REPLICATING TEST TRACK PROTOCOLS IN A SIMULATOR: WHAT NEEDS TO BE MATCHED?

Timothy Brown[1], Dawn Marshall[1], Neil Lerner[2], John Gaspar[1]
[1]University of Iowa, [2]Westat
Email: timothy-l-brown@uiowa.edu

Summary: Many different experimental methods are used to evaluate driving performance as well as to evaluate the effectiveness of various vehicle safety systems but the results often do not match between different experimental approaches. This study aimed to determine the extent to which results can be matched between a driving simulator and a test track when carefully designed studies are used to replicate findings. This study collected simulator data on the National Advanced Driving Simulator (NADS) at the University of Iowa to replicate findings concerning Forward-Crash-Warning interface effectiveness at the Vehicle Research and Test Center (VRTC), East Liberty Ohio. The simulator used a virtual replica of the test track as well as a road course. Event choreography and scanning behavior were compared. Results indicate that results from the simulator were similar to those obtained on the test track. This indicates simulators can replicate findings for the test track and are a valuable tool. Careful experimental design is required to match the event choreography to insure an appropriate comparison. An exact match of the driving environment was not needed for this interface evaluation to obtain comparable results. The extent to which matching motion cues was not evaluated and may prove challenging in simulators without motion systems.

OBJECTIVES

One of the challenges of in comparing or replicating results from different research methodology is understanding the extent to which specific aspects of the scenario matter. Differences are often observed from similar experiments, leaving the research community to try to interpret the implications of disparate results. Consider two studies conducted as part of the same research program to evaluate crash warning interfaces in Forward Crash Warning (FCW) systems. One study conducted using a test-track methodology (Forkenbrock et al., 2011) and another conducted using a driving simulator (Lerner et al., 2011) produced results related to the effectiveness of warning interfaces that appeared divergent. Both efforts found benefits to having a warning for at least one system. The simulator study considered found that both the brake pulse and the auditory/HUD warnings tested led to similar decreases in driver response time compared to no warning. The test track study found that from a range of warnings including visual alert, an audio beeping alert, and a seat belt tensioner that only the seat belt tensioner resulted in a benefit. Additionally, the lack of benefit associated with auditory alerts reported in the test track study conflicted with prior research that showed benefits (e.g., Ho & Spence, 2009; Lee, McGehee, Brown, & Reyes, 2002). Differences in platform and protocol among these studies may explain the difference in sensitivity to warnings.

Prior efforts to examine the differences in performance between the simulator and the test track have produced mixed results. Hoffman, Lee, Brown, and McGehee (2002) attempted to replicate last minute braking responses from test track work completed as part of the Collision Avoidance

Metric Program using a 6 degree of freedom full cab simulator. They found that although general patterns were in agreement and some results were statistically the same, there systematic differences throughout the braking process including a lack of sensitivity to the normal vs hard braking instruction. They attributed these differences to limitations with the visual and vestibular cues. Knapper, Christoph, Hagenzieker, and Brookhuis (2015) used a quarter cab fixed-base simulator and compared results between a field test and simulator that partially replicated the driving environment to assess the impacts of driver distraction. They found that the simulator could provide relative but not absolute validity. Brown, Dow, Marshall, and Allen (2007) compared on-road performance to performance in the National Advanced Driving Simulator for novice drivers in terms of stopping and turning behavior. The simulator used attempted to replicate the on-road environment as closely as possible but did not completely match roadway geometry specifically near intersection. They found that performance most closely matched when roadway geometry matched, but that interaction with traffic on-road caused additional variability that could not be controlled. These results are consistent with other research that shows challenges in precisely matching (Ahlström, Bolling, Sörensen, Eriksson, & Andersson, 2012; De Winter, Van Leuween, & Happee, 2012; Matas, Nettelbeck, & Burns, 2016; Mayhew et al., 2011; Meuleners & Fraser, 2015; Molino et al., 2005; Risto & Martens, 2014; Shechtman, 2010; Shechtman, Classen, Awadzi, & Mann, 2009)

This study was designed to evaluate whether an experimental design on a test track could be replicated in a simulator through careful design. Additionally, this study examined using a parallel protocol whether a simulated road drive instead of a simulated test track, holding other elements of the protocol the same, could also replicate the findings using the same method in the simulator.

METHOD

Experimental Design

This study utilized a 2x3 factorial with between-subject variables: platform and alert modality. Platform was at two levels: simulated test track and simulated road course. Additionally, published data from the VRTC test track study were used for comparison purposes. Alert modality had the levels of no warning, audio, and seat belt tensioner. These alert modalities represented a subset of the data that was previously collected on the test track as part of prior Crash Warnings Interface Metrics research (Forkenbrock et al., 2011). The two levels of platform were a driving simulator scenario replicating the test track drive on NADS-1 and the road drive version on the NADS-1.

Dependent Measures

As the aim of this effort is to compare data to previously collected data from the test track, common dependent measures will be used. Data from the simulator were reduced to provide the same measures used on the test track. The dependent measures considered were speed, distance to the stopped vehicle, time-to-collision (TTC) between the subject vehicle (SV) and the stopped lead vehicle (SLV), time to visual commitment – time from end of the task instruction until the driver was looking at the task, time to end of visual commitment from warning – time from the warning until the driver's eyes were on the road, visual commitment duration – time from looking to the task until looking back to the road, time-to-collision at the end of visual commitment, collision, minimum time to collision, and brake reaction time.

Participants

Thirty-six participants completed all study procedures successfully. Fifty-six participants were enrolled in the study, and twenty were dropped for a variety of reasons most of which were related to the participant being too far away or not looking away when the event began.

Apparatus

The NADS-1 consists of a 24-foot dome in which an entire car is mounted (see **Figure 1**). All participants drove the same vehicle—a 1996 Malibu sedan. The motion system, on which the dome is mounted, provides 400 square meters of horizontal and longitudinal travel and ±330 degrees of rotation. Each of the three front projectors has a resolution of 1600 x 1200; the five rear projectors have a resolution of 1024 x 768. The edge blending between projectors is five-degrees horizontal. The headway display was mounted on the dashboard centered 4.5 inches to the right of the forward line of sight of the participant (see **Figure 2**). The display provided the headway to the lead vehicle in feet. The number recall task used was mounted at 90-degrees to the driver's line of site (see **Figure 3**).

Figure 1. NADS-1 Driving Simulator

Figure 2. Headway Display

Figure 3. Number Recall Display

Inceptive Structure

Compensation was based on that provided by Forkenbrock et al. (2011) , except that it was scaled to meet the needs of local participant recruitment and institutional review considerations concerning coercive compensation. Base compensation was set at $17.50, which represents 50% of the amount used in the VRTC study. Maximum incentive compensation was set at $27.50, which is 42% of the amount used in the VRTC study. This provided a maximum compensation of $45. The base pay was pro-rated. If participation lasted less than 1 hour, participants earned $10. If participation last for over 1 hour, participants earned the full base pay. The incentive was based on the participant's ability to correctly complete the number recall tasks and maintain a safe headway. The ratio of incentive for the headway maintenance and number recall was consistent with the VRTC study: the maximum incentive for headway maintenance was $8.48, the maximum incentive for the number recall task was $19.05, and both are 42% of the amount used by VRTC. For headway maintenance, participants received incentive pay based on how well they maintained a headway of 110 feet. For the number recall task, participants received incentive pay based on how many numbers they got correct. Full details are available in Lerner et al. (2015).

Experimental Procedures

Participants were told that the goal of the research was to examine headway maintenance, when in fact their response to a warning system was being evaluated. Participants were instructed to maintain a headway of 110 feet ± 15 feet. The training presentation included descriptions of the number recall task and headway maintenance task as well as a description of the incentive pay.

Participants were recruited from the NADS registry, which contains over 7,000 individuals. An email was sent to those between the ages of 35 and 55 who had provided an email address. Potential participants were also contacted by telephone. A telephone screening procedure was used to ensure participants met all inclusion requirements. Upon arrival at the NADS facility, participants were escorted to a participant room where informed consent was obtained. Driving history and demographic data were collected. Participants then watched a self-paced PowerPoint presentation describing the driving simulator, the incentive structure, and the task they would be expected to perform while driving. The distraction task was then practiced. The participants then completed a single experimental drive. The experimental drive was followed by an assessment for signs of simulator sickness to determine if they needed to be replaced.

Scenarios

Two driving scenarios were used for this study: test track and road course. Each scenario had a practice segment followed by four segments in which participants maintained headway, engaged in the number recall task which involved reading a series of five sequential numbers and then recalling them aloud, and received feedback. Throughout the four straight-away segments, participants followed a lead vehicle at a nominal headway of 110 feet with a nominal speed of 35 mph. During the fourth segment, a lead vehicle revealed event was triggered where the lead vehicle shifts lanes while the driver is looking to the task and a stopped lead vehicle is revealed. The simulated test-track replicated the geometry of the VRTC test track, and the simulated road course provided longer straightways connected by left curves. The choreography of the lead vehicle revealed event followed the timeline that was laid out in **Figure 4**.
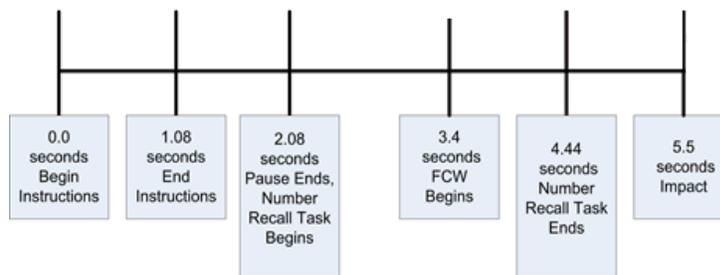
Figure 4. Event Choreography

RESULTS

In comparing the results from the test track with the results from the simulator, both the event choreography and the visual scanning were examined. To that end we examine how closely the data on the simulator matches the data published from the Forkenbrock et al. (2011) study. In looking at TTC to the stopped lead vehicle, the values for the simulator are typically similar to those at VRTC as can be observed in **Table 1**. The results were similar with the distribution of

the data overlapping. This general good correspondence is consistent with what would be expected from the distance data presented above.

Table 1. TTC with Stopped Vehicle Across Critical Points during Lead Vehicle Revealed Event

Description	SV-to-SLV TTC (seconds)								
	Task Instruction		Random Numbers Presented		FCW Alert		VC Concludes		
	VRTC	NADS	VRTC	NADS	VRTC	NADS	VRTC	NADS Eye Tracking	NADS Video Coding
Min	5.070	5.000	2.758	2.880	1.879	1.940	0.319	-1.450	-0.070
Max	5.765	6.210	3.743	3.790	2.325	2.350	1.872	1.330	1.860
Mean	5.412	5.520	3.117	3.330	2.064	2.120	1.030	0.440	1.160
Std Dev	0.165	0.250	0.186	0.190	0.094	0.090	0.466	0.570	0.340
Median	5.410	5.510	3.112	3.330	2.055	2.130	0.927	0.560	1.180
Nominal	5.5	5.5	3.4	3.4	2.1	2.1			

To assess the correspondence visual commitment times between the test track data and the simulator data for the virtual test track course time to visual commitment end from alert was examined. When comparing the end of visual commitment relative to the warning or time the warning would have gone off, the data from NADS shows faster response times and less variance than the VRTC data (see Figure 5). This may reflect the fact that drivers in the NADS study began looking away from the roadway towards the task from 400 to 600 ms sooner on average than drivers on the test track at the time the warning is triggered due to the faster start of the task.
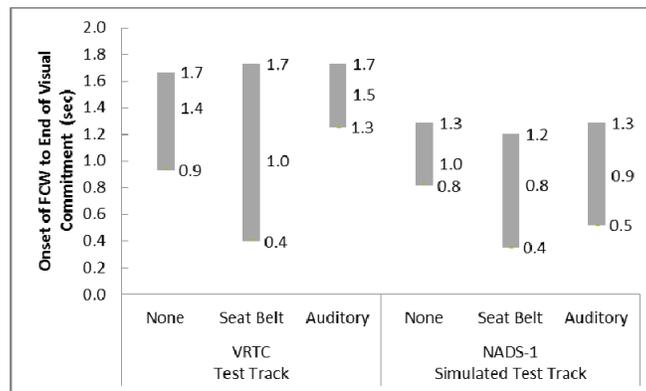
Figure 5. Range and Mean Time to Visual Commitment End from Alert

To assess the correspondence visual commitment times between the simulated test track and simulated road course, measures of visual engagement were examined. There was no significant difference between the road course and the test track for any the vision measures nor an interaction with warning type for time to visual commitment (p=0.2669, p=0.5659), visual commitment duration (p=0.8177, p=0.1941), and time to visual commitment end (p=0.7972, p=0.3419). The results for time to visual commitment end from alert are illustrated in Figure 6 which shows the distribution of the data for each warning type on each course.
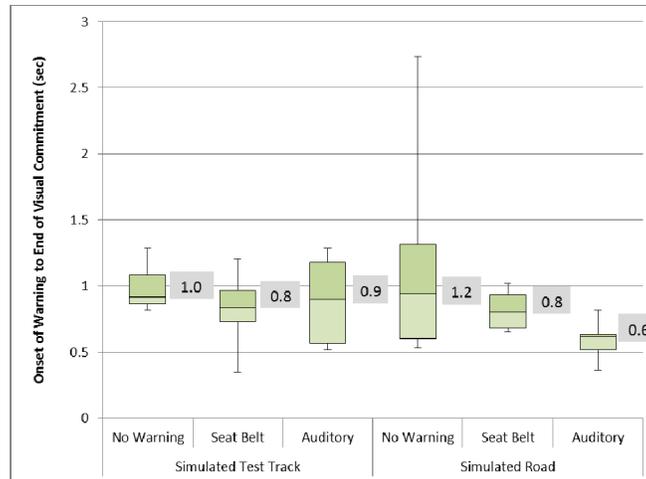
Figure 6.
Comparison of Time to End of Visual Commitment from Alert on Simulated Test Track and Road Courses

CONCLUSIONS

Overall, the performance of subjects on the simulated test track was quite comparable to the performance on the VRTC test track. From a choreography standpoint, the advantages of the simulator actually resulted in a more precise execution of the interaction between the subject and other vehicles. Subjects in the simulator were faster to engage in the numbers task and faster to stop it in response to the alerts, resulting in overall visual commitment duration remaining the same between the two environments. The results suggest that protocols developed on the test track should be able to be translated into the simulator. It also provides hope that the reverse would also be true with simulator scenarios translated to the test track.

When comparing the simulated test track and simulated road environments, results were generally comparable. This was particularly true for the no warning and seat belt tensioner conditions. Although there was no statistical difference, there appeared to be a trend across several measures for a more cautious response from subjects with the auditory alert on the road course compared to the test track environment. Overall, these results point to an opportunity to translate events between road courses and test track in some circumstances if care is taken to match the timing of the events and adjust the protocol to provide a parallel context for the driver relative to the driving situation. These results also point to the fact that, at least for this type of interface evaluation, that an exact match of the driving environment is not required.

It should be noted that this work used one of the worlds most advanced driving simulators which can provide 360-degree horizontal views and realistically replicate the motion cueing that drivers inherently rely on to control the vehicle. This research does not address the extent to which these results are generalizable to small simulators with less overall fidelity. Additional work is needed to evaluate the extent to which these more common simulators can replicate the distribution of results from test-track data collections.

REFERENCES

Ahlström, C., Bolling, A., Sörensen, G., Eriksson, O., & Andersson, A. (2012). Validating speed and road surface realism in VTI driving simulator III.

Brown, T., Dow, B., Marshall, D., & Allen, S. (2007). *Validation of Stopping and Turning Behavior for Novice Drivers in the National Advanced Driving Simulator.* Paper presented at the Driving Simulation Conference North America, Iowa City, IA.

De Winter, J., Van Leuween, P., & Happee, P. (2012). *Advantages and disadvantages of driving simulators: A discussion.* Paper presented at the Proceedings of Measuring Behavior.

Forkenbrock, G., Heitz, M., Hoover, R. L., O'Harra, B., Vasko, S., & Smith, L. (2011). *A test track protocol for assessing forward collision warning driver-vehicle interface effectiveness* (DOT HS 811 501). National Highway Traffic Safety Administration Washington, DC

Ho, C., & Spence, C. (2009). Using Peripersonal Warning Signals to Orient a Driver's Gaze. *Human Factors: The Journal of the Human Factors and Ergonomics Society*. doi:10.1177/0018720809341735

Hoffman, J., Lee, J., Brown, T., & McGehee, D. (2002). Comparison of Driver Braking Responses in a High-Fidelity Simulator and on a Test Track. *Transportation Research Record: Journal of the Transportation Research Board*(1803), 59-65.

Knapper, A. S., Christoph, M., Hagenzieker, M. P., & Brookhuis, K. A. (2015). Comparing a driving simulator to the real road regarding distracted driving speed. *European Journal of Transport and Infrastructure Research (EJTIR), 15 (2), 2015*.

Lee, J. D., McGehee, D. V., Brown, T. L., & Reyes, M. L. (2002). *Driver distraction, warning algorithm parameters, and driver response to imminent rear-end collisions in a high-fidelity driving simulator* (DOT-HS-809-448). National Highway Traffic Safety Administration Washington, DC http://ntl.bts.gov/lib/19...400/19457/PB2002107241.pdf

Lerner, N., Jenness, J., Robinson, E., Brown, T., Baldwin, C., & Llaneras, R. (2011). *Crash Warning Interface Metrics Final Report*. (DOT HS 811 470A). Washington, DC: National Highway Traffic Safety Administration

Lerner, N., Singer, J., Huey, R., Brown, T., Marshall, D., Chrysler, S., . . . Chiang, D. (2015). *Driver-Vehicle Interfaces for Advanced Crash Warning Systems: Research on Evaluation Methods and Warning Signals*. (DOT HS 812 208). Washington, DC: National Highway Traffic Safety Administration

Matas, N., Nettelbeck, T., & Burns, N. (2016). Assessment of driving simulator validity and acceptability for older adult drivers. *Advances in Transportation Studies*(39).

Mayhew, D. R., Simpson, H. M., Wood, K. M., Lonero, L., Clinton, K. M., & Johnson, A. G. (2011). On-road and simulated driving: Concurrent and discriminant validation. *Journal of safety Research, 42*(4), 267-275.

Meuleners, L., & Fraser, M. (2015). A validation study of driving errors using a driving simulator. *Transportation research part F: traffic psychology and behaviour, 29*, 14-21.

Molino, J., Katz, B., Duke, D., Opiela, K., Andersen, C., & Moyer, M. (2005). *Validate first; simulate later: a new approach used at the FHWA highway driving simulator.* Paper presented at the Driver Simulation Conference, North America, Orlando, FL.

Risto, M., & Martens, M. H. (2014). Driver headway choice: A comparison between driving simulator and real-road driving. *Transportation research part F: traffic psychology and behaviour, 25, Part A*, 1-9. doi:http://dx.doi.org/10.1016/j.trf.2014.05.001

Shechtman, O. (2010). Validation of driving simulators. *Advances in Transportation Studies*.

Shechtman, O., Classen, S., Awadzi, K., & Mann, W. (2009). Comparison of driving errors between on-the-road and simulated driving assessment: a validation study. *Traffic injury prevention, 10*(4), 379-385.