One- and Three-Screen Driving Simulator Approaches to Evaluate Driving Capacity: Evidence of Congruence and Participants’ Endorsement

Carrie Gibbons, Nadia Mullen, Bruce Weaver, Paula Reguly, Michel Bédard

OBJECTIVE. We examined the validity of one-screen versus three-screen driving simulators and their acceptability to middle-aged and older drivers.

METHOD. Participants aged 40–55 or 65 and older (N = 32) completed simulated drives first with a single monitor and then with a three-monitor setup, followed by pen-and-paper measures and an interview.

RESULTS. Mean differences between one- and three-screen drives were not statistically significant for Starting/Stopping and Passing/Speed. Correlations between the two drives indicated moderate positive linear relationships with moderate agreement. More errors occurred on the one-screen simulator for Signal Violation/Right of Way/Inattention, Moving in a Roadway, Turning, and Total Scores. However, for Moving in a Roadway, Turning, and Total Scores, correlations between drives indicated strong positive linear relationships. We found no meaningful correlation between workload, computer comfort, simulator discomfort, and performance on either drive. Participants found driving simulators acceptable.

CONCLUSION. Findings support the use of one-screen simulators. Participants were favorable regarding driving simulators for assessment.

Aging is associated with declines in cognitive capacity, vision, and physical abilities that may impair one’s ability to drive safely (Anstey, Wood, Lord, & Walker, 2005; Duchek et al., 2003; Marshall & Man-Son-Hing, 2011). Nevertheless, older adults should not have their driving licenses revoked simply on the basis of age or an age-related medical diagnosis, because many older adults are still safe to drive (Dickerson et al., 2007; Duchek et al., 2003; Ott et al., 2008). Hence, achieving a proper balance between maintaining the driving privilege and ensuring public safety requires a fair and equitable means of assessing driving capacity.

Traditionally, driver assessments have comprised both clinical and on-road driving components. The on-road component, however, creates risk-management issues by placing the driving evaluator, the public, and the driver at risk. Moreover, as the population continues to age, we may lack sufficient capacity to fully evaluate all older drivers requiring on-road assessments (Dickerson, Reistetter, & Gaudy, 2010). Driving simulators provide another avenue to evaluate the driving capacity of older adults. Simulator evaluations can identify those at risk of committing at-fault vehicle crashes (Hoffman & McDowd, 2010) and traffic violations (Lee & Lee, 2005), potentially adding to other clinical tools to determine driving capacity. They also enable evaluation of...
driving skills in a highly standardized fashion that on-road evaluations cannot achieve. Driving simulators are now affordable and may also be more cost-effective and time efficient than on-road driving evaluations.

Simulator validity research has demonstrated a good level of correspondence between simulator and on-road measures of driving behavior such as speed, lateral position, and braking (Bella, 2008; Bédard, Parkkari, Weaver, Riendeau, & Dahlquist, 2010; Hoffman, Lee, Brown, & McGehee, 2003; Lee, Cameron, & Lee, 2003; Mayhew et al., 2011). In a recent literature review, we found that driving behavior in simulators approximates (as opposed to exactly replicating) on-road driving behavior, but that is sufficient for most assessment purposes (Mullen, Charlton, Devlin, & Bédard, 2011).

Note that previous work (Bédard et al., 2010) indicated that demerit points accrued by 8 older adults (aged 67–81 yr) on-road were correlated ($r = .74$) with points accrued on a three-screen STISIM simulator (Systems Technology, Inc., Hawthorne, CA). Shechtman, Classen, Awadzi, and Mann (2009) also found similarities in the number of errors drivers made while making turns during on-road testing compared with a drive in a STISIM simulator integrated in a full vehicle. Lee, Cameron, and Lee (2003) found a positive correlation between older drivers’ performance in a one-screen STISIM simulator and performance in an on-road driving evaluation ($r = .72$). They also found that the simulator could identify older drivers at risk of future traffic violations and that it was sensitive to age-related changes in driving performance (Lee & Lee, 2005).

To further study the utility of simulators within a comprehensive evaluation process, Devos et al. (2007) examined the ability of a clinical screening battery (including assessments of driving history, cognition, and performance on a full-sized Ford Fiesta driving simulator) to predict fitness to drive in participants diagnosed with Parkinson disease (PD). A driving evaluator determined participants’ fitness to drive (pass–fail) after an on-road driving test. Using four measures of health (disease duration, contrast sensitivity, Clinical Dementia Rating [Morris, 1993], and the motor part of the Unified Parkinson’s Disease Rating Scale [Movement Disorder Society Task Force, 2003]), 90% of participants with PD were correctly classified (sensitivity = 91%, specificity = 90%). When simulator performance was added, 97.5% of participants with PD were correctly classified (sensitivity = 91%, specificity = 100%). These results were recently replicated (Devos et al., 2013), demonstrating that driving simulators can improve our ability to accurately classify safe and unsafe drivers and, more important, that they can help reduce the number of false positives (safe drivers classified as “unsafe”).

Although simulators are increasingly affordable, their complexity and cost may still represent barriers. Driving simulators range from fully enclosed cabs with motion platforms and 360˚ screens to simple computer desktop simulators. To ensure that complexity and cost do not act as a deterrent, validation of simpler and cheaper one-screen platforms should be examined. Recent research investigated the performance of 52 young adults who completed identical drives on a one-screen desktop simulator and a STISIM three-screen simulator with a fixed-base car seat and 135˚ field of view (Lemieux, Stinchcombe, Gagnon, & Bédard, 2014). Participants performed similarly in both environments as measured by global indicators of driving performance. Additionally, the correlations between the one- and three-screen configurations were strong for both simulator-recorded errors ($r = .72$) and demerit scores ($r = .73$). These findings indicate that the one-screen desktop option could replace more complex and costly driving simulators in a clinical environment.

A second issue is that it is unclear how older adults view driving simulators as tools to evaluate driving capacity. Liu, Watson, and Miyazaki (1999) assessed a virtual reality driving simulator and concluded that use of this technology with older adults was feasible. Their conclusion, however, was based on the use of a head-mounted display, a steering wheel mounted on a table with pedals underneath, and drivers ranging in age from 13 to 76 yr or older. The use of nonsenior drivers in the study limits the generalizability of these findings to our population of interest. Moreover, it is unclear whether participants felt that simulated driving provided an accurate picture of their driving ability.

In another study, researchers questioned participants aged 50 and older after they viewed but did not drive a three-screen simulator to gather opinions on the use of driving simulators in rehabilitation settings (Crisler et al., 2012). Participants were generally positive about the use of simulators for rehabilitation but less so for evaluation or screening purposes. Participants did not drive the simulator, however, so no data were collected about actual simulated driving experiences.

The objectives of this study were to examine (1) the validity of a one-screen simulator setup relative to a three-screen setup and (2) the acceptability of driving simulators as a tool to assess driving capacity from the perspective of both middle-aged (age 40–55) and older (age 65 or older) drivers.
Method

Research Design

Our study comprised a one-group pretest–posttest design to examine simulator validity on one-screen and three-screen setups and a qualitative open-ended interview to explore acceptability of simulator assessment. Measures were completed after participants finished two simulated drives. Participants took approximately 1.5 hr to complete the project, and they were reimbursed $50. The project received ethics approval from the research ethics boards at Lakehead University and St. Joseph’s Care Group in Thunder Bay, Ontario, Canada. All participants provided informed consent before taking part in the study.

Participants

Participants were recruited through word of mouth or from our research center’s database of people interested in taking part in driving research. We recruited drivers from two age groups, age 40 to 55 yr and age 65 yr and older. Inclusion criteria were the ability to speak English, possession of a valid driver’s license, and regularly driving at least 3 times per week.

Measures

Participants completed the National Aeronautics Space Administration Task Load Index (NASA TLX), a workload assessment tool used to assess simulated environments (Hart & Staveland, 1988). NASA TLX comprises six questions and includes the following subscales: Mental Demands, Physical Demands, Temporal Demands, Own Performance, Effort, and Frustration. Higher scores indicate greater demand. Hart and Staveland (1988) found test–retest reliability of .83, and Rubio, Diaz, Martin, and Puente (2004) found it correlated with other measures of workload ($r = .98–.99$) and performance ($r = .65$).

Participants provided basic demographic information (e.g., age, gender) and information about their driving history and habits (e.g., self-reported driving frequency and distance, crash history, driving self_restrictions, usual driving speed). They also completed a 16-item questionnaire concerning their use of computers and comfort level with them. Questions were selected from the Attitudes Toward Computers Questionnaire (Jay & Willis, 1992) and scored on a 5-point Likert-type scale, with responses ranging from 1 (strongly agree) to 5 (strongly disagree). Higher scores indicate greater comfort with computers. Finally, participants completed the Simulator Sickness Questionnaire (SSQ; Kennedy, Lane, Berbaum, & Lilienthal, 1993), a 16-item questionnaire designed to assess discomfort that participants may feel while driving in a simulated environment. Response options range from none to severe with a total possible score of 235.62 based on Kennedy et al.’s (1993) scoring system. Higher scores indicate more discomfort. Kennedy and colleagues (1999, cited in Kennedy et al., 2001) reported that the SSQ has a split-half correlation of $r = .80$ for 200 participants after exposure to a virtual environment.

Participants were interviewed by a research assistant (RA) using semistructured questions that the authors developed. Questions focused on how realistic the driving experience felt, any physical discomfort that the simulator caused, whether the simulated drive provided a reasonable reflection of their driving skills and abilities, their preferences and opinions of the one-screen compared with the three-screen setup, and whether participants felt that such an approach could enhance current procedures to assess driving ability. After the questions were pilot tested by the RA in the first few interviews, they were reviewed again by the research team; no changes were made. The interviews were recorded and generally took 10–20 min to complete.

Procedures

After receiving instructions from the RA, participants completed a 5-km (3.1-mi), 15-min practice drive followed by the one-screen and three-screen drives. Participants completed the one-screen component first because evidence suggested that simulator discomfort is less likely with this setup (Johnson, 2005; Mollenhauer, 2004). Thus, we hoped to collect data for at least the one-screen drive for participants who might experience simulator discomfort during the three-screen drive. The two drives were identical, took about 20 min to complete, and were based on a standard Ontario G2 driving assessment circuit. Auditory instructions were provided for participants. We gave a 5-min rest break to participants between drives while the RA adjusted the monitors.

Drives were completed on a STISIM Drive® M400 simulator (Systems Technology, Inc., Hawthorne, CA) consisting of a driver’s seat, passenger seat, steering console with horn, brake and accelerator foot pedals, signal light, and dash board (including speedometer and tachometer). For the one-screen setup, the driver’s view was presented on one 17-in. monitor (45˚ field of view) with a rearview mirror displayed near the center. The three-screen setup displayed the driver’s view across three 17-in. monitors (135˚ field of view) with the rearview mirror displayed on the central monitor and sideview mirrors displayed on the outer monitors.
For each drive, the RA scored participants by using the Manitoba Road Test (MRT). The test can be separated into five components: Starting/Stopping, Signal Violation/Right of Way/Inattention, Moving in a Roadway, Passing/Speeding, and Turning. Driving errors were detected by the simulator software for each drive in 10 categories: off-road crashes, collisions, pedestrians hit, exceeding the speed limit, speeding tickets, traffic light tickets, stop signs missed, centerline crossings, road edge excursions, and illegal turns.

Data Analysis

Descriptive statistics are presented for participants’ characteristics in the Results section. To examine the mean number of errors committed by participants on each of the drives, we used \( t \) tests. We used both Pearson’s \( r \) and the intraclass correlation (ICC) to examine the concordance between one-screen and three-screen drives with both MRT scores and the number of errors recorded by the simulator. We also used Pearson’s \( r \) to examine the associations between MRT scores and the number of errors committed on the two drives and the results from the pen-and-paper tests (NASA TLX comfort with computers questionnaire, and the SSQ). All analyses were performed using IBM SPSS (Version 20; IBM Corporation, Armonk, NY). Confidence intervals for Pearson correlations were computed using the rhoCI macro (Weaver & Koopman, in press).

Data collected from the participant interviews were thematically analyzed by age group (40–55 and >65). The recorded interviews were divided between two of the researchers (Reguly and Gibbons) and played back. Salient data from participant responses were extracted and summarized. Interrater reliability was assessed with 10 of the 32 interviews being reviewed by both researchers (Reguly and Gibbons) and reliability being confirmed by a third researcher (Mullen). The few discrepancies identified were subsequently discussed until agreement was reached. Reliability was intermittently measured throughout interview data retrieval (i.e., reliability was assessed on approximately every third recording) to account for any observer drift.

Results

Participants

Thirty-two drivers took part in the study; 16 (8 men, 8 women) were in the younger age group and 16 (8 men, 8 women) were in the older age group. In the younger group, ages ranged from 40 to 55 (mean \( M = 46.87 \), standard deviation \( SD = 5.16 \)); in the older group, participants were aged 65–87 (\( M = 74.94, SD = 7.09 \)). The majority of younger (\( n = 12, 75.0\% \)) and older (\( n = 10, 62.5\% \)) participants drove more than 50 km (31 mi) per week. Participants were asked about self-imposed restrictions on driving. Only 1 (6.25%) younger participant restricted driving (daytime only). Although 10 (62.5%) participants in the older group had no restrictions, the remaining 6 (37.5%) indicated that they restricted driving to daytime (\( n = 2 \)), outside of rush hour (\( n = 4 \)), local routes (\( n = 1 \)), fair weather (\( n = 1 \)), and other (\( n = 2 \)).

Simulated Driving Performance

We examined the concordance between the two simulated drives with scores on the five components and total score of the MRT (see Table 1). The mean differences between the one- and three-screen drives were not statistically significant for Starting/Stopping, \( t(25) = 0.98, p = .334 \), and Passing/Speeding, \( t(25) = 1.54, p = .136 \). For each of these variables, correlations between the scores on the two drives indicated a moderate to strong positive linear relationship, and the ICCs indicated moderate to strong agreement. Although the difference between the scores for the Moving in a Roadway variable was statistically significant, \( t(25) = 2.30, p = .030 \), the Pearson \( r \) and ICC reflected a similar pattern.

For Signal Violation/Right of Way/Inattention, \( t(25) = 2.75, p = .011 \), and Turning, \( t(25) = 6.18, p < .001 \), scores were significantly higher (indicating a greater number of errors) for the one-screen drive. For Signal Violation/Right of Way/Inattention, the Pearson \( r \) demonstrated a weak linear relationship between the two drives, and the ICC indicated only fair agreement. However, participants registered few demerit points for this domain on the three-screen drive, effectively distinguishing our ability to demonstrate a correlation. For the Turning component, the correlation between scores on the one-screen and three-screen drives indicated a strong positive linear relationship. Because there was a significantly greater number of demerit points during the one-screen drive, the ICC was weaker than the Pearson \( r \). (The magnitude of the ICC is affected by both linear association between the variables and the size of the mean difference.)

We also examined the total scores on the drives and found a significantly greater number of errors on the one-screen simulator, \( t(25) = 6.47, p \leq .001 \). The Pearson \( r \) indicated a strong positive linear relationship between the two total scores but, as we saw for Turning, the ICC was lower because of a significantly greater number of demerit points during the one-screen drive. We found a similar
pattern for the simulator-recorded mistakes, \( t(25) = 2.57, p = .016 \) (see Table 1 and Figures 1 and 2).

**Workload Assessment**

The scores on the NASA TLX ranged from 2.67 to 15.83 (\( M = 9.39, SD = 3.27 \)). Two participants were unable to complete this component of the testing because of simulator discomfort. Participants’ subjective workload evaluation of driving the simulator was not significantly related to scores on the MRT or the simulator-generated number of errors on either the one-screen or the three-screen drives (Table 2).

**Comfort With Computers**

Total scores for the computer comfort questionnaire were approximately normally distributed with a (\( M = 66.19, SD = 7.47 \)). Total scores on this questionnaire were not correlated with total scores on the MRT or the total number of simulator-computed driver errors on either the one- or three-screen simulator (see Table 2).

**Simulator Discomfort**

The total weighted score on the SSQ ranged from 0 to 63.58 (\( M = 12.74, SD = 20.59 \)). Ten participants (31.3%) experienced simulator discomfort (3 younger and 7 older participants). Two (6.3%) participants experienced simulator discomfort during the orientation

![Figure 1. Total demerits on the Manitoba Road Test: one screen and three screens.](image-url)

The solid line is the least squares regression line, and the dashed line is the line of perfect agreement. For points below the dashed line, the one-screen score is greater than the corresponding three-screens score, and for points above the dashed line, the three-screens score is greater than the corresponding one-screen score.
of younger and 60% of older drivers who answered this question preferred the three-screen over the one-screen setup. Participants liked having the side views and felt that it was more comfortable and realistic. When asked about the overall realism of the simulator, 14 respondents (87.5%) in the younger age group felt that it was “moderate” to “very” realistic as opposed to only 4 (25%) of the older drivers. Concerns presented by drivers in both age groups included difficulties adjusting to the braking and acceleration of the simulator (n = 7 younger, 8 older) and the sensitivity of the steering wheel when turning (n = 7 younger, 12 older). Some participants also identified concerns with the physical environment, such as poor positioning of the monitors (n = 2 younger, 3 older), feeling too close to the screens (n = 3 younger, 2 older), and misalignment of the steering wheel and seat (n = 3 older). Six participants (37.5%) in the younger group and 10 (62.5%) in the older group mentioned that they felt the simulator required more effort than regular driving.

When we asked participants about their experiences on the simulated drives, most participants in the younger group (n = 13, 81.3%) felt that the simulated drives provided a reasonable reflection of their driving skills, whereas fewer than half (n = 6, 42.9%) of the older group felt this way (2 participants in the older group did not answer this question). When queried about whether they thought better drivers on the road would also be better drivers on the simulator, 9 participants in the younger group and 5 in the older group indicated yes; 5 in the younger group and 6 in the older group responded no, and the rest were undecided. We also asked participants whether they thought that an evaluation on a driving simulator could enhance current procedures for examining fitness to drive. Seventy-five percent of the younger participants and 56.3% (n = 9) of the older participants indicated that they felt this would be the case. We further found that participants (n = 12 younger, n = 10 older) felt that the simulator would be both

Table 2. Correlations Between Driver Errors and Study Instruments on One- and Three-Screen Drives

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Manitoba Road Test Score</th>
<th>Simulator-Computed Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 Screen</td>
<td>3 Screen</td>
</tr>
<tr>
<td>NASA TLX</td>
<td>0.39 [−0.02, 0.67]</td>
<td>0.27 [−0.13, 0.59]</td>
</tr>
<tr>
<td>Computer Comfort</td>
<td>−0.23 [−0.57, 0.18]</td>
<td>−0.16 [−0.52, 0.24]</td>
</tr>
<tr>
<td>SSQ</td>
<td>−0.12 [−0.49, 0.28]</td>
<td>−0.19 [−0.54, 0.21]</td>
</tr>
</tbody>
</table>

Note. The data from participants who completed both drives are presented (n = 26). Five participants who experienced simulator discomfort did not complete both drives, and driving data from 1 participant were incorrectly recorded and not usable. Cases in which the 95% CI does not include 0 are statistically significant at the .05 level; cases in which the 95% CI does include 0 are not statistically significant, p > .05. NASA TLX = National Aeronautics Space Administration Task Load Index; SSQ = Simulator Sickness Questionnaire.
an acceptable and useful training and teaching tool for all age groups to improve driving abilities. Note that 7 of the participants in the younger group felt that older adults (i.e., those >65 yr old) would have difficulty with the simulator because of unfamiliarity with computers, but only 4 participants in the older group identified this as a concern.

Discussion

Our findings add further support for the use of one-screen simulators (Lee, Cameron, & Lee, 2003; Lee & Lee, 2005; Lee, Lee, & Cameron, 2003; Lee, Lee, Cameron, & Li-Tsang, 2003; Lemieux et al., 2014). Drivers scored statistically significantly fewer errors on the three-screen simulator on three variables (i.e., Signal Violation/Right of Way/Inattention, Moving in a Roadway, Turning) and on the total score from both the MRT and the simulator-generated number of errors. One possible explanation for this difference is that all participants completed the one-screen scenario before the three-screen scenario, and they improved with practice; only a fully counterbalanced design could verify whether a practice effect exists independent of the number of screens. With the exception of Signal Violation/Right of Way/Inattention and Turning, the ICCs for all variables from both the MRT and the simulator-generated errors indicated moderate to strong agreement between the scores obtained on the two drives.

Mayhew et al. (2011) examined differences between one- and three-screen drives for younger drivers and found no performance discrepancies between the two. That being said, the discrepancies between the two setups illustrate the importance of carefully adjusting the scoring algorithms to reflect the differences. A person may score differently if tested on two different simulators, and a score used to make a fitness-to-drive determination on one setup cannot be assumed to be the same on another.

Moreover, we found no meaningful correlation between measures of workload, computer comfort, simulator discomfort symptoms, and performance on the simulator whether it was a one- or a three-screen setup. This finding is important because performance on the simulator should be independent of these variables. However, participants who volunteered for the study, particularly those in the older age group, may be more computer literate and thus more comfortable with such technology than the typical older adult.

Some participants experienced simulator discomfort while operating the simulator, but only 5 were unable to complete the two drives. Precautions can be taken to minimize the likelihood of simulator discomfort. Stern et al. (2006) have made recommendations that involve person, environment, occupation, and performance factors. Person factors include requesting participants arrive comfortably full and rescheduling if participants are experiencing headache, nausea, dizziness, gross fatigue, or vertigo before the session. Environment factors include cool room temperature, low lighting, and a spacious simulator setup. Occupation factors include decreasing the road texture and starting with simple drives. Performance factors include explaining controls and displays before the orientation drive and outlining the protocol should simulator discomfort occur.

Although evidence exists that older adults experience simulator discomfort more frequently than younger drivers (Classen, Bewernitz, & Shechtman, 2011), research with older adults driving on a three-screen STISIM simulator has revealed that the incidence of simulator discomfort is low (i.e., approximately 11% in a sample of 284 adults aged 60–99 yr; Freund & Green, 2006). We have also established that older drivers who are unable to complete a simulated drive because of simulator discomfort are not the drivers who have the poorest on-road driving performance (Mullen, Weaver, Riendeau, Morrison, & Bédard, 2010). Hence, simulator discomfort should not prevent assessment of drivers who most need evaluation. Moreover, in this study we found that mild simulator discomfort symptoms (e.g., dizziness, discomfort) did not appear to affect simulated driving performance; other researchers have found similar results (Rizzo, Sheffield, Stierman, & Dawson, 2003). Therefore, simulator discomfort should not affect driver evaluation.

Our findings should be interpreted in light of participants’ preference for the three-screen setup. Some participants indicated that they found the simulator unrealistic, particularly those in the older age group. Participants further identified areas for improvement (e.g., alignment of the seat and wheel, the touchiness of the wheel). These comments provided valuable feedback that can be incorporated to improve the simulator experience at minimal or no additional cost. Regardless of the shortcomings identified, we were encouraged that the majority of participants felt that a simulator would be an acceptable training and teaching tool. The inclusion of middle-aged drivers provided data on how future older drivers may feel about the use of driving simulators to assess fitness to drive. The findings from the interview portion of the study indicated that in general this group was amenable to the simulator experience and felt that it provided a reasonable reflection of their driving ability.
Limitations and Future Research

Our participant group comprised people interested in driving research; thus, our findings may not be representative of the general population. In addition, some participants may have previously completed sessions on the driving simulator for other studies within our lab and may have been more comfortable taking part in driving simulator assessments. Future research should focus on developing valid evaluation algorithms for driving simulators to enrich current driving evaluations. We are conducting a multisite study of a three-tiered testing process that includes one-screen simulator assessments to assess driving fitness in older adults.

Implications for Occupational Therapy Practice

The results of this study have the following implications for occupational therapy practice:

- Achieving a balance between maintaining the driving privilege and ensuring public safety requires a fair and equitable means of assessing driving capacity—driving simulators may assist in this endeavor.
- As indicated by similar scores on both the one-screen and the three-screen setups, a one-screen simulator may meet the needs of occupational therapists completing driving assessments.
- Both middle-aged and older adults may be amenable to the use of a driving simulator to evaluate driving ability. ▲

Acknowledgment

Michel Bédard was a Canada Research Chair in Aging and Health (http://www.chairs.gc.ca) at the time of this work and acknowledges the support of the Canada Research Chair Program.

References


