) have shown a robust and reliable association with safe driving (Horswill & McKenna, 2004; McKenna & Crick, 1991; Ross Jones et al., 2014).

Tests of hazard perception assess an individual's ability to detect and respond to a potentially dangerous element in the roadway before a collision occurs. Assessments within various age groups have revealed that, even after controlling for generally slower responses, older drivers are slower than younger drivers at identifying and responding to dangerous elements (Horswill, Anstey, Hatherly, & Wood, 2010; Horswill et al., 2008; Scialfa et al., 2012b). Furthermore, it is a skill that may be improved with training (Chapman, Underwood, & Roberts, 2002; Fisher, Pollatsek, & Pradhan, 2006; Horswill, Kemala, Wetton, Scialfa, & Pachana, 2010) and such improvements lead to improved driving safety (Horswill, Taylor, Newnam, Wetton, & Hill, 2013).

Of particular interest in safe driving is the Useful Field of View (UFOV), a cognitive measure of visual processing speed, which is often assessed under divided attention conditions (Roegner, Cissell, Ball, Wadley, & Edwards, 2003). The UFOV has demonstrated a moderate relationship to collisions in both retrospective and prospective studies, where those with slower speed of processing have greater collision involvement (Ball & Owsley, 1993; Ball, Owsley, Sloane, Roegner, & Bruni, 1993; Goode et al., 1998; Owsley, Ball, Sloane, Roegner, & Bruni, 1991; Like hazard perception, UFOV performance has demonstrated improvements after training (Ball et al., 2002; Willis et al., 2006; Roegner, et al., 2003). As a result, there has been interest in determining if UFOV training in older drivers results in safer driving behaviors and decreased risk of collision.

There also has been some examination of the influence of multiple-object tracking on driving safety among older drivers (Trick, Perl, & Sethi, 2005; Lochner & Trick, 2011). Safe driving requires that one is able to track a number of objects in motion (e.g., vehicles, pedestrians, etc.) while operating the vehicle. However, the number

of targets adults are able to track declines with age (Trick et al., 2005; Sekuler, McLaughlin, & Yotsumoto, 2008). It has been suggested that deficits in multiple-object tracking may be related to the increased collision risk in complex driving scenarios such as left turns at intersections (Trick et al., 2005). Additionally, multiple-object tracking in older drivers has been found to be predictive of driving performance as determined by an on-road evaluation (Bowers et al., 2013).

A number of training programs have been developed in an effort to improve driving fitness and delay driving cessation in older drivers. Classroom-based training such as the American Automobile Association's (AAA) and Canadian Automobile Association's (CAA) Driver 55 Plus target knowledge of changing roadway rules and regulations, driving strategies such as increased shoulder and mirror checks, and information on how aging impacts driving.

In contrast, computer-based programs using cognitive exercises that emphasize speed of processing have also been developed to improve on-road safety. A subset of the sample from the Advanced Cognitive Training for Independent and Vital Elderly (ACTIVE) study received speed-of-processing training (Edwards, Delahunt, & Mahncke, 2009). This training progressed over nine sessions and included tasks such as locating and identifying increasingly demanding stimuli. A follow-up after five years revealed fewer state-reported collisions for the training group than the controls. Other longitudinal assessments of cognitive training found that after three years, only nine per cent of those who completed speed-of-processing training had ceased driving compared to 14 per cent of a control group. Similarly, Roegner et al. (2003) assessed UFOV training by incorporating three computer-based tasks: central stimulus identification; central and peripheral stimulus identification; and central stimulus identification among distractors. Each participant completed a variable amount of training to reach a threshold of 17 ms for the first task and 75 per cent accuracy for the second and third tasks. Post-training assessments and a follow-up at 18 months revealed that training significantly reduced the UFOV impairment and improved reaction time. However, these improvements were not reflected in behavioral measures during on-road evaluations.

Recently, commercial programs for training skills related to safe driving have been made available to consumers. DriveSharp, developed by Posit Science (<https://www.drivesharp.com/>), is a cognitive training program marketed for older adults that focuses on UFOV training, multiple-object tracking, working memory, and divided and selective attention. As support for its effectiveness, specifically the UFOV training, DriveSharp cites a number of studies that demonstrate improvements

in identification and reaction time (Ball, Edwards, Ross, & McGwin, 2010; Owsley et al., 1998; Roenker et al., 2003; Rubin et al., 2007; Sims, McGwin, Allman, Ball, & Owsley, 2000) and a reduction in risk of collisions up to 50 per cent (Ball et al., 2002).

The program involves three activities: Jewel Diver, Road Tour, and Sweep Seeker. Jewel Diver is intended to improve divided attention by training multiple-object tracking. The game progresses by increasing the speed and number of targets, as well as the similarity between the objects and background. Road Tour focuses on expanding the UFOV by using a double-decision task: while fixating on a central object, participants must discriminate two target objects among distractors in the peripheral field. The game progresses by presenting the objects further in the periphery, increasing similarity between targets and distractors, and increasing the number of distractors in the background. The final exercise, Sweep Seeker, is the speed-of-processing training component. Participants are asked to indicate if two sine-wave gratings are oriented in the same direction. The game progresses by decreasing the allowed decision time. Total training time for each game lasts between two to four hours, and it is recommended that individuals complete 20 minutes of training three times a week until all the games have been completed. However, the marketing and instructions for DriveSharp indicate that “dramatic improvements” in driving safety occur after a minimum of 10 hours of training (Posit Science, 2010, p. 5).

The study discussed in this article assessed the usability and effectiveness of the DriveSharp program for improving the skills trained (i.e., UFOV training, speed of processing, divided attention, multiple-object tracking). We also determined if there was practice-based improvement in performance on tests that are associated with driving safety. The program was assessed with a group of older drivers who completed DriveSharp training

for two hours a week over a five-week period in a facilitated environment. Their performance was compared to a control group in pre- and post-training assessments using Trails A and B, and a dynamic HPT.

## Methods

### Participants

A total of 53 older adults with a current driver’s license were recruited through the Kerby Centre, an older adult activity, education, and resource centre. They were offered in-class, computer-based training to assist with safe driving. Thirty-five of the participants were enrolled in the course; of these, five withdrew before commencement. As well, two individuals could not complete the testing, two missed pre-testing sessions, and another two did not complete the post-test session. Thus, a total of 24 participants ( $M_{\text{age}} = 75.29$  yrs,  $SD = 6.65$  yrs) completed both pre- and post-testing sessions in addition to attending the training. The remaining participants volunteered as a part of the control group. Of the 23 who were recruited, five did not complete post-testing. The remaining 18 participants with complete data sets ( $M_{\text{age}} = 68.83$  yrs,  $SD = 6.37$  yrs) were included in the analyses. Demographic information is presented in Table 1.

Although the experimental group was significantly older than controls, the two groups did not differ on other measures. On average, the control group had approximately one more year of formal education. Both groups were in good self-reported physical health, although a minority reported experiencing problems in mental health. Mini-Mental State Exam (MMSE) scores did not differ significantly and indicated that on average they did not demonstrate any obvious signs of cognitive impairment. Their driving history suggested that they were, in a global sense, “safe behind the wheel”.

**Table 1: Experimental and control group demographic information**

	Experimental Group	Control Group	p-value
Age	75.29 (6.65)	68.83 (6.37)	.003*
Gender ratio (M/F)	9/15	10/8	.157
Presence of mental illness (%)	13%	11%	.896
Years of education	13.17 (2.37)	14.28 (2.68)	.163
Collisions within 2 yrs	.25 (0.61)	.11 (0.32)	.385
Moving violations within 2 yrs	.17 (0.38)	.28 (0.56)	.456
Km driven per year	12,479.6 (12,847.1)	12,947.1 (13,588.7)	.913
Physical health rating	4.25 (0.68)	3.94 (.94)	.226
MMSE	28 (2.04)	28.06 (2.1)	.932

\* Indicates significant differences between experimental and control group on that measure ( $p < .05$ ).

MMSE = Mini-Mental State Exam.

Unless otherwise indicated, cell values are arithmetic means (standard deviations).

## Materials and Apparatus

### Hazard Perception Test

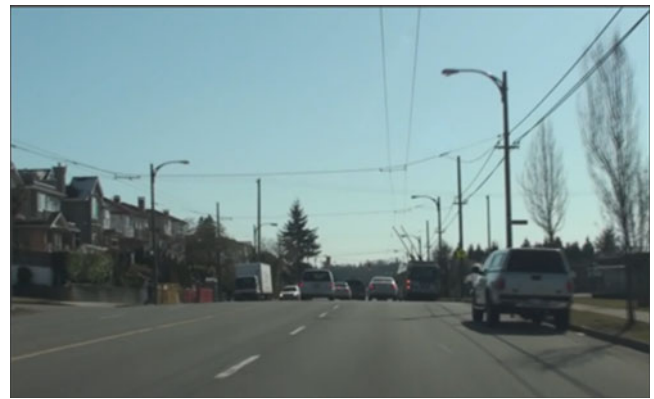
The current version of the Hazard Perception Test uses selected scenes from the dynamic HPT developed by Scialfa et al. (2012b), which has good internal consistency and short-term reliability (Scialfa et al., 2014). The test differentiates novice and experienced young adults (Scialfa et al., 2012b) and experienced young and older adults (Horswill et al., 2010).

Two brief versions of the test lasting approximately 15 minutes each were created. These were used in place of the original HPT, which takes over one hour to complete. A brief version has clear advantages for testing large groups of people. It can be more easily used in a multi-faceted assessment battery without inducing fatigue or taking too much time. This brief version has been shown (Ross et al., 2014) to predict on-road performance in healthy older adults.

Each test comprises a series of 26 silent driving scenes lasting between 10 to 62 seconds filmed in Vancouver, B.C., Canada, and surrounding areas using a Sony Handycam Camcorder, model HDR-SR11 in AVCHD 16M (FH) format at a resolution of 1920 × 1080/60i. The camera was mounted inside a 2005 Subaru Impreza and secured to the inside door window on the passenger side of the vehicle. An extendable arm allowed the videotaped scenes to give a “driver’s eye” view. Filming occurred in March and April, 2009, during daylight hours, generally under clear skies and dry roadway conditions in a variety of frequently encountered environments (e.g., residential, limited-access freeway). Each driving scene was edited from original files using Sony Vegas Movie Studio Platinum software (version 9.0a) at a resolution of 1280 × 720. Only one traffic device found in the scenarios differed from those found in Alberta, a flashing green signal light. Participants were instructed to treat the flashing light as a solid green light. For a detailed description of the types of hazards presented, please see Scialfa et al. (2012a).

Of the 26 driving scenes in each test, 17 (65%) and 18 (69%) respectively contained a *traffic conflict*, defined as a situation in which the camera car was required to take evasive action such as slowing, stopping, or steering to avoid a collision with a road user or stationary object. Examples of the traffic conflicts include a braking lead vehicle, pedestrian incursion, and construction equipment in the driving lane (see Figure 1). The remaining scenes did not contain a traffic conflict and were included to increase uncertainty about hazard presence, as would be the case in normal driving.

At onset of the traffic conflict, the object in the scene had a height ranging between 1 and 10 deg ( $M = 3.0$  deg)



**Figure 1: Static screenshot of a traffic conflict scene from the dynamic hazard perception test (Scialfa et al., 2012b) (resolution is better than depicted)**

and a width between 1.6 and 14.8 deg ( $M = 4.4$  deg) at a nominal viewing distance of 50 cm. The eccentricity of the objects relative to screen centre ranged between  $-0.9$  and  $3.4$  deg on the vertical axis ( $M = 1.0$  deg) and  $-16.2$  and  $10.9$  deg on the horizontal axis ( $M = -1$  deg). Thus, objects in traffic conflicts are quite varied in their size and location but, on average, do not require excellent acuity or peripheral vision.

Custom software defined the onset, offset, and spatial extent of the traffic conflicts of each scene (see Marrington, Horswill, & Wood, 2008). This same software was used to present scenes to participants and record the spatial coordinates of their responses and their reaction times. A 17" LCD desktop monitor with a resolution of 1280 × 1024 set at a viewing distance of approximately 50 cm was used to present the HPT and collect responses. Participants were instructed to identify the traffic conflict as quickly and accurately as possible using the mouse. The average reaction time to traffic conflict scenes was used as the dependent measure because errors were infrequent.

*Trails A and B.* These brief, paper-and-pencil tests require that one draw a line correctly connecting a series of numbers (Trails A) or alternating between numbers and letters (Trails B). The test administrator corrects errors during performance, and the time taken for completion is used as the dependent measure. Trails A and B are commonly used in the neuropsychological literature as a measure of processing speed, working memory, and executive control and are also used in commercially available assessment tools for driver risk such as the Roadwise Review (e.g., Scialfa, Ference, Boone, Tay, & Hudson, 2010). To facilitate comparison with our samples, Tombaugh (2004) reports that average Trails A and Trails B completion times for those aged 70–74 are approximately 42 and 109 sec respectively.

### DriveSharp

Participants completed the DriveSharp training sessions in a classroom environment on individual desktop computers. During the first session, the program included a brief tutorial for each exercise and prompted participants to complete an initial baseline assessment. After this, the program automatically prompted them to complete further assessments when a variable number of trials had been completed, unless a participant actively selected an assessment to be done. Each of the games progressed based on the improvement of the individual. The algorithms used to determine progress are not freely available, but Dobres et al. (2013) have reported that the programs use a staircase to adjust the difficulty of the task, converging on a 70.7 per cent correct performance criterion. There are three exercises (see Figure 2).

*Jewel Diver* incorporates divided attention and multiple-object tracking. Participants are shown a number of targets (jewels), which are then masked (by bubbles). The goal is to track the bubbles as they move around the screen and identify which occlude the target jewels. Similarly, the assessment requires that participants track jewels occluded by fish. Improvement scores are based on the average number of jewels correctly tracked.

*Road Tour* is intended to enlarge the UFOV. Participants focus on the centre of the screen and are briefly

shown a target vehicle. They must then identify which vehicle they saw at the centre of the screen as well as indicate the location of a secondary target (a traffic sign) in their periphery. Improvement is operationalized as the average exposure duration of the target stimuli when it is correctly identified, in conjunction with the correct localization of the peripheral stimuli.

*Sweep Seeker* trains speed of processing by asking participants to identify the orientation of sine-wave gratings that vary in contrast and spatial frequency. During the game, participants are shown a series of tiles. They must select three matching tiles to clear the tiles from the screen. After clearing a selection of tiles, a window that displays two gratings is shown. Participants are asked to identify the movement of the gratings (i.e., inward or outward optic flow). Responses are timed, and a score is obtained from the speed with which they determine the orientation. After completing the task, they are returned to the original “tile-breaking” task. The goal is to clear the screen of all the tiles. The assessment requires that participants only determine the orientation of the gratings. Improvements are tracked by averaging the exposure duration of iterations where the participant successfully indicated the direction of the sine-wave gratings.



**Figure 2:** A. Screenshot of the DriveSharp Jewel Diver activity (on-screen resolution is better than depicted). B. Screenshot of the DriveSharp Road Tour activity (on-screen resolution is better than depicted). C. Screenshot of the DriveSharp Sweep Seeker activity (on-screen resolution is better than depicted)

### Usability Assessment

The primary outcome variables in this study were those related to the DriveSharp exercises (e.g., time spent on them and performance improvement), in addition to response times for Trails A, Trails B, and the Hazard Perception Test. While this information provides insight into the behavioral changes that occurred as a result of the DriveSharp training modules, it does not assess participants' general opinions and beliefs regarding the usability of the program. Effective adoption of a technology-based training program requires that the target audience is able to use the technology. People are more likely to discontinue a less usable training program before they reach the recommended training times, and, as a result, any potential improvements to driving would be reduced. Additionally, ease of use is especially important for a program directed towards older adults, who may have lower computer literacy (Nielsen, 2013).

To gather usability data, participants completed a brief, 20-statement survey (See Table 2). The statements were either positive or negative remarks about the program, training routine, or computers in general. A 5-point Likert scale, with 1 representing "strongly disagree" and 5 representing "strongly agree" was used.

### Procedure

In addition to the training described in the discussion herein, there were two experimental sessions lasting approximately one hour each. During the pre-testing

session, participants completed a demographic questionnaire assessing collision history, annual distance driven, and information regarding any medical issues that may affect their driving. Additionally, the MMSE (Folstein, Folstein, & McHugh, 1975) was administered to assess their cognitive status. Participants also completed Trails A, Trails B, and an HPT.

For the HPT, to ensure that all participants were familiar with using a computer mouse, a practice screen was displayed. The practice screen contained a series of 10 targets; participants were asked to select each target using the mouse. After familiarization, they were instructed on how to complete the hazard perception test. They were given a short practice trial of eight scenes to familiarize themselves with the test requirements. Traffic conflicts similar to those occurring during the experimental trials were used.

The DriveSharp course lasted five weeks, during which participants were scheduled to attend two one-hour sessions a week for a total of 10 hours of training. Sessions were led by a facilitator who assisted with the initial introduction to the program and any difficulties operating the computer. Participants were instructed to follow the recommended schedule for the DriveSharp software as much as possible, however, as the sessions were longer than the recommended 20 min, they were free to select the exercises on which they would focus. As such, there was considerable variability in the exercises they chose to complete.

After completing the five weeks of training, participants attended a post-testing session where they completed a usability questionnaire, Trails A and B, and

**Table 2: Responses to usability questions regarding the DriveSharp training software**

Statement	Mean	Median	Range
I found the system unnecessarily complex	2.14	2	1-5
I understand how this will improve my driving skills	3.96	4	1-5
I thought the system was easy to use	4.11	4	2-5
I think that I would need the support of a technical person to be able to use this system	2.61	2.5	1-5
I would buy this program	2.57	2.5	1-5
I would imagine that most people would learn to use this system very quickly	3.75	4	1-5
I would recommend this program to a friend	3.79	4	1-5
I understood how to play Jewel Diver	4.22	5	2-5
I understood how to play Road Tour	3.64	4	2-5
I understood how to play Sweep Seeker	4.32	5	2-5
I would like to have more instruction before beginning	3.21	3	1-5
I felt I had enough time to complete the day's training	3.89	4	1-5
I had difficulties using the mouse to play the games	2	1	1-5
The registration process was difficult	1.64	1	1-5
I understand the difference between the "assessment" and "training" steps	3.82	4	1-5
I knew how to pause the game if I needed to	3.89	4	1-5
I understood that my eyes had to be at a specific distance from the computer	3.79	4	1-5
I understood how to score high points in all the games	3.07	3	1-5
I would like to use this program again	3.32	4	1-5
I am very good with computers	2.32	2	1-5

a second version of the HPT. The majority of participants completed post-testing within one week of ending the DriveSharp training.

With the exception of the usability questionnaire, the protocol was the same for the control group's pre- and post-testing sessions, including the five-week interval between testing.

## Results

### *Attendance and Training Times*

Posit Science has stated that participants must complete at least 10 hours of training for the program to have an effect on driving behaviors (Posit Science, 2010, p. 5). To satisfy this requirement we scheduled 10 one-hour sessions for the participants over a five-week period. The average attendance of the participants who were post-tested was  $M = 9.43$  sessions (this includes two participants from the first course attendees who attended 11 sessions).

Although the participants were scheduled for 10 hours of training time, the average time spent specifically on the training component of the DriveSharp program was only  $M = 4.2$  hrs.,  $SD = 1.45$  (Table 3). This is less than one half the time recommended. Additionally, some participants had zero training hours logged for some of the games. Training times of zero do not indicate that participants had no experience with the task; due to misunderstandings during the training sessions, a number of participants spent their time completing assessments instead of the training exercises.

### *Learning as Shown by Assessments*

The number of assessments varied greatly for each exercise: Sweep Seeker ( $M = 3.17$ ,  $SD = 1.56$ ), Jewel Diver ( $M = 2.67$ ,  $SD = 1.2$ ), and Road Tour ( $M = 2.71$ ,  $SD = 1.3$ ). When using recommended scheduling, DriveSharp indicated that after a participant completed a given level of training, the program would prompt them for an assessment. However, as participants' training sessions lasted longer than 20 minutes, they were free to select not only the exercises but also additional assessments. To determine if the number of assessments related to performance at the individual level, the percentage improvement (from baseline raw

scores) for each exercise was modeled in relation to the number of assessments (see Figure 3).

These data demonstrate that the participants varied greatly in the number of assessments taken, and improvements are quite variable both across and within individuals. Road Tour and Jewel Diver appear to have a general linear trend indicating further improvement beyond the first assessment. Although it seems natural to predict that participants would improve more with each assessment because they had received more training between assessments, such an assumption would be inappropriate. As we have noted, it is possible that no training at all occurred between a participant's assessments. Recurrent exposure to an assessment would presumably lead to some improvements, but these improvements are likely to be less robust than if the participant spent time training between assessments.

That being said, to determine if there was a significant improvement in performance, separate repeated-measures analyses of variance (ANOVAs) were conducted to analyse the per cent improvement for the first three assessments. Performance during only the first three assessments was used as these cells were the most populated; only 40 per cent of participants completed a fourth assessment for the Sweep Seeker game, and only 21 per cent for both Jewel Diver and Road Tour. Only data from participants who had completed all three assessments were used. The Greenhouse-Geisser correction was used where sphericity was violated.

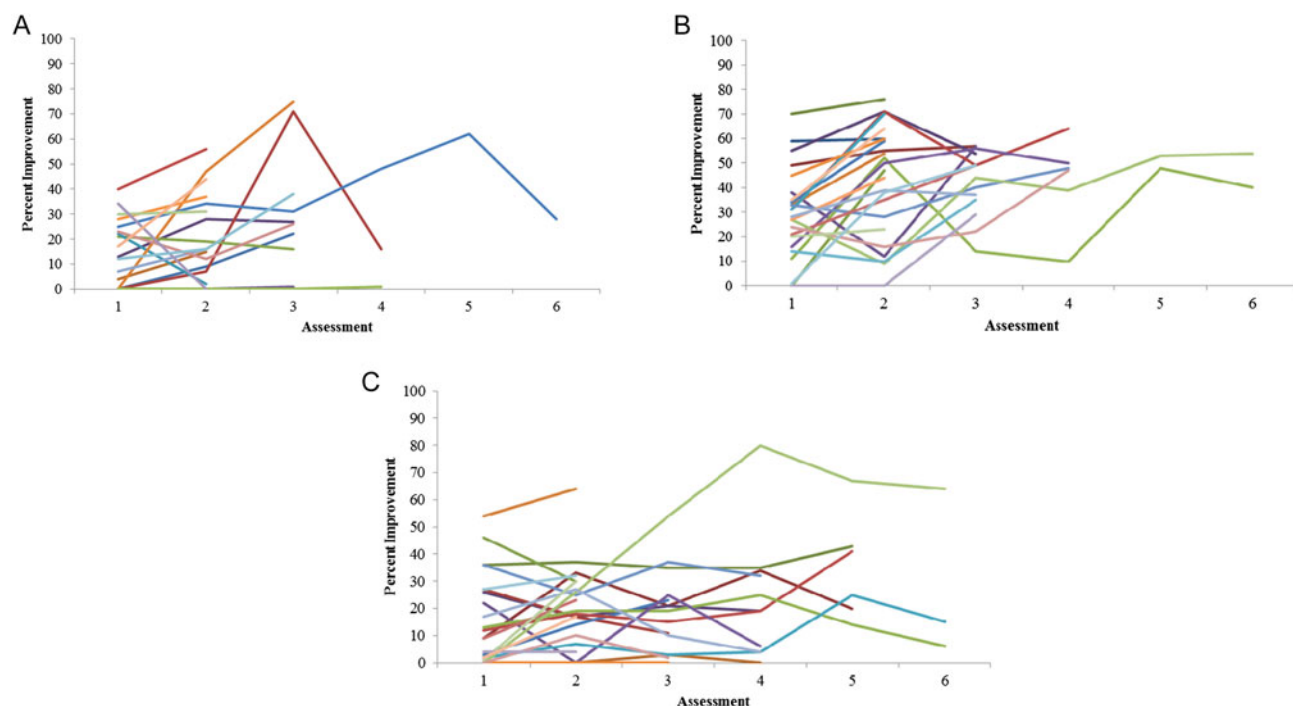
Descriptive statistics for each analysis are presented in Table 4. Analysis of the Sweep Seeker game indicated there was no significant improvement in performance,  $F(2, 32) = 1.63$ ,  $p = .213$ ,  $\eta^2 = .092$ . Analysis of Jewel Diver data ( $n = 15$ ) indicated significant improvement –  $F(1, 14) = 4.78$ ,  $p = .046$ ,  $\eta^2 = .254$ . Finally, analysis of Road Tour assessments ( $n = 10$ ) was non-significant,  $F(1, 10) = 1.43$ ,  $p = .266$ ,  $\eta^2 = .137$ . Thus, there is no strong evidence of improvement with time on task, even though the time spent was not unambiguously associated with training, as opposed to assessments.

### *Analysis of Outcome Variables*

If DriveSharp training is beneficial for skills related to driving safety, then one would expect the experimental group to show improvement in HPT latencies, as well as performance on Trails A and B, two tests that are associated with driving safety. Furthermore, if this improvement in post-test outcome variables is the result of training, and not merely an artifact of practice on the outcome variables, then the experimental group should demonstrate greater gains than the control group.

**Table 3: Training times in minutes**

Test	Mean (SD)	Range
Sweep Seeker	80.71 (25.79)	37–121
Jewel Diver	83.21 (42.38)	0–157
Road Tour	87.82 (46.94)	0–169
Total	252.54 (87.19)	98–376



**Figure 3:** A. Per cent improvement of participants for each assessment of Jewel Diver performance. B. Per cent improvement of participants for each assessment of Road Tour performance. C. Per cent improvement of participants for each assessment of Sweep Seeker performance. Participants vary on the number of assessments they took. DriveSharp did not report the exact value if a per cent improvement was less than zero (these values are recorded and displayed as a “zero” improvement).

In order to assess this, we compared the experimental and control groups in their pre-test versus post-test differences on the HPT, Trails A and Trails B.

Only those with complete data sets were included in the analysis; participants with results missing from either pre- or post-testing were excluded. One participant was removed from the experimental group because their pre-test HPT scores were over 3 standard deviations greater than the sample average (23 s. vs. 3.55 s.). The final analysis included  $n = 23$  participants from the experimental group and  $n = 18$  from the control group. Analyses of the Hazard Perception Test were completed using reaction time for the scenes containing a traffic conflict.

*Pre-Test Group Differences*

A comparison of the pre-test assessments (Trails A, Trails B, and the HPT) was conducted to examine what

group differences, if any, existed at baseline. The descriptive statistics and significance tests for the pre-test analyses are displayed in Table 5. The analyses indicated that the control and the experimental group differed significantly on their pre-test scores for Trails A ( $t[40] = 2.02, p = .050$ ), Trails B ( $t[40] = 2.59, p = .013$ ), and the HPT ( $t[40] = 2.96, p = .005$ ). The experimental group demonstrated slower performance than the control group on all tests. As a consequence, there is a need to control for these pre-test differences while assessing post-test differences in performance.

*Post-Test Group Differences*

The descriptive statistics and significance tests for the post-test analyses are found in Table 5. The analysis indicated that there was no significant difference between the control and the experimental groups’ Trails A test results ( $t[40] = 1.17, p = .249$ ). However, there was a

**Table 4: Descriptive statistics for the analysis of performance improvement (%)**

No. of Assessments	Sweep Seeker ( $n = 17$ )	Jewel Diver ( $n = 15$ )	Road Tour ( $n = 14$ )
Assessment	Mean (SD)	Mean (SD)	Mean (SD)
1	11.94 (13.12)	10.53 (11.18)	24.86 (16.11)
2	14.71 (12.29)	16.20 (12.67)	34.71 (23.19)
3	17.88 (14.92)	26.40 (25.02)	42.29 (13.51)



**Table 5: Pre- and post-test experimental and control group performance on Trails A, B, and HPT**

	Experimental Group Mean (SD)	Control Group Mean (SD)	p-value
Pre-test Trails A	38.91 (15.69)	30.84 (7.20)	.050*
Pre-test Trails B	108.03 (64.76)	67.275 (18.31)	.013*
Pre-HPT	3.82 (1.01)	2.98 (.73)	.005*
Post-test Trails A	34.75 (13.43)	30.29 (10.39)	.249
Post-test Trails B	90.49 (54.28)	65.05 (16.39)	.039*
Post-HPT	3.55 (1.75)	2.65 (.68)	.028*

\* Indicates significant difference between experimental and control group on that measure ( $p < .05$ ).

significant group difference on Trails B ( $t[28.37] = 2.17$ ,  $p = .039$ ) and the HPT ( $t[31.47] = 2.29$ ,  $p = .028$ ), where the experimental group demonstrated slower performance.

A comparison of pre-test and post-test data indicates that both groups demonstrated improved speed on all outcome variables, as might be expected from practice effects alone. However if the DriveSharp training influenced performance as expected, the experimental group should reveal significantly greater post-test improvements.

#### Gain Score Analysis

To account for both practice effects and the slower pre-test scores for the experimental group, we conducted a third test, an examination of the difference between pre-test and post-test scores. This test, referred to as *gain score analysis*, determined if the rate of learning was equivalent between groups. If DriveSharp does have an effect on performance, then we would expect to see a larger change for the experimental than the control group.

Descriptive statistics and significance values are shown in Table 6. Analyses revealed that there were no differences in the rate of learning for Trails A, Trails B, or the HPT. The mean gain scores on both Trails A and B indicate that the experimental group had greater average improvement; however, there was considerable variance around these scores as well. The mean change in the HPT performance revealed a trend contrary to expectations, in that the control group showed greater gains.

Because the groups differed on age at testing, an analysis was carried out on gain scores using age as a co-variate. Descriptive statistics and significance values are provided in Table 6. After adjusting for the group differences due to age, there was a non-significant trend towards greater improvement in the experimental group. However, in no case were the group differences significant. We also adjusted for individual differences in age, MMSE score, and number of collisions within the past two years (see Table 5). Again, no significant group differences were found.

#### Training Times and Trails A, B, and HPT Gain Scores

If the DriveSharp program produces improvements in skills related to safe driving, then one would expect those who spend more time in the exercises to show the greatest gains in driving-related skills. To examine this hypothesis, we evaluated the relationship between the time spent training on each DriveSharp game and the difference scores from pre-test to post-test (i.e., gain scores) for Trails A, Trails B, and the HPT.

Regression analyses revealed that there was no significant relationship between time spent training on the Sweep Seeker component for either Trails A gain scores ( $F[1, 22] = .01$ ,  $p = .911$ ), Trails B gain ( $F[1, 22] = .94$ ,  $p = .342$ ), or HPT gain ( $F[1, 22] = .66$ ,  $p = .425$ ); the Jewel Diver component for Trails A ( $F[1, 22] = .00$ ,  $p = .990$ ), Trails B ( $F[1, 22] = .33$ ,  $p = .574$ ), or the HPT ( $F[1, 22] = .38$ ,  $p = .544$ ); or the Road Tour component for Trails A ( $F[1, 22] = .01$ ,  $p = .921$ ), Trails B ( $F[1, 22] = 1.37$ ,  $p = .253$ ),

**Table 6: Gain score (pre- and post-test) values for experimental and control group performance on Trails A, B, and HPT**

	Experimental Group	Control Group	p-value; $\eta^2$		
			No Co-variables	Age	Age, MMSE, Collisions
Trails A Gain	4.16 (14.19)	.63 (5.75)	.278; .024	.120; .061	.103; .070
Trails B Gain	17.54 (60.92)	1.56 (19.56)	.238; .028	.492; .012	.444; .016
HPT Gain	.27 (1.47)	.33 (.62)	.840; .001	.321; .025	.335; .025

**Note:** Cell values are arithmetic means and standard deviations.

\* Indicates significant difference between experimental and control group on that measure ( $p < .05$ ).

HPT = hazard perception test.

MMSE = Mini-Mental State Exam.

or the HPT ( $F[1, 22] = .837, p = .370$ ). Additionally, there was no relationship between total time spent training and either Trails A gain scores ( $F[1, 22] = .00, p = .993$ ), Trails B gain ( $F[1, 22] = .00, p = .998$ ), or the HPT gain ( $F[1, 22] = .851, p = .366$ ). These findings indicate that neither the time participants spent training on each component of DriveSharp, nor the total time training, were significantly related to their improvement scores.

### Usability

Responses to the usability questionnaire are provided in Table 2. Generally, respondents found the system easy to use and believed it was not unnecessarily complex. They also understood how it would improve their driving skills and would recommend it to a friend. Participants believed most people would learn the system quickly and that the registration process was not difficult. However, most users would not buy this program nor would they be likely to use the program again.

On a *per game* basis, respondents believed that they understood how to play Sweep Seeker and Jewel Diver, while Road Tour was more difficult. They felt that they understood the difference between “assessment” and “training” steps, and they knew how to pause the games if needed.

It is important to note (see the range statistics in Table 2) that there was considerable variability in the responses given to these usability questions. Some participants were quite positive in their evaluations, while others were very negative. This variability indicates that experiences with the program are person-specific, and it is likely that compliance and effectiveness will depend on the match between the demands of the software and the capabilities of the user.

Self-reports are not necessarily accurate. Answers may be exaggerated, omitted because a respondent is embarrassed, or inaccurate because of forgotten details. For example, results from the usability questionnaire indicated that the participants believed they understood the difference between assessment and training, but their data demonstrate fairly clearly that this was not the case, as they took, on average, many more assessments than requested. This does not mean that the usability data have no value. Regardless of level of true comprehension, the self-perception of comprehension is necessary for user acceptance.

## Discussion

Usability assessments of the program indicated that participants rated the training experience favorably, and that they believed it had a positive effect on their driving behaviors. However, analyses of performance data

did not reveal any significant benefits to the exercises. After adjusting for group differences due to age, MMSE scores and number of collisions within the past two years, we found that – although there was a trend towards greater improvement in the experimental group – effects sizes remained low, and the difference was still non-significant (see Table 6). Additionally, although there was significant improvement in Jewel Diver from baseline to the third assessment, neither the time participants spent training on each component of DriveSharp, nor the total time training, was significantly related to their improvement scores.

Non-random assignment and variability in training times are both challenges to interpreting the results of the study. Participants contacted the centre specifically either to register for the DriveSharp program or to volunteer as part of the control group. They may have self-selected such that only those who felt that they were most in need of training agreed to join the training group. This self-selection may have worked against any demonstration of improvement. However, if the training group was “extreme”, then they should have shown more improvement, if only as an artifact of “regression towards the mean”. Second, in applied settings, it is very likely that those who feel themselves to be most in need of training will be more likely to enroll in training programs. Thus, the present results reflect the realities of training efforts for driver improvement.

Additionally, training times, which were an average of 4.2 hours, are problematic as any driving skills the program may improve are indicated to be a result of receiving a *minimum* of 10 hours of training. However, it is noteworthy that other studies of skill training (e.g., Horswill et al., 2010; Roenker et al., 2003) reported improvements in performance after much shorter periods of time. In fact, Roenker et al. (2003) demonstrated that speed-of-processing training of 4.2 hrs on average resulted in significant improvements in on-road driving.

One possibility may be that performance improves over time to some asymptotic level and that improvements in driving skills may be positively related to the amount of time spent completing the exercises. However, although the data from the current study suggests initial improvement from baseline to the first assessment, after this, participants’ improvement leveled off. As noted previously, the amount of training was not consistent between participants. Additionally, it is clear that despite the continual presence of a facilitator, there was considerable variability in time spent training on individual games and, thus, in the total training received. Despite these limitations, the findings of this study are consistent with other assessments of DriveSharp’s impact on driving performance (Dobres et al., 2013).

From some initial pilot testing, it appeared that DriveSharp was designed to require participants to complete a certain amount of training for each game. It was possible, however, for participants to play whichever game they chose. This is likely the cause of the low (in some cases non-existent) training times observed for some of the games. However, it does not account for the low overall training times observed. One possible explanation is that there is invariably a certain amount of time spent at the start of each class to turn on the computers, discuss the program, and converse about other topics. During a post-testing session, one participant reported 20 minutes on average per class to begin training, leaving only 40 minutes for the exercises themselves.

It is also worth noting that total training time as recorded by the software is only a reflection of time spent with the exercises; time spent in “assessment” sessions is not recorded. Some participants completed as many as eight assessments of one game, and assessments can be quite lengthy. However, when reporting individual results, DriveSharp records and reports on a maximum of six assessments. Assessments are important to provide an evaluation of progress; however, they should not be substituted for training sessions as their design features and practice benefits likely differ from the training sessions. For future DriveSharp studies, we recommend limiting the number of assessments to three per game – one at the beginning, one halfway through the course, and one at the termination.

Although the usability data reflect a generally positive attitude towards the program, a limitation in their interpretation is participant attrition. There were six participants who dropped out from the DriveSharp experimental group, and we were unable to obtain usability data from these individuals. Anecdotally, some of those individuals dropped out because they found the program too difficult and/or they did not understand how it would improve their driving. As DriveSharp assesses performance from mouse speed and accuracy calculations, individuals with no previous computer experience are at a disadvantage compared to those comfortable with the technology.

Clearly then, DriveSharp faces challenges regarding compliance, usability, and effectiveness. The compliance and usability problems our study uncovered have implications for implementation of the exercises. If it is difficult in a facilitated setting to ensure that participants understand the instructions and comply with them, then how much more difficult will it be if individuals work without a facilitator? Given that more than 10 hours of involvement with the product did not produce meaningful change in performance, how likely is it that older adults, many of whom are not “savvy” with video games, will invest even more

time and effort with the programs? Is the time and effort commensurate with gains, or would it be more sensible to have individualized, on-road instruction that focuses on those sub-skills that are most in need of remediation? These questions must be addressed in additional, independent research.

## References

- Ball, K., Berch, D. B., Helmers, K. F., Jobe, J. B., Leveck, M. D., Marsiske, M., et al. (2002). Effects of cognitive training interventions with older adults. A randomized controlled study. *The Journal of the American Medical Association*, 288, 2271–2281.
- Ball, K., & Owsley, C. (1993). The useful field of view test: A new technique for evaluating age-related declines in visual function. *Journal of the American Optometry Association*, 64, 71–79.
- Ball, K., Owsley, C., Sloane, M. E., Roenker, D. L., & Bruni, J. R. (1993). Visual attention problems as a predictor of vehicle crashes in older drivers. *Investigative Ophthalmology and Visual Science*, 34, 3110–3123.
- Ball, K., Roenker, D., Bruni, J., Owsley, C., Sloane, M., Ball, D., et al. (1991). Driving and visual search: Expanding the useful field of view. *Investigative Ophthalmology and Visual Science* (Suppl. 32), 1041.
- Ball, K. K., Edwards, J. D., Ross, L. A., & McGwin, G. (2010). Cognitive training decreases motor vehicle collision involvement among older drivers. *Journal of the American Geriatrics Society*, 58, 2107–2113.
- Ball, K. K., Roenker, D. L., Wadley, V. G., Edwards, J. D., Roth, D. L., McGwin, G., et al. (2006). Can high-risk older drivers be identified through performance-based measures in a Department of Motor Vehicles setting? *Journal of American Geriatric Society*, 54, 77–84.
- Bowers, A. R., Anastasio, R. J., Sheldon, S. S., O'Connor, M. G., Hollis, A. M., Howe, P. D., et al. (2013). Can we improve clinical prediction of at-risk older drivers? *Accident Analysis and Prevention*, 59, 537–547.
- Chapman, P., Underwood, G., & Roberts, K. (2002). Visual search patterns in trained and untrained novice drivers. *Transportation Research Part F, Traffic Psychology and Behavior*, 5, 157–167.
- Chipman, M., MacGregor, C., Smiley, A., & Lee-Gosselin, M. (1992). Time vs. distance as measures of exposure in driving surveys. *Accident Analysis & Prevention*, 24(6), 679–684.
- Dobbs, B., & Schopflocher, D. (2010). The introduction of a new screening tool for the identification of cognitively impaired medically at-risk drivers: The SIMARD A modification of the DemTect. *Journal of Primary Care and Community Health*, 1, 119–127.
- Dobres, J., Potter, A., Reimer, B., Mehler, B., Mehler, A., & Coughlin, J. (2013). Assessing the impact of “Brain Training” on driving performance, visual behavior, and

- neuropsychological measures. *Proceedings of the Seventh International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, Bolton Landing NY, June 17–20. Ames, IA: Public Policy Center of the University of Iowa (pp. 50–56). Available at <http://drivingassessment.uiowa.edu/2013/proceedings>
- Edwards, J. D., Delahunt, P. B., & Mahncke, H. W. (2009). Cognitive speed of processing training delays driving cessation. *Journal of Gerontology A Biological Science Medical Science*, *64*, 1262–1267.
- Edwards, K. J., Leonard, K., Lunsman, M., Dodson, J., Bradley, S., Myers, C., et al. (2008). Acceptability and validity of older driver screening with the driving health inventory. *Accident Analysis and Prevention*, *40*, 1157–1163.
- Evans, L., (2004). *Traffic Safety*. Bloomfield Hills, MI: Science Serving Society.
- Fisher, D., Pollatsek, A., & Pradhan, A. (2006). Can novice drivers be trained to scan for information that will reduce their likelihood of a crash? *Injury Prevention*, *12*, 25–29.
- 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.
- Goode, K. T., Ball, K., Sloane, M., Roenker, D. L., Roth, D. L., Myers, R. S., et al. (1998). Useful field of view and other neurocognitive indicators of crash risk in older adults. *Journal of Clinical Psychology in Medical Settings*, *5*, 425–440.
- Hakamies-Blomqvist, L., Raitanen, T., & O'Neill, D. (2002). Driver ageing does not cause higher accidents per km. *Transportation Research Part F: Traffic Psychology and Behavior*, *5*, 271–274.
- 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*, 939–944.
- Horswill, M. S., Kemala, C. N., Wetton, M., Scialfa, C. T., & Pachana, N. A. (2010). Improving older drivers' hazard perception ability. *Psychology and Aging*, *25*, 464–469.
- Horswill, M. S., Mannington, S., McCullough, C., Wood, J., Pachana, N., McWilliam, J., et al. (2008). The hazard perception ability of older drivers. *Journal of Gerontology: Psychological Sciences*, *63B*, P212–P218.
- Horswill, M. S., & McKenna, F. (2004). Drivers' hazard perception ability: Situation awareness on the road. In S. Banbury, & S. Tremblay (Eds.), *A cognitive approach to situation awareness* (pp. 155–175). Aldershot, UK: Ashgate.
- Horswill, M. S., & Taylor, K., Newnam, S., Wetton, M., & Hill, A. (2013). Even highly experienced drivers benefit from a brief hazard perception training intervention. *Accident Analysis and Prevention*, *52*, 100–110.
- Li, G., Braver, E. R., & Chen, L. H. (2003). Fragility versus excessive crash involvement as determinants of high death rates per mile driven in older drivers. *Accident Analysis and Prevention*, *35*, 227–235.
- Lochner, M., & Trick, L. M. (2011). Attentional tracking of multiple vehicles in a highway driving scenario. *Proceedings for 6th International Symposium on Human Factors in Driving Assessment, Training, and Vehicle Design*.
- Marrington, S. A., Horswill, M. S., & Wood, J. M. (2008). The effect of simulated cataracts on drivers' hazard perception ability. *Ophthalmology Vision Science*, *85*, 1121–1127.
- McKenna, F. P., & Crick, J. L. (1991). *Hazard perception in drivers: A methodology for testing and training*. Final Report, Behavioural Studies Unit, Crowthorne, UK: Transport and Road Research Laboratory.
- Meuleners, L. B., Harding, A., Lee, A. H., & Legge, M. (2006). Fragility and crash over-representation among older drivers in Western Australia. *Accident Analysis and Prevention*, *38*, 1006–1010.
- Nielsen, J. (2013). *Seniors as web users*. Nielsen Norman Group. Retrieved September 27, 2015 from <http://www.nngroup.com/articles/usability-for-senior-citizens/>
- Owsley, C., Ball, K., Sloane, M. E., Roenker, D. L., & Bruni, J. R. (1991). Visual/cognitive correlates of vehicle accidents in older drivers. *Psychology of Aging*, *6*, 403–415.
- Posit Science. (2010). *DriveSharp overview and installation*. Retrieved September 27, 2015 from [http://www.brainhq.com/sites/default/files/pdfs/ds\\_install.pdf](http://www.brainhq.com/sites/default/files/pdfs/ds_install.pdf)
- Roenker, D. L., Cissell, G. M., Ball, K. K., Wadley, V. G., & Edwards, J. D. (2003). Speed-of-processing and driving simulator training result in improved driving performance. *Human Factors*, *45*, 218–233.
- Ross, R., Cordazzo, S., & Scialfa, C. (2014). Predicting on-road driving performance and safety in healthy older adults. *Journal of Safety Research*, *51*, 73–80.
- Rubin, G. S., Ng, E. S., Bandeen-Roche, K., Keyl, P. M., Freeman, E. E., & West, S. K. (2007). A prospective, population-based study of the role of visual impairment in motor vehicle crashes among older drivers: The SEE study. *Investigative Ophthalmology and Vision Science*, *48*, 1483–1491.
- Scialfa, C., Borkenhagen, D., Lyon, J., Deschênes, M., Horswill, M., & Wetton, M. (2012a). The effects of driving experience on responses to a static hazard perception test. *Accident Analysis and Prevention*, *45*, 547–553.
- Scialfa, C., Deschênes, M., Ference, J., Boone, J., Horswill, M., & Wetton, M. (2012b). Hazard perception in older drivers. *International Journal of Human Factors and Ergonomics*, *1*, 221–233.
- Sekuler, R., McLaughlin, C., & Yotsumoto, Y. (2008). Age-related changes in attentional tracking of multiple moving objects. *Perception*, *37*, 867–876.

- Sims, R. V., McGwin, G., Allman, R. M., Ball, K., & Owsley, C. (2000). Exploratory study of incident vehicle crashes among older drivers. *Journal of Gerontology, 55*, M22–M27.
- Tombaugh, T. (2004). Trail making test A and B: Normative data stratified by age and education. *Archives of Clinical Neuropsychology, 19*, 204–213.
- Trick, L. M., Perl, T., & Sethi, N. (2005). Age-related differences in multiple-object tracking. *Journal of Gerontology: Series B: Psychological Sciences and Social Sciences, 60B*, 102–105.
- Willis, S. L., Tennstedt, S. L., Marsiske, M., Ball, K., Elias, J., Koepke, K. M., et al. (2006). Long-term effects of cognitive training on everyday functional outcomes in older adults. *The Journal of the American Medical Association, 296*, 2805–2814.
- 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.