Comparative sensitivity of a simulated driving task to self-report, physiological, and other performance measures during prolonged wakefulness

J. Todd Arnedt*, M. Ainsley C. Geddes, Alistair W. MacLean

Department of Psychology, Queen’s University, Kingston, ON, Canada

Received 28 April 2003; accepted 12 May 2004

Abstract

Objectives: The objectives of this study were to compare (1) the sensitivity of simulated driving to self-report measures, nocturnal sleep latency tests (SLTs), and an auditory vigilance task and (2) urban and motorway driving. Methods: Healthy males 18 to 35 years maintained wakefulness for one night and were tested at 2400, 0230, 0500 and 0730 h. In Study 1 (n = 11), the SLTs were followed by auditory vigilance and simulated driving tasks; in Study 2 (n = 18), the SLTs were preceded and followed by simulated driving on motorway and urban routes. Results: In Study 1, speed variability, tracking variability, and driving off the road on the driving simulator had comparable sensitivity to d’ on the auditory vigilance task. In Study 2, driving performance was consistently worse on the motorway route. Conclusion: The driving simulator was equally sensitive to another performance measure during prolonged wakefulness and impairments were greater with motorway driving.

Keywords: Prolonged wakefulness; Sleepiness; Driving; Simulated driving; Sensitivity

Introduction

The contribution of sleepiness to large-scale disasters, errors in the workplace, and motor vehicle accidents has been increasingly recognized and studied [1–3]. Conditions that have been associated with an increased risk for sleep-related accidents include shift work, the use of alcohol or other drugs, and sleep disorders, in particular, obstructive sleep apnea, narcolepsy, and chronic insomnia [4]. Estimates of the number of automobile crashes attributable to sleepiness range widely, but there is converging evidence that from 16% to 23% are sleep-related [2,5]. Among the demographic and driving behaviour characteristics associated with increased risk for fall-asleep crashes are younger age, male gender, higher education level and annual mileage driven, and more time spent on motorway roads [6–8].

Clearly, there exists a need to develop measures adequate to detect the presence of sleepiness. Currently, introspective estimates and electrophysiological measures, such as the Multiple Sleep Latency Test [9], are the primary instruments used for this purpose. However, Dinges and Kribbs [10] have argued that measures of functional capacity might be more suitable for detecting the consequences of increased sleepiness. A number of tests are well known to be sensitive to conditions provoking sleepiness, such as sleep loss, and the characteristics of such tests have been widely described. These characteristics include long duration [11,12], no knowledge of results [13], complex yet uninteresting [14], monotonous [15], familiar [13], and experimenter paced [14,16]. These features have been used for the development of a number of performance tests, including vigilance [11], simple [17] and choice [18] reaction time, logical reasoning [19], and memory [20].

Broadbent [21] has suggested that a battery of tests assessing a variety of task-relevant characteristics should be
the approach of choice when assessing complex perform-
ance. Such tasks are generally well tolerated by student
subjects; however, patients frequently indicate that they lack
face validity and may therefore be less inclined to give their
best performance or may refuse to participate altogether. In
addition, these tests frequently have large practice effects,
making them impractical in clinical settings [22,23], and
their relative sensitivity to conditions causing performance
decrements in comparison with other measures of sleepiness
is unknown. The ideal performance measure would have
high face validity, involve a highly practised skill, and
assess a number of relevant aspects of functional capacity.
For these reasons, a driving simulator provides a possible
alternative to the use of a test battery. Driving simulators of
various levels of sophistication have been used to examine
the consequences of a variety of conditions, including sleep
depression [24,25], alcohol ingestion [24,26,27], and sleep
disorders [28,29].

The York Driving Simulator (York Computer Technolo-
gies, Kingston, ON, Canada) is a low-cost, portable driving
simulator that offers a flexible data output, allowing for the
measurement of a number of driving-related constituent
skills, such as vehicle control (e.g., steering and pedal
control) and on-the-road measures [e.g., lane position
(tracking) and speed deviation]. Such a device could prove
valuable in a clinical setting to provide objective information
regarding a patient’s ability to operate a motor vehicle safely.
Currently, most clinicians rely on self-report and physiolog-
ical measures, such as the Multiple Sleep Latency Test, to
guide such decisions. There seems considerable potential,
however, for the use of performance measures to provide a
more objective and relevant assessment. Driving perform-
ance on the York Driving Simulator has previously been
found to deteriorate following partial sleep deprivation (4.5-h
time in bed) and to ameliorate with slow-release caffeine
[30]. In addition, minimal practice effects on this simulator
were found in a sample of young participants with little
driving experience [31]. We sought to determine the degree to
which the York Driving Simulator is sensitive to the effects of
prolonged wakefulness in comparison with other self-report,
physiological, and behavioural measures of sleepiness.

This paper describes two studies. The purpose of the first
study was to compare the sensitivity of the York Driving
Simulator to other measures known to be sensitive to
sleepiness and to decrements in performance. The aim of the
second study was to compare the decrements occurring
while driving under urban and motorway conditions to
evaluate more closely the degree to which task character-
istics affect simulator performance.

Study 1

The aim of this study was to compare the sensitivity of
the York Driving Simulator to self-report measures, repeated
nocturnal sleep latency tests (SLTs), and a 30-min auditory
vigilance task under conditions likely to produce perform-
ance decrements.

Methods

Participants

Eleven male participants recruited from the Introductory
Psychology Subject Pool (mean age = 20.4 years, S.D. = 2.2,
range 18–25; mean Epworth Sleepiness Scale [32] score 8.6,
S.D. = 4.7, range 2–18), who gave their informed consent,
experienced one night of prolonged wakefulness. Inclusion
criteria were a regular sleep schedule, neither extreme
evening nor morning type [33], good health, the absence of
medications that affect daytime sleep schedule, neither extreme
evening nor morning type, and willingness to comply with the
conditions of the experiment. One participant had an Epworth Sleepiness
Scale score of 18, which exceeds the clinical cutoff for
excessive daytime sleepiness [32]. Participants were asked to
maintain a regular sleep schedule, with bedtime between
2300 and 0100 h and waking between 0700 and 0900 h, for
the week prior to the study and to refrain from the use of
caffeine for 24 h and alcohol and other drugs for 48 h prior to
the study. Compliance with the protocol was confirmed using
sleep logs and an activity/substance use diary. Consistent
with the policies of the Department of Psychology, partic-
ipants earned a 1% increase to their class mark in
Introductory Psychology for each 1 h of participation in
departmental research activities, up to a maximum of 5%. Ac-
Accordingly, for this study, they were awarded an extra 5% to
their class mark. Of 14 participants who volunteered for the
study, two did not meet the inclusion criteria and one became
ill on the test night for reasons unrelated to the experiment.

Dependent measures

Self-report

The Stanford Sleepiness Scale [34] and a modified form
of the Stanford Scale [35] were used to assess subjective
sleepiness. The latter test yields two factors provisionally
described as, respectively, “an energetic or activating factor”
(Factor 1; range 0–13) and a factor “related to conscious-
ness, sleepiness, and a loss of control over remaining awake”
(Factor 2; range 0–11; Ref. [35], p. 38). Higher scores on
both factors indicate higher levels of subjective sleepiness.

Physiological

Repeated nocturnal SLTs were conducted consistent with
procedures outlined for the Multiple Sleep Latency Test [9].
Continuous monitoring of electroencephalogram (EEG),
electrooculogram (EOG), and electromyogram (EMG) was
conducted for each test. The test was terminated after 2 min
of Stage 2 sleep, the first epoch of any other stage of sleep, or
after 20 min without sleep onset. The resulting records were
scored independently by two scorers, and any discrepancies
not resolved by consultation were decided by a third scorer. Electrophysiological activity was also recorded while participants performed a 30-min auditory vigilance task, and records were scored in 30-s epochs using standard procedures [36]. A score of 30 min was given to participants who showed no signs of sleep onset on this task. Consistent with guidelines for the Multiple Sleep Latency Test, the dependent variable for both physiological tests is the latency from lights out to the first epoch scored as Stage 1 sleep.

**Performance**

The auditory vigilance task [37] required participants to listen to a series of blocks of 10 tones of constant pitch (1000 Hz) with a fixed intertone interval of 5 s. Noise tones had a 1.0-s duration, while signal tones were 1.2-s long. Each had an 82-dB SPL intensity and was played against a background noise of 32 dB SPL. The signal tones occurred once in each block of 10 tones and were randomly interspersed among the nine noise tones. Participants were instructed to sit comfortably and press a microswitch with the preferred hand when they thought they heard a signal tone. The dependent variables reported here are the widely used signal detection parameters $d'$ and Beta ($\beta$). The former is typically interpreted as an individual’s intrinsic ability to discriminate signal from noise tones, with higher scores representing greater sensitivity to signal tones. On the other hand, $\beta$ is an index of an individual’s decision criterion, or degree of caution, used to respond to a signal. Higher $\beta$ scores indicate a more cautious response style.

The York Driving Simulator runs on a personal computer system, and an electronic interface provides driver inputs to the software (steering, brakes, and accelerator). Simple line graphics are used to portray objects and to induce a sense of apparent motion by using real-time perspective generation. The simulator presents a forward view from the driver’s seat of either an urban or a motorway road scene (Fig. 1), with standard lane markings and signs and signals appropriate for the specified road environment. The participant navigates within the computer driving environment interactively just as one would navigate a car in the real world, while being told to obey all road signs and traffic signals and to maintain the car in the appropriate lane on the road. The participants drove in a simulated night environment and were instructed to drive as far as they could in 30 min while keeping the car’s position in the middle of the right hand lane and obeying traffic lights, speed restrictions, stop signs, and other road signs. Each 30-min drive consisted of an area of urban driving followed immediately by motorway driving, each section taking approximately 15 min.

Vehicle and road dimensions in the York Driving Simulator are measured in simulator units, where 1 ft (0.3048 m) = 100 simulator units (su). The simulated car was 600 su in width and 1000 su in length. The entire simulated road is 3000 su wide; each lane on the two-lane road is 1100 su wide, and 400 su on either side is devoted to the road shoulder.

**Procedure**

Prior to the experiment, participants attended two training sessions. At each training session, they completed the self-report assessments and were given a 15-min practice trial on the auditory vigilance task and a 20-min trial on the driving simulator. At the first training session, the participants also had electrodes attached as they would for the SLTs. Prior to the first test session on the experimental night, participants were given a further 30-min practice trial on the driving simulator and a further demonstration of the auditory vigilance task.

Test sessions took place at 2400, 0230, 0500, and 0730 h; each session took about 90 min to complete. The participants came to the laboratory about 2 h prior to the first session to be prepared for continuous recording of EEG (C3/A2, O2/A1), referential EOG, and submental EMG activity. They remained awake under supervision in the laboratory, indulging in quiet activities between test sessions. At each test session, the SLT was given first followed by the two performance tests (auditory vigilance and simulated driving tests); the order of the latter two tests was counterbalanced across test sessions. Self-report sleepi-
ness ratings were obtained prior to each SLT and each performance test.

Results

In general, the data were analysed using repeated measures analysis of variance, in which time of testing (2400, 0230, 0500, and 0730 h) was extracted as a within-subject factor. For the Stanford and Modified Stanford Sleepiness Scales, a second within-subject factor was also extracted corresponding to the time of administration (prior to the SLT and the two performance tests). In the case of the auditory vigilance and the simulated driving tasks, time-on-task was extracted by analysing the data in 5-min blocks. Where appropriate, Huynh–Feldt [38] corrections for lack of sphericity were applied, and the data were further analysed using orthogonal polynomials or the Tukey Honestly Significant Difference Test. The level of significance for all analyses was set at .05.

Sensitivity of tests to experimental conditions

The summary statistics for all of the measures at the four test sessions are reported in Table 1.

Self-report

With increasing time awake, participants reported themselves as increasingly sleepy on the Stanford Sleepiness Scale \(F(3,30)=16.23, P=.0001\) and on both Factors 1 \(F(3,30)=15.07, P<.0005\) and 2 of the Modified Stanford Scale \(F(3,30)=8.58, P=.003\). There was also a small, but statistically significant, difference between reports of subjective sleepiness [Stanford Sleepiness Scale: \(F(2,20)=26.33, P<.0001\]; Modified Stanford Scale: Factor 1: \(F(2,20)=20.00, P<.0005\); Factor 2: \(F(2,20)=12.26, P<.0005\)] before the SLT and the performance tests, but no difference in sleepiness ratings between the two performance tests. For example, for the Stanford Sleepiness Scale, the mean values before the SLT and the auditory vigilance task were 4.1, 4.8, and 4.7, respectively.

Physiological measures

On the SLT, latency to Stage 1 sleep decreased with increasing time awake \(F(3,30)=15.80, P<.0005\), as was true also for the auditory vigilance task \(F(3,30)=6.31, P=.002\).

Performance

On the auditory vigilance task, there was a decline in \(d'\) with increasing time awake \(F(3,30)=15.70, P=.0001\) and time-on-task \(F(5,50)=9.05, P=.0001\). Beta did not change significantly with either increasing time awake \(F(3,30)=0.49, P>.05\) or time-on-task \(F(5,50)=0.94, P>.05\).

With increasing time awake on the simulated driving task, the drivers drove increasingly to the left of the centre of their lane \(F(3,30)=6.81, P<.005\), the variability of their tracking increased \(F(3,30)=9.54, P<.005\), the speed variability increased \(F(3,30)=4.96, P<.009\), and the number of times that they drove off the road increased linearly \(F(1,9)=7.70, P=.022\). With increasing time-on-task during each session, drivers again tracked to the left of the lane centre \(F(5,50)=7.78, P=.01\), and tracking variability increased \(F(5,50)=16.72, P<.005\). At the beginning of the task, drivers tended to drive below the speed limit, but, as task duration increased, they increased their speed until they were driving above the posted limit \(F(5,50)=5.04, P<.03\).

Comparative sensitivity

To compare the various dependent measures on a common scale, Cohen’s standardised difference measure [39] was used. This was defined as (mean at other

<table>
<thead>
<tr>
<th>Test time</th>
<th>2400 h</th>
<th>0230 h</th>
<th>0500 h</th>
<th>0730 h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-report measures</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stanford Sleepiness</td>
<td>Mean 3.5</td>
<td>S.E.M. 0.4</td>
<td>4.2</td>
<td>5.0</td>
</tr>
<tr>
<td>Modified Stanford Sleepiness Scale: Factor 1</td>
<td>Mean 2.0</td>
<td>S.E.M. 0.9</td>
<td>3.2</td>
<td>5.1</td>
</tr>
<tr>
<td>Modified Stanford Sleepiness Scale: Factor 2</td>
<td>Mean 9.0</td>
<td>S.E.M. 1.0</td>
<td>10.7</td>
<td>12.1</td>
</tr>
<tr>
<td>Physiological measures</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sleep latency on the SLT (min)</td>
<td>Mean 13.8</td>
<td>S.E.M. 1.9</td>
<td>6.4</td>
<td>4.4</td>
</tr>
<tr>
<td>Sleep latency during the Auditory Vigilance Test (min)</td>
<td>Mean 28.0</td>
<td>S.E.M. 1.4</td>
<td>23.8</td>
<td>22.1</td>
</tr>
<tr>
<td>Tracking deviation from lane centre (m)</td>
<td>Mean 2.63</td>
<td>S.E.M. 0.28</td>
<td>2.02</td>
<td>0.27</td>
</tr>
<tr>
<td>Tracking variability (m)</td>
<td>Mean 3.36</td>
<td>S.E.M. 0.53</td>
<td>3.84</td>
<td>0.46</td>
</tr>
<tr>
<td>Deviation from posted speed (km/h)</td>
<td>Mean 13.8</td>
<td>S.E.M. 2.0</td>
<td>6.5</td>
<td>3.2</td>
</tr>
<tr>
<td>Variability of speed deviation (km/h)</td>
<td>Mean 9.0</td>
<td>S.E.M. 1.0</td>
<td>10.7</td>
<td>0.4</td>
</tr>
<tr>
<td>Off-road incidents (number/5 min)</td>
<td>Mean 2.0</td>
<td>S.E.M. 0.1</td>
<td>0.9</td>
<td>1.0</td>
</tr>
</tbody>
</table>

### Table 1

Mean and standard error of the mean (S.E.M.) on the self-report, physiological, auditory vigilance, and simulated driving measures across the four test times in Study 1.
times—mean at 2400 h)/pooled standard deviation. As suggested by Cohen [39], a standardised difference of 0.2 was considered small, 0.5 was considered medium, and 0.8 was considered a large difference. The results are shown in Fig. 2.

Discussion

While the driving simulator was not the most sensitive test, it did appear to reach a level of sensitivity sufficient to warrant further investigation. These findings are considered in more detail in the General Discussion below. However, it was first necessary to resolve an important confound in the present study. An examination of the driving simulator data for time-on-task effects indicates that the greatest change in performance occurred during the second half of the 30-min test. This happened to be the time when participants were driving on the motorway route. There is thus a confound between driving conditions and time-on-task. To determine which of these factors is responsible for the decline in performance, a second study was carried out as described below.

Study 2

The aim of the second study was to eliminate the confound of route and task length by counterbalancing the order of presentation of urban and motorway routes.

Methods

Participants

Participants were recruited from the Introductory Psychology Subject Pool. Twenty healthy males, who agreed to comply with the requirements of the study, gave their informed consent to participate. Inclusion criteria were a valid driver’s license, a regular sleep schedule, a moderate to good sleeper [40], neither extreme morning nor evening type ([33]; mean 48.9, S.D. = 8.7, range 32–62), no evidence of excessive daytime sleepiness ([32]; mean 6.8, S.D. = 2.3, range 4–12), and no medications with sleep-altering qualities. One participant included in the sample had an Epworth score above the clinical cutoff for excessive daytime sleepiness [32]. Of the original 20 participants, two participants were excluded from the analyses because of failure to comply with the instructions, resulting in a final sample of 18 participants (mean age = 21.9 years, S.D. = 3.5, range 19–32 years).

Participants were instructed to refrain from using alcohol and other drugs for 48 h and caffeine for 24 h prior to the experimental night. They were also asked to maintain a regular sleep schedule, retiring between 2300 and 0100 h and rising between 0700 and 0900 h, beginning 48 h prior to testing. Compliance with the conditions was assessed using sleep logs and an activity/substance use diary. Participants were either paid CAD$30 or awarded a 5% addition to their overall mark in the Introductory Psychology course for participation.
Dependent measures

Self-report

The self-report variables were the same as those used in Study 1.

Physiological

The SLT procedure was as described in Study 1. Electrophysiological recordings were also made during the simulated driving task; however, these data will not be reported here. As in Study 1, the dependent variable was the latency to Stage 1 sleep.

Performance

The York Driving Simulator was used to assess driving performance. The urban and motorway routes were similar to those used previously. As in Study 1, the dependent variables were mean tracking, tracking variability, speed deviation, speed variability, and the number of off-road occurrences.

Procedure

Forty-eight hours prior to testing, the participants came to the Sleep Laboratory for a 30-min training session on the driving simulator, 15 min each on the urban and motorway routes. A second similar training session took place at the beginning of the experimental night.

On the experimental night, participants arrived at the Sleep Laboratory between 2100 and 2130 h, where electrodes were applied for continuous monitoring of EEG (C3/A2, O2/A1), EOG, and EMG activity. Following training, four 90-min test sessions took place at 2400, 0230, 0500, and 0730 h. For each test session, the participants drove for 30 min on either the urban or motorway route of the driving simulator, underwent the SLT, and then drove the other route for another 30 min. The order of the urban and motorway routes was counterbalanced across participants. Participants completed the Stanford Sleepiness and the Modified Stanford Scales before and after each simulated drive, and before each SLT. During breaks, the participants engaged in quiet activities in the laboratory under supervision. At the conclusion of testing, the participants were debriefed and remunerated for their participation.

As part of a protocol separate from this study, 10 randomly selected participants wore a behavioural device, the Magellan Monitor, on their nondominant wrist throughout the experimental night, except during the SLT. The monitor emitted vibratory stimuli, on average, every 45 s, with a minimum and maximum interstimulus interval of 30 and 60 s, respectively. The participants were required to respond to the stimuli as quickly as possible by flexing their index finger to activate a piezoelectric response sensor.

Statistical analyses

In general, the data from the four test sessions were analysed using split plot analyses of variance, with the wearing of the Magellan Monitor (Magellan, no Magellan) extracted as a between-subjects source of variance, and test session (2400, 0230, 0500, and 0730 h) extracted as a within-subjects source of variance. For the self-report measures, two analyses were conducted. The first extracted the time of questionnaire administration (before and after the driving sessions and before the SLT) as an additional within-subjects factor to examine self-reported changes in sleepiness with prolonged wakefulness. The second analysis extracted route (motorway and urban) and time of administration (before and after the driving sessions) as additional within-subjects factors. For the simulated driving task, route (motorway and urban) and the six 5-min blocks of the task were extracted as within-subjects effects.

Univariate F ratios were evaluated using the Huynh–Feldt correction factor for sphericity [38]. Orthogonal polynomials and Bonferroni-adjusted t tests were used where appropriate. The level of significance for all analyses was set at .05.

Results

Effects of prolonged wakefulness

Self-report

With increasing time awake, participants rated themselves as more sleepy on all three self-report sleepiness measures: the Stanford Sleepiness Scale \( F(3,48) = 86.35, P < .001 \) and Factors 1 \( F(3,48) = 40.69, P < .001 \) and 2 \( F(3,48) = 36.68, P < .001 \) of the Modified Stanford Scale. Post hoc analyses indicated that self-reported sleepiness ratings on all three scales increased significantly from 2400 to 0730 h, with significant differences between ratings at each successive test time, except between 0500 and 0730 h \( P < .001 \). There were also significant time effects on the Stanford Sleepiness Scale \( F(2,32) = 21.71, P < .001 \) and on Factors 1 \( F(2,32) = 12.78, P < .001 \) and 2 \( F(2,32) = 47.27, P < .001 \) of the Modified Stanford Scale. In all cases, ratings were significantly higher before the SLT and after the driving task than before the driving task \( P < .01 \). Sleepiness ratings before the SLT for Factor 2 of the Modified Stanford Scale were also significantly higher than after the simulated driving task \( P = .013 \). There were no significant Time×Magellan interactions or three-way interactions involving test, time, and Magellan, but a significant Test×Magellan interaction was found for Factor 2 of the Modified Stanford Scale \( F(3,49) = 3.41, P < .05 \), and Test×Time interactions were evident for both the Stanford Sleepiness Scale \( F(6,96) = 2.92, P < .05 \) and...
Factor 2 of the Modified Stanford \[ F(6,96)=3.65, P<.01\].

**Physiological measures**

On the SLT, sleep onset latency to Stage 1 decreased with increasing time awake \[ F(3,48)=26.97, P<.001 \]. There were no significant main effects of Magellan or Test $\times$ Magellan interactions.

**Driving performance**

There were no main effects of wearing the Magellan monitor on any of the dependent variables of the driving task. Consistent with the findings from Study 1, with increasing time awake, the participants drifted to the left of the centre of the lane \[ F(3,48)=10.14, P<.001 \], tracking became more variable \[ F(5,80)=16.97, P<.001 \], speed variability increased \[ F(3,48)=21.80, P<.001 \], and off-road incidents increased linearly \[ F(1,16)=4.85, P<.05 \]. Post hoc analyses indicated that mean tracking, tracking variability, and speed variability were worse at the 0730-h test time relative to both the 2400- and 0230-h test times \( P<.003 \). Additionally, mean tracking and speed variability were worse at 0730 h compared with the test at 0500 h \( P<.008 \). There was a significant Test $\times$ Magellan interaction only for speed variability \[ F(3,48)=3.78, P<.05 \], which revealed worse performance at the 0730-h test time in the group wearing the device. Performance on the following dependent variables also deteriorated with increasing time-on-task: mean tracking \[ F(5,80)=5.17, P=.001 \], tracking variability \[ F(5,80)=10.95, P<.001 \], speed variability \[ F(5,80)=3.55, P<.01 \], and off-road occurrences \[ F(5,80)=3.50, P<.05 \]. Mean tracking \[ F(15,240)=2.73, P<.01 \], tracking variability \[ F(15,240)=4.31, P<.001 \], speed variability \[ F(15,240)=4.27, P<.001 \], and off-road events \[ F(15,240)=2.74, P<.05 \] deteriorated more rapidly during the 30-min driving task as the night progressed. There were no significant Magellan $\times$ Block interactions, but speed deviation \[ F(15,240)=2.69, P<.01 \] and speed variability \[ F(15,240)=1.71, P=.05 \] deteriorated more rapidly across the 30-min driving task at the later test times for participants wearing the Magellan monitor.

**Effects of route type on measures of driving performance**

Descriptive data for the self-report measures and the simulated driving task with effect sizes (Cohen’s $d$) are presented in Table 2.

**Self-report**

Ratings on the three self-report sleepiness measures did not differ by route type, but ratings were consistently higher after the driving task than before: Stanford Sleepiness Scale \[ F(1,16)=80.30, P<.001 \] and Factors 1 \[ F(1,16)=29.90, P<.001 \] and 2 \[ F(1,16)=94.69, P<.001 \]. There were no significant two- or three-way interactions involving Magellan, route, and time.

**Driving performance**

The significant differences in driving parameters between the urban and motorway routes are shown in Fig. 3.

Performance was significantly worse on the motorway route for all five simulated driving variables: mean tracking \[ F(1,16)=8.35, P=.01 \], tracking variability \[ F(1,16)=26.23, P<.001 \], mean speed deviation from the posted speed signs \[ F(1,16)=70.38, P<.001 \], speed variability \[ F(1,16)=148.33, P<.001 \], and the number of off-road incidents \[ F(1,16)=15.47, P<.001 \]. Across the six 5-min blocks of the driving task, off-road events \[ F(5,80)=2.53, P<.05 \] increased more rapidly on the motorway route. There were no Route $\times$ Test or Magellan $\times$ Route interactions, but a significant three-way interaction involving Magellan, route, and block was found for off-road incidents \[ F(5,80)=2.66, P<.05 \].

**General Discussion**

The primary objective of the first study was to evaluate the sensitivity of a face valid measure of driving ability, the York Driving Simulator, relative to measures that are widely used and known to be sensitive to the effects of sleep loss. The second study was conducted to clarify an important confound in the first study, as well as to examine more closely the degree to which task characteristics (in particular, monotony) affect simulated driving performance in a manner consistent with other performance measures.

**Sensitivity of the York Driving Simulator**

By placing the various sleepiness measures on a common metric in Study 1, it was possible to compare their relative sensitivity to the experimental manipulation. The findings
were generally consistent with the model of sleepiness outlined by Carskadon and Dement \cite{41}. Sleep latency on the SLT was the most sensitive measure, presumably due, in large part, to the fact that the test is conducted in an environment that is explicitly devoid of external stimulation—the lights are dimmed, the participant is supine, noise levels are low, room temperature is constant—and the participant is instructed to try to fall asleep. As such, the SLT is the most direct measure of physiological or basal sleepiness. The four parameters of the York Driving Simulator that showed significant deterioration with increasing time awake in both studies—mean tracking, tracking variability, speed deviation, speed variability, and off-road incidents—were considerably less sensitive than the SLT and slightly less sensitive than the other performance measure, $d'$ on the auditory vigilance task. These findings too would be predicted by the model of Carskadon and Dement \cite{41}. The auditory vigilance task was administered while the participant sat upright in a chair and had to make a simple motor response; in contrast, the driving task required sustained attention to a more complex stimulus and the manipulation of steering wheel and pedals.

**Self-report and performance measures**

As expected, self-reported sleepiness increased monotonically across the night of prolonged wakefulness in both studies. These findings are consistent with several studies.

---

![Graphs of mean tracking, tracking variability, speed deviation, speed variability, and off-road events](image)

*Fig. 3. Comparison of overall driving performance on urban and motorway routes of the York Driving Simulator. Results for the five dependent measures of the driving task—mean tracking, tracking variability, speed deviation, speed variability, and off-road incidents—are plotted as means±S.E.M.s.*
showing increased subjective sleepiness during the night assessed in experimentally induced sleep deprivation studies (e.g., Ref. [42]), constant routine protocols (e.g., Ref. [43]), and studies of shift workers in a variety of industries [44,45]. In addition, it has been shown repeatedly that alertness follows a self-sustained circadian pattern reaching its nadir in the early morning around 0500 h [46,47]. Interestingly, although objective performance was consistently worse on the motorway route of the driving task, there were no significant differences in self-reported sleepiness between the motorway and urban routes. This finding suggests that the degree of impairment associated with sleepiness may be underrecognized, which has implications for decisions about one’s capacity to operate safely a motor vehicle. The inability to appreciate fully the dangers associated with drowsy driving has also been highlighted by other researchers [48].

Consistent with previous studies [49,50], latency to Stage 1 sleep onset declined linearly across the night on the SLT.

The decline in $d'$ on the auditory vigilance task with increasing time awake is consistent with previous findings and has typically been interpreted as indicating a reduction in the intrinsic capacity to discriminate a signal [51–53]. Naitoh [54] has cautioned against a simple interpretation of the signal detection measures $d'$ and $\beta$ and has questioned the appropriateness of applying signal detection theory to sleep loss studies. He argues that independent evidence, such as physiological measures, is necessary to ensure “that the subject observe and attend to the stimuli” (Ref. [54], p. 360). Because no such evidence was gathered in this study, the findings may indicate either a true reduction of intrinsic capacity or a loss of attentional capacity.

*Simulated driving performance during prolonged wakefulness*

Simulated driving performance deteriorated progressively across the night in both studies. More specifically, with increasing time awake, participants drove closer to the middle of the road, tracking and speed became more variable, and the number of off-road incidents increased. This overall decline in driving ability during periods of increased sleepiness is consistent with the findings from both simulated driving [55] and nondriving performance studies conducted during the circadian trough of alertness [56,57]. The findings from the present studies are also consistent with the well-identified nighttime peak for sleepiness-related motor vehicle accidents [8,58,59] and with the occurrence of several large-scale catastrophic events during the early morning hours [3]. One significant shortcoming with the current study is that the experimental design did not permit a systematic analysis of the relative contribution of circadian and homeostatic influences to the observed decrements in performance across the night. This is an important issue, and one that could be addressed in future studies experimentally by, e.g., testing participants at the same circadian phase, but altering the length of wakefulness prior to testing.

In addition to the deficits observed across the night, there were characteristic changes in driving performance within the 30-min driving task. With increasing time-on-task, participants again drifted towards the middle of the road, and their tracking became more variable. These findings are consistent with previous studies showing deviations in lateral position as a result of prolonged driving [60,61]. Although the driving task used in the present studies was considerably shorter (30 min) in duration than those in the latter two studies (2 h), it is likely that time-on-task fatigue was a significant contributor to the observed deterioration in performance. Others have noted that fatigue is manifested more rapidly in driving simulators than in actual driving conditions (e.g., Ref. [55]), and participants in the present study rated themselves as significantly sleepier after the driving task than before. From a practical perspective, a test that is sensitive to detecting the decrements related to sleepiness and of a relatively short duration is more likely to be preferred by both clinicians and patients.

Contrary to the present findings, Lenné et al. [25,55] have found that participants drifted towards the outer edge rather than towards the midline of the road with increasing task duration. Their simulated driving task may have been more alerting, however, because participants were required to perform a secondary reaction time task while driving, and there were other simulated cars on the road. In a more monotonous driving task without other simulated cars, the sleepy driver may approach the road midline in an effort to avoid an off-road incident. Stewart and MacLean [62] have also observed that when other cars are introduced, participants drive towards the outer edge of the road, perhaps, to reduce the likelihood of a collision with another vehicle.

*Effects of route type on driving performance*

The results from Study 2 demonstrated that driving performance was consistently worse on the motorway route. More specifically, participants’ ability to maintain a consistent road position and a constant speed was more variable, their speed deviated more from the posted speed limit, and they drove off the motorway road more frequently. These findings may have been due to the greater monotony of the motorway route, which has few turns and no road signs or stop lights. In addition to being consistent with predictions, the deterioration in driving performance is supported by accident data demonstrating a 20–25% higher incidence of sleep-related motor vehicle accidents on motorway than on urban routes [2,48]. Studies of the effects of sleep loss on performance have consistently shown that monotonous tasks are more sensitive to the effects of sleepiness than are intrinsically more interesting tasks [11,13,15,63]. Moreover, the greater decline in performance on the motorway route is consistent with the model of Carskadon and Dement [41] of the constructs underlying
sleepiness because physiological sleepiness is theoretically unmasked in the absence of alerting factors [64]. The results of the second study, in particular, suggest that simulated motorway routes may be more sensitive to the decrements in performance observed with sleepiness than urban routes are and may therefore be more appropriate for assessing driving-related impairments associated with sleepiness.

Finally, half of the participants in Study 2 were required to respond to the Magellan Monitor during the driving task, which could have resulted in better performance in this group due to increased arousal. There were, however, no main effects of the Magellan on driving performance, and the only significant two-way interaction found was for performance in the group wearing the device. Thus, the Magellan seemed to have little direct effect on arousal level while driving. Comparing the concordance of the Magellan data with the other measures of sleepiness will reveal whether this device might serve as an on-board real-time indicator of a driver’s level of alertness.

In summary, the findings from both investigations suggest that the York Driving Simulator is sensitive to the decrements in performance evident with increasing time awake, and it appears to have at least comparable sensitivity under these conditions to an auditory vigilance task. Moreover, consistent with what has been demonstrated with other performance measures, increasing the monotony of the task produces an even more marked deterioration in driving performance. The simulated driving task offers advantages over self-report, physiological, and other performance measures in the assessment of driving performance, such as providing a more face valid measure of driving and using a highly practised skill. We have recently shown that performance impairments on the York Driving Simulator following prolonged wakefulness are equivalent to the decrements produced by levels of alcohol intoxication that are illegal for operating a motor vehicle [65,66]. These findings suggest that such a driving task may have clinical utility in the assessment of driving ability in patients who are at an increased risk for sleepiness-related accidents.

Acknowledgments

The support of the Ontario Ministry of Transportation (Grant No. 9505) and of Queen’s University is gratefully acknowledged, as is the technical assistance of J. Brown, S. Ferguson, and M. York.

References

[4] Lyznicki JM, Doee GC, Davis RM, Williams MA. Sleepiness, driving, and motor vehicle crashes Council on Scientific Affairs, American Medical Association [see comments]. JAMA 1998;279:1908–13.
[12] Wilkinson RT. Effects of up to 60 hours’ sleep deprivation on different types of work. Ergonomics 1964;7:175–86.
[24] Lenne MG, Triggs TJ, Redman JR. Interactive effects of sleep deprivation, time of day, and driving experience on a driving task. Sleep 1998;21:38–44.


