Jun 20th, 12:00 AM

Comparison of Driver Distraction Evaluations across Two Simulator Platforms and an Instrumented Vehicle

Susan T. Chrysler
University of Iowa, Iowa City, IA

Joel Cooper
Precision Driving Research, Inc., Salt Lake City, UT

Daniel V. McGehee
University of Iowa, Iowa City, IA

Christine Yager
Texas A&M Transportation Institute, College Station, TX

Follow this and additional works at: https://ir.uiowa.edu/drivingassessment

COMPARISON OF DRIVER DISTRACTION EVALUATIONS ACROSS TWO SIMULATOR PLATFORMS AND AN INSTRUMENTED VEHICLE

Susan T. Chrysler[1], Joel Cooper[2], Daniel V. McGehee[3] & Christine Yager[4]
[1]National Advanced Driving Simulator, University of Iowa, Iowa City, Iowa, USA
[2]Precision Driving Research, Inc., Salt Lake City, Utah, USA
[3]Public Policy Center, University of Iowa, Iowa City, Iowa, USA
[4]Texas A&M Transportation Institute, College Station, Texas, USA
Email: susan-chrysler@uiowa.edu

Summary: The purpose of this work was to assess the cross-platform validity of two driving simulators and an instrumented vehicle operated on a closed driving course. Characteristics of vehicle speed and performance to an Alert Response Task were evaluated using a MiniSim, manufactured by the National Advanced Driving Simulator group, a Realtime Technologies, Inc. desktop simulator, and an instrumented 2005 Toyota Highlander. Results indicate a high degree of relative validity between the three research platforms with mean and standard deviation of vehicle speeds showing near identical patterns under various secondary task demands. Performance on an auditory Alert Response Task also showed a high degree of consistency across the three research platforms. Performance on a visual Alert Response Task appeared to be highly reactive with the testing conditions present in the instrumented vehicle evaluations. These data have practical implications for the use of driving simulators in experimentally controlled research and also make suggestions about the use of visual warnings to elicit emergency response behaviors in drivers.

As more and more technology makes its way into vehicles, the need for simple, cost effective, reliable, and valid distraction evaluation protocols becomes increasingly important. One commonly utilized research tool that meets many of the above criteria is driving simulation. Although simulation has been widely adopted for controlled driving evaluation, only a handful of studies have ever compared results between different driving simulators (see Jamson, 2011), while many more have compared results between simulators and a real-world driving task (see Mullen, Charlton, Devlin, & Bédard, 2011 for a comprehensive review). Validating driving simulator data across platforms and with on-road data is an important step toward understanding how findings between various research platforms generalize to each other and the real world. In this paper, we present a subset of dependent measures from a series of studies that looked at the cross platform validity of three different driving research platforms. In each experiment, a near identical procedure and experimental configuration was used. This included identical participant instructions, identical secondary-task stimuli, and identical secondary task controls. This highly consistent experimental setup allowed us to compare and contrast the performance accuracy of three different research platforms in different locations. A MiniSim, manufactured by the National Advanced Driving Simulator group, was used at the Iowa site, while an Realtime Technologies (RTI) desktop simulator and an instrumented vehicle were used at the Texas Transportation Institute (TTI) locations.

Simulator validation is complicated by a number of factors and has long been a subject of active research (Blaauw, 1982; Harms, 1996; Tornos, 1998; Godley, Triggs, & Fildes, 2002; Molino,

Opiela, Katz, & Moyer, 2005). Among the important factors that should be considered are the various ways in which the fidelity of the research platforms differ and the importance that these differences might have on the obtained results. There is a modest literature in simulator validation—from speed and tracking to sign recognition. However, absolute validation can be a challenge as driver-in-the-loop simulation is often used in cases that are too dangerous to carry out on-road and so relative validation is often more realistic. According to Mullen et al., (2011), absolute validation is achieved when the data obtained on disparate research platforms is statistically indistinguishable while relative validation is attained when the data show identical sensitivity to experimental manipulation. Because of the various difficulties in establishing absolute validity, this research will explore the relative validity of two simulation platforms with an instrumented vehicle on closed course test track.

Methods

A total of 121 subjects participated in this research. A breakdown of gender and age across sites is provided in Table 1.

Table 1. Participant gender and age ranges across research studies

Location: Test Platform	N	Age Range	Gender
TTI: Realtime Technologies Desktop Simulator	40	25-55	18 Men, 22 Women
TTI: Instrumented Vehicle on Test Track	41	25-70	16 Men, 16 women (ages 25-55)
			4 Men, 5 Women (ages 55-70)
Iowa: MiniSim Desktop Simulator	40	18-35	21 Men, 19 Women

Research was conducted on three distinct platforms, two of which were driving simulators (RTI and MiniSim), the other was a 2005 instrumented Toyota Highlander driven on a closed course. The RTI desktop driving simulator used at TTI consists of 3, 22" monitors placed on a table with Logitech steering, gas, and brake controls. For this research, mirrors were disabled from view and a minimalist analog speedometer was digitally rendered to the bottom center of the forward screen. Driving scenarios were authored using SimCreator from RTI. The MiniSim desktop driving simulator used at the National Advanced Driving Simulator had 3, slightly smaller, 19" screens with steering, gas, and brake controls manufactured by ECCI. In addition, the MiniSim featured a reconfigurable dash screen mounted just below the central monitor. Both simulators gathered and stored a variety of driving performance measures sixty times per second. Nearly identical driving scenarios were authored for the two simulation platforms. In each case, a straight, rural, two-lane road with a shoulder was simulated. Several objects, such as buildings, traffic cones, trees, etc., were placed along the sides of the roadway to provide visual interest and to break up the monotony of the drive. Each driving segment began with the simulators stopped in the right lane.

The 2005 Toyota Highlander at TTI was instrumented with 4 cameras, a 3-axis accelerometer, high accuracy GPS, and sensors on the steering, gas, and brakes. Data collection and storage was handled by a data stream integrator (Dewetron 5000) which was configured to save data sixty times per second. The closed driving course was located on a former airfield. Similar to the simulated driving environments, long straight sections of runway and taxiway were driven.

*Secondary tasks***:** Identical secondary tasks were used in each of the three experiments. In these, A Sign Display Task, an Alert Response Task, and an Information Search Task were periodically activated during driving. The overall project was aimed at assessing distraction due to tasks enabled by Connected Vehicles technologies. The results shown in this paper focus solely on the comparison across platforms and, as such, the specific tasks should be considered as examples of visual-manual tasks typical of any touchscreen center stack display.

In the Sign Display Task, two types of roadway signs were displayed on a Xenarc 10.2-inch touch screen monitor: speed limit signs and work zone signs. The displayed sign alternated between each sign type once every 40 seconds with a 10-second standard deviation. For the simulator studies, the two speeds shown were 40 or 55 mph. For the test track study, speeds of 35 or 40 mph were displayed. Each newly displayed sign had a 50% probability of changing from the previous sign of that type. That is, if the prior speed limit sign displayed 40mph, the next speed limit sign had a 50% probability of changing to 55mph and a 50% probability of remaining the same. Drivers were instructed to follow the speed shown on the display screen.

For the Alert Response Task, drivers were given either an auditory or visual warning every 20-30 seconds where the probability of an auditory or visual task was equal and randomly determined. Once alerted, drivers directed their attention to a softly illuminated "decision light" that was mounted above the center screens in the driving simulators or on the hood in the instrumented vehicle (See Figure 1). If the decision light was red (80% probability), participants were instructed to press a response button that was mounted just to the right of the steering wheel, if it was green (20% probability), they did nothing. Alert tones were generated as needed using the simple harmonic combination of C3, C4, and C5, at a pulse rate of 3.3hz (200ms on, 100ms off). Volume was calibrated to play ~20db above ambient road noise. This resulted in 75-90db alerts across each of the research sites. This task was meant to emulate a crash warning system where the alert would direct the drivers attention back to the forward roadway and the decision light provides a response choice (go/ no go) similar to one a driver may make in a developing crash situation.

An information search task, designed to mimic potential mobility or sustainability functions in a Connected Vehicles environment, was displayed on a SoundGraph 4.3-inch touch screen. Sixteen audio recordings were made which posed simple yes/no questions that could be answered through information found by making menu selections and interpreting maps and graphs. Every 80 seconds, with a 15 second standard deviation, a new question was asked. Once the participants found the answers by navigating through the menus, they responded by pressing either the Yes or No button on the screen.
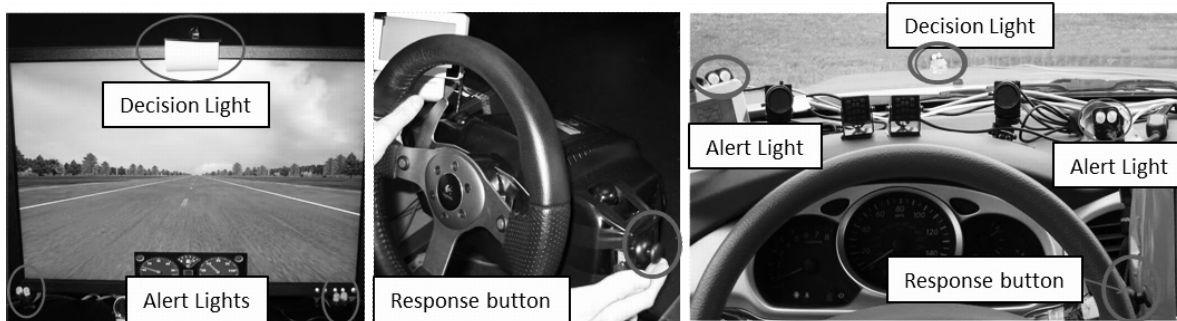
Figure 1. Alert task setup for simulator and instrumented vehicle studies

Software to control task timing and to collect participant responses was written in Python. The random onset of each task resulted in a random intermixing of task presentation to participants such that no participant experienced the exact same timing as any other. Responses to the alert task were collected using a Phidgets 8/8/8 Interface Kit and an analogue response button. Timing tests indicated approximately 1ms response accuracy to screen or button press detection.

Procedure: Each participant completed three driving segments (See Figure 2). Segment A began with a 90 second baseline drive, followed by 30 seconds of just the Alert Task, then 30 seconds of the Alert + Sign Tasks, and then 400 seconds of the Alert + Sign + Information Search Tasks. Segment B consisted of 400 seconds where all three tasks were eligible to execute. Segment C began with 90 seconds of baseline driving and then consisted of 400 seconds where all three tasks could execute. Data presented in this report was generated during the driving segments when all tasks could execute. At the end of each segment workload and situation awareness questions were administered, but are not reported here due to space limitations.
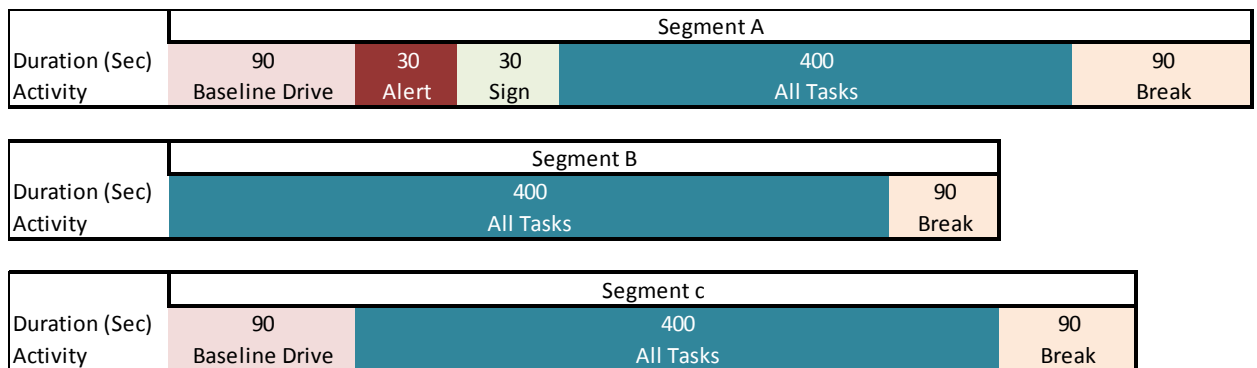

Figure 2. Schematic of driving segments

Measures and Results

As previously stated, the intent of this report is not to provide an exhaustive evaluation of all dependent variables across research platforms and secondary tasks, but rather, to highlight just a few key relationships and patterns. To this end, two common speed control measures and two common secondary task response measures are reported. These are: *Mean Speed*, defined as the average of speed minus posted speed; *Standard Deviation of Speed*, defined as the standard deviation of speed minus posted speed; *Reaction Time*, defined as time it took participants to press the response buttons when the decision light illuminated red; and *Missed Events*, defined as

the percentage of red decision light activations where participants did not respond; *Responses to False Alarms*, defined as the percent of green decision light activations which elicited a button press by participants.

Mean Speed: Across all research sites and platforms, the highest mean speed was recorded during the baseline driving condition and the lowest reported speed occurred during the information search task (See Figure 3 and Table 2). The relative significance of each of the secondary tasks differed slightly depending on the research platform and location, but in each case, mean speed was significantly higher in the baseline driving condition than during the information search task (Repeated measures ANOVA, all p's < .01).

*Standard Deviation of Speed***:** In a nearly identical manner to results on mean speed, the standard deviation of vehicle speed was consistently lowest during the baseline driving condition and consistently highest during the information search task. This pattern was also statistically significant across each of the experimental platforms (Repeated measures ANOVA, all p's < .01). Thus, across each research platform, drivers consistently maintained higher speed and less speed variability in the baseline driving condition and maintained slower, more variable speed, while completing the information search task.
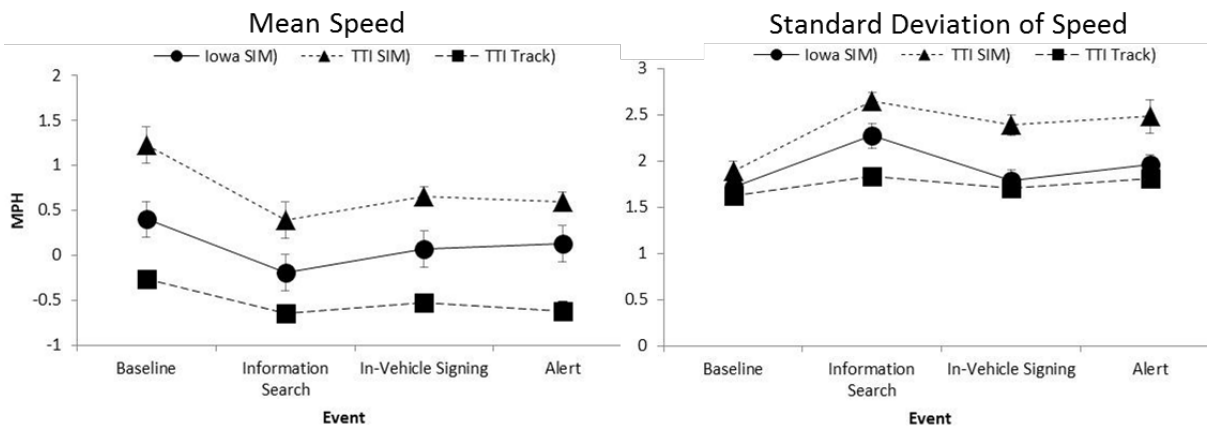

Figure 3. Mean speed deviations from posted speed limit

Table 2. Speed Measures: within-platform speed pairwise comparisons, p < .05

Measure		Baseline	Information Search	In-Vehicle Signing	Alert
Mean Speed	TTI-Sim	3	1	2	1,2
	TTI – Track	3	2	2	2
	Iowa - Sim	3	2	2	2
SD Speed	TTI-Sim	1	3	2	2,3
	TTI – Track	1	2	1	2
	Iowa - Sim	1,2	3	1	2

Reaction Time: The alerting tasks were embedded in the secondary tasks, making their presentation nearly identical across each of the research platforms. This led to a great deal of consistency in their resulting values, with some notable exceptions. Namely, response times to

red lights in the alerting task did not differ by modality (visual or auditory) in either of the simulated driving environments (Iowa: t(39)=1.13, p > .05; TTI: t(39)=.414, p>.05) . However, on the test track response times to visual alerts were significantly slower than to auditory alerts (t(39)=8.17, p<.001).

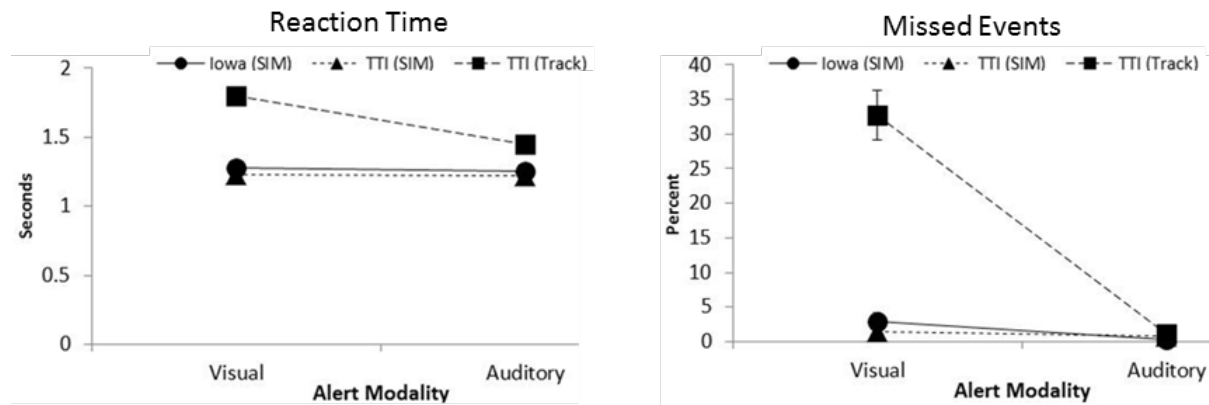**Figure 4. Reaction Time and Missed Events**

*Missed Events***:** In each of the three experimental settings, the average number of missed alerts was greater in the visual than auditory conditions. This pattern reached statistical significance on the Iowa driving simulator (2.9% visual rate, .9% auditory rate: Wilcoxon sign rank test, p < .05) and the TTI instrumented vehicle (32.7% visual, 1.1% auditory, p < .001) but not on the TTI driving simulator (1.5% visual, .7% auditory: p = .126).

Discussion

In this research we evaluated speed and response time measures that were recorded using two driving simulator platforms and an instrumented vehicle. Results from these investigations help to address the question of how performance on a consistent secondary task might be expected to vary across different driving research platforms. Data from these studies indicated that the driving simulators and the instrumented vehicle generally showed a high degree of concurrence, this was especially true with the speed measures across all three platforms and the reaction time and missed events in the driving simulators. The most notable difference between platforms occurred on the Alert Response Task in the instrumented vehicle. Specifically, in the instrumented vehicle, visual alerts led to the slowest response time and resulted in the greatest amount of missed events. These findings have implications for efforts to standardize testing and evaluation of secondary task performance across a variety of research platforms.

One of the most interesting outcomes of this research was how similarly participants controlled their speed in each of the driving platforms. As a reminder, for these analyses we calculated speed as the difference in the subject vehicle velocity from the posted speed limit. This manipulation allowed us to more directly compare the various speeds across the two driving simulators and the instrumented vehicle. Due to safety concerns and the length of available roadway, speed limits in the closed course driving varied between 35 mph and 40 mph while speeds in the driving simulators varied between 40 mph and 55 mph. Notwithstanding these differences, we found that in each of the conditions, participants responded with reliable speed modulation that was consistent in the simulators and on the test track. This speed similarity of control was evident in both the mean and standard deviation of vehicle speed with general

reductions in mean speed associated with secondary task processing as well as general increases in speed variability. People often have trouble controlling their speed in small, fixed-base simulators, so the fact that in both simulators they were able to manage their speed as well as they did on the test track speaks to the consistency in response across the testing platforms.

Similarly to speed control, we found strikingly parallel task performance in the Alert Response Task between the driving simulators. However, responses to visual alerts were significantly slower in the instrumented vehicle. While there are a number of potential accounts for these findings, observations made by the research assistant, and comments from many of the participants, indicate that participants may have had a difficult time noticing the blinking visual alerts in some daytime lighting conditions. For consistency, the visual alert presented on the track was identical to that used in both of the simulated environments. However, the ambient light was significantly greater, and more variable, on the test track than the driving simulators. This created situations where the visual alert was more or less visible, depending on the time of day, changes in weather, and angle of the vehicle with respect to the sun. This finding suggests that caution should be used when generalizing the results of visually demanding tasks from the simulator to the real world as lighting conditions in the real world may dramatically affect performance.

ACKNOWLEDGEMENTS

REFERENCES

Blaauw, G.J., 1982. Driving experience and task demands in simulator and instrumented car: a validation study. Human Factors 24 (4), 473–486.

Harms, L., 1996. Driving performance on a real road and in a driving simulator: Results of a validation study. In: Gale, A.G., Brown, I.D., Haslegrave, C.M., Taylor, S. (Eds.). Vision in Vehicles V. Elsevier/North-Holland, Amsterdam, pp. 19–26.

Jamson, H. (2011). Cross-Platform Validation Issues. In Fisher, D., Rizzo, M., Caird, J., & Lee, J (Eds.), *Driving Simulation for engineering, medicine, and psychology* (pp. 13-1: 13-18). Boca Raton, FL: Taylor & Francis Group.

Godley, S.T., Triggs, T.J., Fildes, B.N. (2002). Driving simulator validation for speed research. Accident Analysis and Prevention. 34. pp 589-600.

Molino, J., Opiela, K., Katz, B., Moyer, M.J., (2005). Validate First; Simulate Later: A New Approach Used at the FHWA Highway Driving Simulator. Proceedings of the Driving Simulator Conference North America.

Mullen, N., Charlton, J., Devlin, A., & Bédard, M. (2011). Simulator Validity: Behaviors Observed on the Simulator and on the Road. In Fisher, D., Rizzo, M., Caird, J., & Lee, J (Eds.), *Driving Simulation for engineering, medicine, and psychology* (pp. 13-1: 13-18). Boca Raton, FL: Taylor & Francis Group.