# REPEATABILITY OF THE NHTSA VISUAL-MANUAL GUIDELINES ASSESSMENT PROCEDURE

**Annie Rydström**
**Mikael Ljung Aust**
**Robert Broström**
**Trent Victor**
Volvo Car Corporation
Sweden

Paper Number 15-0240

## ABSTRACT

The Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices includes the Eye Glance Measurement Using Driving Simulator Testing (EGDS) task acceptance test. This paper describes the outcome of two EGDS tests, where both tests assessed the same set of 10 in-vehicle tasks, though each test employed a different group of 24 test participants, randomly sampled according to the proscribed EGDS procedure. Thus, in total 48 test participants drove on a simulated motorway in a lead vehicle following scenario while performing tasks such as changing temperature, setting destination in the navigation system and selecting and playing music using a prototype in-vehicle infotainment system located on the center panel. When comparing the test outcomes between the two groups, it was found that for 6 of the 10 tested tasks, pass/fail outcomes differed between the two groups on one or more of the proposed criteria. This high level of inconsistency in outcome between two identical tests using ten identical tasks raises questions regarding the repeatability of the proposed NHTSA EGDS procedure.

## INTRODUCTION

Human-Machine Interface (HMI) guidelines and standards to promote safe design are important tools during the design of in-car interfaces. Several design guidelines and standards have been published, including the European Statement of Principles [1] the JAMA guidelines [2] and the Alliance guidelines [3] (see [4] for an extensive overview of guidelines and standards to use to enhance safety and ease of use, and reduce distraction). Lately authorities have expressed increased concern about a potential intensification in hazardous driver behavior following the functionality growth in infotainment systems in cars and smartphones. As a part of the Driver Distraction Program the National Highway Traffic Safety Administration (NHTSA) in 2012 issued additional guidelines, the Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices. The guidelines provide interface recommendations and performance-based acceptance criteria. After some criticism from both the car industry and the academia on first and foremost the task acceptance testing described in the guidelines, the final version of the guidelines was published April 26 2013 [5] (and some additional clarifications has been published after that [6]). The guidelines are valid for all new cars sold in the USA 3 years after the date of publication.

The guidelines include the task acceptance testing procedure Eye Glance Measurement Using Driving Simulator Testing (EGDS). Volvo Car Corporation (VCC) has for many years used test methods similar to EGDS to evaluate in-car HMIs during their development phase, and has in previous years performed numerous EGDS tests. However, when analyzing the test data according to the EGDS acceptance criteria, it has been found that a task which passes the test at one test occasion does not necessarily pass at another occasion.

A key property of any performance assessment procedure is reliability, i.e. for outcomes to be consistently similar when a test is repeated. If results are not repeatable, the whole idea behind the test fails. The lack of test repeatability when applying the NHTSA EGDS criteria at VCC has therefore raised concerns. However, since the testing performed at VCC has not strictly followed the prescribed EGDS procedure, it is possible that the lack of repeatability may be due to differences in test setups rather than an inherent property of the criteria formulation.

The objective of this study was to assess whether the lack of repeatable outcomes previously found at VCC is methodological or structural in character. To this end, two identical EGDS task acceptance tests were performed at a professional test facility external to VCC. The same 10 in-vehicle tasks were used for both tests, and the NHTSA

proposed EGDS test protocol was strictly followed in terms of recruiting and test setup, in order to have exactly the same trial conditions for the two test groups.

## METHOD

The test was performed in the test facilities of Ergoneers GmbH in Manching, Germany.

### Participants

A total of 48 participants, i.e. two test samples, were recruited in the area around Manching in Germany. Each test sample conformed to the test participant recommendations in the NHTSA guidelines in terms of general criteria, test participant impartiality, mix of ages (six participants in each of the age groups 18-24, 25-39, 40-54, and 55+) and an equal balance of men and women [5].

### Equipment

The test was conducted in an Ergoneers Sim-Lab left-hand drive, fixed base and open cab driving simulator (Figure 1). The simulator is designed as a replica of an actual production vehicle and is adapted to the driving simulator recommendations in the NHTSA guidelines [5].

***Figure 1: The Ergoneers Sim-Lab driving simulator***

The driving simulation used in the study was a car following scenario on an undivided, four lane wide highway with a speed limit of 50 mph, designed in accordance with the driving simulator scenario recommendations [4]. The Ergoneers Dikablis head mounted eye tracking system was used to record the eye movements of the participants. A touch-screen tablet with a prototype infotainment HMI implemented on it was mounted at the center panel of the simulator. Tasks 1-9 were performed on this tablet. An additional, conventional Panasonic car stereo was installed in the simulator to be able to test the AAM radio manual tuning reference task [3].

### Tasks

The prototype infotainment system was designed to replicate a typical in-car center stack interface. Nine tasks were included for the in-vehicle interface, and an additional radio manual tuning reference task was added for the conventional car stereo (Table 1). I line with the guideline recommendations, all tasks started from a home screen view and ended when the task had been executed correctly [5].

*Table1: Tasks*

| Task no. | Task | Task description |
|---|---|---|
| 1 | Activate vehicle function | Access menu and activate a specific vehicle function (on/off option) |
| 2 | Activate USB | Access menu and activate USB |
| 3 | Set a new destination | Access navigation menu and activate a search field and tap in a destination of 10 characters and start guidance |
| 4 | Seta a favorite destination | Access navigation menu and a favorite menu, choose a specific destination and start guidance |
| 5 | Select and play music from USB | Access menu and activate USB and find a specified song in a list |
| 6 | Call contact from recent calls list | Access phone menu and recent calls list and choose a specified contact in the list to call |
| 7 | Dial a phone number manually | Access phone menu and keypad and tap in a 10 digit phone number and call |
| 8 | Change temperature | Access climate and increase temperature with 2 degrees |
| 9 | Set seat heat | Access seat heat and set it to a mid-level |
| 10 – AAM reference task | Manual radio tuning | Access FM1 and adjust to 106.7 (starting position FM3, presetting 97.4) |

**Procedure**

The test procedure, including test participant training, was arranged in alignment with the recommendations in the guidelines [5]. Upon arriving to the test facilities, the participants were given a brief description on the test procedure and they were asked to fill in a consent form and a demographic questionnaire. Then they were asked to sit down in the simulator and adjust the seat. They got to put on the eye tracking glasses and the eye tracking system was calibrated. The user interface and the tasks were demonstrated to the participants and they were able to practice each task as many times as they wanted to. The participants were then given instructions on the driving scenario and got to practice driving the simulator. When they felt comfortable driving the simulator they got to practice to perform the tasks as many times as needed while driving. After the training session the data trial was performed, i.e. the participants performed the tasks and data was collected.

**Acceptance criteria**

Eye movements were analyzed in accordance with the EGDS acceptance criteria [5, 6], which proscribes that no more than 3 out of 24 of the participants shall fail to meet each of the following criteria:

- **Acceptance criteria 1 (AC1):** For at least 21 of the 24 test participants, no more than 15 percent (rounded up to the next whole number) of each participant's total number of eye glances away from the forward road scene have durations of greater than 2.0 seconds while performing the testable task one time.
- **Acceptance criteria 2 (AC2):** For at least 21 of the 24 test participants, the mean duration of each participant's eye glances away from the forward road scene is less than or equal to 2.0 seconds while performing the testable task one time.
- **Acceptance criteria 3 (AC3**): For at least 21 of the 24 test participants, the sum of the durations of each individual participant's eye glances away from the forward road scene is less than or equal to 12.0 seconds while performing the testable task one time.

## RESULTS

When comparing the acceptance test outcomes between the two groups of test participants, results show that for the 10 tested tasks, outcomes were the same in four instances and different in six instances.

Tasks 1, 2, 3 and 9 had the same outcome in both test groups. Tasks 1, 2 and 9 passed on all criteria, while task 3 failed on all criteria (Table 2). The other tasks had different outcomes. Task 8 failed on AC1 and AC2 with group 1, but passed on AC1 and AC2 with group 2. Tasks 4 and 6 passed on AC1 but failed on AC2 with group 1, while the inverse was true for group 2, i.e. failed on AC1 but passed on AC2. Tasks 5 and 7 failed on AC1 and AC2 with

group 1 but passed on AC2 with group 2. Task 10 failed on AC1 with group 1 but passed on AC1 with group 2 and failed on AC3 with both groups.

***Table 2: The number of participants that failed to meet criteria, green = pass and red = fail***

| Task no. | Task | Group 1 | | | | Group 2 | | |
|---|---|---|---|---|---|---|---|---|
| | | AC1 | AC2 | AC3 | | AC1 | AC2 | AC3 |
| 1 | Activate vehicle function | 2 | 2 | 0 | | 3 | 3 | 0 |
| 2 | Activate USB | 3 | 3 | 0 | | 1 | 1 | 0 |
| 3 | Set a new destination | 8 | 6 | 9 | | 10 | 4 | 13 |
| 4 | Seta a favorite destination | 2 | 6 | 0 | | 4 | 2 | 0 |
| 5 | Select and play music from USB | 8 | 6 | 3 | | 6 | 3 | 1 |
| 6 | Call contact from recents list | 3 | 4 | 0 | | 4 | 3 | 0 |
| 7 | Dial a phone number manually | 11 | 9 | 3 | | 10 | 3 | 3 |
| 8 | Change temperature | 6 | 6 | 0 | | 2 | 2 | 0 |
| 9 | Set seat heat | 1 | 1 | 0 | | 0 | 0 | 0 |
| 10 | AAM ref. task: Manual radio tuning | 5 | 2 | 12 | | 3 | 1 | 16 |

## DISCUSSION AND CONCLUSIONS

The objective of this study was to assess whether the lack of repeatable outcomes previously found at VCC when applying the NHTSA acceptance criteria to EGDS test type methods was primarily methodological or structural in character. The results of the two identical EGDS acceptance tests reported above indicate that the lack of repeatability stems from how the criteria are designed, rather than how the test is performed. In this case, the two tests were set up following the proscribed protocol as strictly as possible, and the tested tasks were the same between the groups. Yet 6 out of 10 tasks came out different in terms of pass or fail on at least one of the proposed criteria.

From the above table, it mainly seems like acceptance criteria 1 and 2 have the most inconsistent outcomes, while acceptance criteria 3 (total eyes-off-road time) seems quite stable. Notable is the inconsistent outcome on task 10, which is a radio task that is supposed to work as a reference task, i.e. a task that has been prevalent throughout traffic through the ages and which drivers are thought to accept the risk of performing.

It is of course possible that some of the inconsistencies may still be due to methodological issues. In a perfect world, if two samples are recruited from the same population, and the same task is tested using the same test leaders, test method, equipment and logging apparatus, the outcome of a test should be consistent. In the real world on the other hand, differences may still occur due to different test leaders sometimes giving instructions in slightly different manners, and the environmental conditions and other factors may vary slightly. However, in this study, since the outmost care was taken to achieve an identical test setup for the two groups, it is unlikely that any unintended difference in the methodology would by itself lead to such a differentiated outcome. Moreover, since it would be a very large effort to reach a higher level of between test consistencies than the one in the current study, going for higher methodological precision does not seem to be an option. For an acceptance test to work in practice, it has to be robust enough to be applied at any certified test facility, without leading to inconsistent test outcomes.

Given that this study did have a high level of between test consistency, it thus seems like the problem with inconsistent outcomes have to do with how the acceptance criteria are formulated rather than with inconsistencies in

the EGDS testing. It is thus recommendable to investigate what it is in the criteria formulation that leads to these inconsistent outcomes, before applying the NHTSA EGDS method for assessment of production vehicles.

The results have some additional interesting properties. For one, the only tasks that passed all three criteria were the really simple ones that require only one or two button presses, such as activate a vehicle function, activate USB and set the seat heat. While that these simple tasks would pass was expected, it was not expected that a fourth task that presumably should belong to the same category of simple tasks, i.e. change the temperature in the vehicle, would fail (though it only failed in group 1, it passed in group 2). Looking at the detailed glace data from the two groups above, it can be seen that many drivers have long mean glance durations for these simple tasks. This suggests that drivers strive to complete these tasks in one go, rather than use multiple off-road glances.

This further highlights the need to understand what it is in the acceptance criteria formulation that leads to inconsistent outcomes. If future studies verify that mean glance durations indeed are longer for short and simple tasks, then this illustrates a deliberate driver strategy that has to be accommodated both in task development and testing. NHTSA has partially taken this into account by allowing a limited number of glances over 2.0 seconds for a task [6].

However, the results from this study shows that this topic still needs further investigation. For example, as the criteria are currently formulated, a person who uses a single glance of 3 seconds to complete a task will fail on both AC1 and AC2, but if that person uses a single glance of 3 seconds and then adds a check glance of 0.5 seconds (e.g. to verify that a setting has been updated), that person will pass both on AC1 and AC2. In the raw data from the two studies above, there are many examples of drivers who consistently fall into either of these two categories (i.e. they either use one long glance, or one long glance combined with a check glance). The behavior thus clearly is there empirically speaking, and therefore has to be considered in the acceptance criteria formulation.

In summary then, the difference in outcome between two identical tests using ten identical tasks raises questions regarding the repeatability of assessment outcome of the proposed NHTSA EGDS procedure and criteria. Since the need for an assessment method of this type is clear, and projected to lead to increased safety on the roads, further research into what exactly would need reformulation in criteria and method for the assessment procedure to fulfill its intended purpose is urgently needed.

## REFERENCES


[1] European Commission. 2008. “Commission Recommendation of 26 May 2008 on Safe and Efficient In-Vehicle Information and Communication Systems; Update of the European Statement of Principles on Human-Machine Interfaces.”

[2] JAMA. 2004. “Guideline for In-Vehicle Display Systems, Version 3.0”. Japan Automobile Manufacturers Association, Tokyo, Japan, August 2004.

[3] Alliance of Automobile Manufacturers. 2006. “Statement of Principles, Criteria and Verification Procedures on Driver-Interactions with Advanced In-Vehicle Information and Communication Systems”, Driver Focus-Telematics Working Group, Washington, DC., June 26, 2006 update.

[4] Green, P. 2009. “Driver Interface Safety and Usability Standards: An Overview”. In K. Young, J. D. Lee and M. A. Regan (Eds.) Driver Distraction: Theory, Effects, and Mitigation. Boca Raton, FL: CRC Press.

[5] Visual-manual NHTSA Driver Distraction Guidelines for in-Vehicle Electronic Devices (2013). Docket no. NHTSA-2010-0053, National Highway Traffic Safety Administration, April 2013.

[6] Visual-manual NHTSA Driver Distraction Guidelines for in-Vehicle Electronic Devices (2014). Docket no. NHTSA-2014-0088, National Highway Traffic Safety Administration, September 2014.