U.S. Department
of Transportation
National Highway
Traffic Safety
Administration

www.nhtsa.gov

DOT HS 811 726

March 2013

Driver Behavior During Visual-Manual Secondary Task Performance: Occlusion Method Versus Simulated Driving

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT HS 811 726		
4. Title and Subtitle Driver Behavior During Visual-Manual Secondary Task Performance: Occlusion Method Versus Simulated Driving		5. Report Date April 2012 6. Performing Organization Code NVS
7. Author(s) Thomas A. Ranney, G. H. Scott Baldwin, and Larry A. Smith, Transportation Research Center Inc.; John Martin, Ohio State University; Elizabeth N. Mazzae, National Highway Traffic Safety Administration		8. Performing Organization Report No.
9. Performing Organization Name and Address U.S. Department of Transportation National Highway Traffic Safety Administration 1200 New Jersey Avenue SE. Washington, DC 20590		10. Work Unit No. (TRAIS) 11. Contract or Grant No.
12. Sponsoring Agency Name and Address National Highway Traffic Safety Administration 1200 New Jersey Avenue SE. Washington, DC 20590		13. Type of Report and Period Covered 14. Sponsoring Agency Code

15. Supplementary Notes

The authors wish to thank Ed Parmer, Adam Andrella, Jodi Clark, Mike Rhoades, Matt Hostetler and Judy Weiser for contributing to this work.

16. Abstract

Experiment: An experiment was conducted to assess proposed acceptance criteria of NHTSA's Visual-Manual Driver Distraction Guidelines. Seventy two participants, ages 18-75 (3 age groups, balanced gender), performed secondary tasks (radio tuning, destination entry, 7-digit phone dialing, and 10-digit phone dialing) in two protocols: (1) driving a low-fidelity simulator; and (2) wearing occlusion goggles. Secondary tasks were performed using the integrated in-vehicle information and communication system of a 2010 Toyota Prius V. The Dynamic Following and Detection (DFD) driving simulator protocol combines car following with detection of visual targets presented in the roadway display during a single 3-minute drive. Performance metrics included: total eyes-off-road time (TEORT), mean glance duration, proportion of long glances, standard deviation of lane position (SDLP), car following delay, target detection accuracy (proportion correct), and target detection response time. The occlusion protocol required participants to perform 5 instances of each secondary task while wearing occlusion goggles that switched every 1.5 seconds between open and closed (occluded) states (as defined in ISO 16673). Total shutter open time (TSOT) was the performance metric.

Results: The full sample (N = 72) was used to determine "ground truth" test outcomes for each metric. Test results were compared with expectations that radio tuning is acceptable for driving and destination entry is not. Occlusion TSOT values were greater than the 9-second proposed NHTSA acceptance criterion, indicating all four secondary tasks did not meet the acceptance criterion. TEORT values were less than the 12-second proposed NHTSA acceptance criterion for 7-digit dialing and radio tuning, but not for destination entry and 10-digit dialing. Criterion values for mean glance duration and proportion of long glances were not sensitive to differences among secondary tasks. The observed TEORT/TSOT ratio was close to 1.0, considerably less than the predicted value (1.33).

DFD driving performance metrics were compared to fixed and benchmark acceptance criteria. Results for fixed criteria included: 1) car following delay – all tasks met acceptance criterion; 2) SDLP – destination entry did not meet acceptance criterion, other tasks met criterion; 3) target detection response time – all tasks met acceptance criterion; and 4) target detection accuracy – destination entry did not meet acceptance criterion, other tasks met criterion. Results relative to the destination entry benchmark included: 1) car following delay – no tasks met acceptance criterion; 2) SDLP – all tasks met acceptance criterion; 3) target detection response time –acceptable for 10-digit dialing, but unacceptable for other tasks; and 4) target detection accuracy – all tasks met acceptance criterion. Results were not entirely consistent for fixed vs. benchmark approaches.

Analyses of 3 subsets (N = 24) were performed to assess consistency of results over repeated testing. Results were consistent across groups for the glance-based metrics in both the DFD and occlusion protocols, however because criteria for some metrics did not discriminate among tasks as predicted, these tests were not strong. Results for several driving performance metrics were not consistent across groups.

Strong age effects were apparent for all metrics. Younger drivers (18-25) were most likely to provide outcomes meeting the acceptance criteria; older drivers (60-75) were most likely to provide outcomes not meeting the acceptance criteria. Strong and consistent age effects reflect heterogeneity of sample construction, which could reduce test sensitivity when using relatively small samples.

Overall, 3 metrics found destination entry unacceptable and radio tuning acceptable: (1) TEORT (driving simulator); (2) lane position variability (SDLP); and (3) target detection accuracy. The distribution of TSOT values suggests that increasing the proposed criterion could provide the expected outcomes. The proportion of long glances appears more promising than mean glance duration for ensuring tasks can be performed without long glances. Samples of 24 cannot be expected to provide consistent results over repeated testing for all metrics. Both occlusion and DFD protocols are too demanding for some older drivers. Other procedural issues that could influence test outcome were identified.

17. Key Words		18. Distribution Statement	
Driver performance, in-vehicle information systems (IVIS), visual-manual distraction		Document is available to the public from the National Technical Information Service www.ntis.gov	
19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	92	

Form DOT F 1700.7 (8-72) **Reproduction of completed page authorized**

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

Background

On February 24, 2012, NHTSA released a proposed set of Driver Distraction Guidelines covering original equipment in-vehicle device secondary tasks (where secondary tasks are defined as communications, entertainment, information gathering, and navigation tasks not required to drive) performed by the driver through visual-manual means (meaning the driver looking at a device, manipulating a device-related control with the driver's hand, and watching for visual feedback). These Guidelines will be used to identify secondary tasks that interfere with a driver's ability to safely control a vehicle (National Highway Traffic Safety Administration [NHTSA], 2012). The NHTSA Guidelines specify multiple test methods with acceptance criteria. While including a total of seven test protocols in its Notice of Proposed Federal Guidelines (NHTSA, 2012), NHTSA indicates that it prefers the following two test protocols and their associated acceptance criteria:

- Option EGDS: Eye Glance Testing Using a Driving Simulator, and
- Option OCC: Occlusion Testing.

Following an approach set forth by the Alliance of Automobile Manufacturers' (Alliance) and the growing body of research indicating that eye glance behavior is a prominent indicator of driver distraction, these preferred test protocols focus on eye glance behavior as a method for determining which visual or visual-manual secondary tasks are not acceptable for performance by a driver while operating a vehicle.

Objectives

The purpose of this research was to support NHTSA's driver distraction guideline development efforts. The specific objectives included:

1. Evaluate the proposed criteria using glance data obtained directly from a protocol in which drivers perform secondary tasks while driving a simulator.
2. Evaluate the occlusion protocol and the proposed criterion.
3. Determine the relation between these two metrics and in particular, whether this relation remains constant across a variety of secondary tasks.
4. Use specified driving performance metrics to determine the test outcomes of different secondary tasks with both fixed acceptance and benchmark criteria. A second part of this objective was to compare the test outcomes derived using driving performance metrics with those derived using the glance-based metrics.
5. Determine how the test outcomes differ when using different sample sizes and whether repeated testing using independent samples provides consistent results.

Experiment

An experiment was conducted using two protocols to assess the procedural details and proposed acceptance criteria of NHTSA's Driver Distraction Guidelines for visual manual secondary tasks performed using in-vehicle electronic devices. It is important to note that the procedural details of the work reported herein are not entirely consistent with the details specified in the recently-released draft Guidelines. This research was designed and implemented while the Guidelines specifications were evolving. One notable difference is that the age group definitions used for the present work differ slightly from those included in the draft Guidelines. Three age groups were used in the present work, while four are specified in the Guidelines.[1] One possible effect of this difference is that the Guidelines specifications

[1] Test sample composition specifications contained in the NHTSA Guidelines require samples of 24 participants, with 6 (3 females, 3 males) in each of the following age ranges: 18-24; 25-39; 40-54; 55 and older.

allow the use of slightly younger participants in the oldest group than were used in this study. Other differences will be noted as appropriate in this report.

The experiment used a within-participants' repeated-measures design, in which all participants completed all task conditions. Secondary tasks, including radio tuning, destination entry, 7-digit phone number dialing, and 10-digit phone number dialing, were performed using the integrated in-vehicle information and communication system of a 2010 Toyota Prius V. Seventy-two participants, ages 18-75 (3 groups: 18-25; 26-59; 60-75, each group balanced by gender), completed a single session lasting approximately 4 hours. Participants performed the secondary tasks in each of two protocols: 1) while driving a low-fidelity simulator in the DFD protocol; and 2) while wearing occlusion goggles that periodically masked the secondary task interface in accordance with the protocol defined in ISO 16673. The DFD driving simulator protocol combined car following with target detection, in which drivers responded to simple visual targets presented frequently in the simulated roadway display. Each combination of primary (driving) and secondary task was performed during a single 3-minute drive. Driving protocol metrics included: total eyes-off-road time (TEORT), standard deviation of lane position (SDLP), car following delay, target detection accuracy (proportion correct), and target detection response time. The occlusion protocol required participants to perform 5 instances of each secondary task while wearing occlusion goggles that switched every 1.5 seconds between open and closed (occluded) states. The total shutter open time (TSOT) was the performance metric.

Results

The proposed acceptance criteria of NHTSA's Driver Distraction Guidelines for visual manual secondary tasks were assessed relative to the expectations that radio tuning is acceptable and destination entry is not acceptable for performance by the driver while driving. Acceptance criteria providing results consistent with these expectations were considered successful. Results from the full sample (N = 72) were interpreted to represent "ground truth." Accordingly, results from small sample testing were assessed relative to the full sample results. NHTSA's proposed 9-second occlusion TSOT criterion was evaluated using a large sample with the proposed acceptance criterion (21 of 24: 87.5%). It was found that none of the four secondary tasks met the acceptance criterion. Analyses were conducted using three independent samples of 24 participants to assess the test-retest reliability, or consistency of test outcomes. Each sample was constructed to satisfy NHTSA's earlier sample composition criteria in terms of age and gender. Results were consistent across the three 24-participant samples; no tasks met the acceptance criteria in any of the three groups. Additional analyses were later performed 1) with the oldest age group removed, and 2) using the revised age groups contained in the NHTSA Guidelines and both showed radio tuning to meet the acceptance criteria.

The total eyes-off-road time (TEORT) obtained in the DFD driving simulator protocol was computed for the same participants performing the same secondary tasks. Mean values, computed over multiple task instances[2] for each participant, were compared with NHTSA's proposed 12-second criterion. With the full sample (N = 72), two of the four secondary tasks (radio tuning and 7-digit dialing) had outcomes consistent with meeting the (87.5%) acceptance criterion. The acceptance criteria outcome decisions were consistent across the three 24-participant groups. Age differences were strong and statistically significant. Older participants were significantly more likely to provide outcomes that did not meet acceptance criteria than younger participants.

The use of the same participants for occlusion and DFD testing allowed a direct comparison of the TSOT and TEORT distributions associated with the various secondary tasks. The proposed criterion values for TEORT and TSOT were based on the expectation that the relation between these two metrics would be 12/9 or 1.33. Across all secondary tasks, TSOT values were highly correlated with TEORT values (R square = 0.81). Two regression equations were used to explore the relation between the two metrics. The

[2] The NHTSA Guidelines specify the use of a single task instance for computing TEORT.

results revealed that the TSOT and TEORT values were close to being equivalent across all secondary tasks. Specifically, when predicting TEORT from TSOT, it was found that TEORT = 8.66 seconds when TSOT = 9 seconds; when predicting TSOT from TEORT, it was found that TSOT = 11.75 seconds when TEORT = 12 seconds. The resulting ratio values (TEORT/TSOT) were 1.04 and 0.98, respectively, which are considerably smaller than the predicted ratio (4/3 = 1.33). Within the individual secondary tasks, correlations between TSOT and TEORT were very low, indicating weak relations between the two metrics. Therefore, mean values were compared for individual tasks rather than regression parameters.

TEORT/TSOT ratios computed using mean values for the individual tasks were: destination entry = 1.12; 10-digit dialing = 0.90; 7-digit dialing = 0.87; and radio tuning = 0.94. Again, these values are considerably smaller than the predicted ratio (1.33).

Analyses were performed using the proportions of long glances (> 2 seconds) from the DFD data. When considered in the aggregate, the overall proportion of long glances was less than 5% for all tasks. In contrast, all tasks had at least some participants with more than 15% of glances away from forward roadway view being longer than 2 seconds. When the data were summarized for each participant, all secondary tasks had more than 87.5% of participants with less than 15% of glances longer than 2 seconds. Thus, all secondary tasks were found to be acceptable relative to the proposed criterion. Analyses of mean glance durations were also performed. None of the tasks had any participants with mean values longer than 2.0 seconds, which is consistent with the conclusion that all tasks were acceptable for this metric.

Two sets of analyses were performed using the four primary DFD driving performance metrics (car following delay, lane position variability (SDLP), detection task response time and detection task accuracy). Data from all 72 participants were used for these analyses. The first set of tests compared metric means relative to proposed fixed acceptance criterion values. Three of the tests were structured such that acceptance required demonstration that sample means were statistically not different from an established criterion value; one metric (target detection task accuracy) required performance better than the criterion level for acceptance. For car following delay, no task had a mean value greater than the proposed criterion (4.6 seconds); thus, all tasks had acceptable levels of car following delay. For lane position variability (SDLP), destination entry had a mean value statistically greater than the proposed criterion (1.0 foot); thus, destination entry did not meet the acceptance criterion for SDLP while all other tasks met this acceptance criterion. For detection task response time, all tasks had acceptable levels of detection task response time. For detection task accuracy, destination entry had mean accuracy less than the (0.8) criterion; thus, destination entry did not meet the acceptance criterion while all other tasks were acceptable. These analyses were repeated for each of the three groups of 24 participants. Some differences were observed in outcomes among groups of 24 participants for SDLP and detection task accuracy (DT proportion correct).

The second set of DFD driving performance metric analyses involved comparisons using destination entry as the benchmark representing an unacceptable level of distraction potential. These tests were constructed such that for a task to be acceptable, metric values would have to be significantly better (reflecting significantly less distraction potential) than those associated with the destination entry task. For SDLP, all three secondary tasks (radio tuning, 7-digit dialing, and 10-digit dialing) were statistically better than destination entry, thus all 3 had outcomes meeting the acceptance criterion. For car following delay, none of the tasks were statistically better than destination entry, thus none would meet the acceptance criterion. For detection task response time, the 10-digit dialing task was statistically better than destination entry and would thus meet the acceptance criterion. For detection task accuracy, all three secondary tasks were statistically better than destination entry, so all three tasks met the acceptance criterion.

Age differences in the full sample were statistically significant. Older and younger participants differed significantly in their test outcomes. For one task (radio tuning), both younger and middle-aged participants had test outcomes consistent with meeting the acceptance criteria, while older participants' results were consistent with not meeting the acceptance criteria.

Age effects were strong in the DFD driving performance metrics. One of four metrics exhibited significant Age x Secondary Task interaction, indicating that age effects were not consistent across tasks for this metric. A summary of the test outcomes for all metrics based on the large sample results is presented in the following table (Table 1).

Table 1. Summary of Test Outcomes for Metrics using the Full Sample (N = 72)

Metrics	**Destination Entry**	**Radio Tuning**	**7-Digit Phone Dialing**	**10-Digit Phone Dialing**
TSOT < 9 seconds	Does Not Meet Criteria	Does Not Meet Criteria	Does Not Meet Criteria	Does Not Meet Criteria
TEORT < 12 seconds	Does Not Meet Criteria	Meets Criteria	Meets Criteria	Does Not Meet Criteria
PLG < 15% glances longer than 2 seconds	Meets Criteria	Meets Criteria	Meets Criteria	Meets Criteria
MGD < 2 seconds	Meets Criteria	Meets Criteria	Meets Criteria	Meets Criteria
Delay < 4.6 seconds	Meets Criteria	Meets Criteria	Meets Criteria	Meets Criteria
SDLP < 1 foot	Does Not Meet Criteria	Meets Criteria	Meets Criteria	Meets Criteria
Detection Task Response Time < 1.0 second	Meets Criteria	Meets Criteria	Meets Criteria	Meets Criteria
Detection Task Accuracy > 0.8	Does Not Meet Criteria	Meets Criteria	Meets Criteria	Meets Criteria
Delay (s) < Destination. Entry	NA	Does Not Meet Criteria	Does Not Meet Criteria	Does Not Meet Criteria
SDLP < Destination. Entry	NA	Meets Criteria	Meets Criteria	Meets Criteria
Detection Task Response Time (s) < Destination Entry	NA	Does Not Meet Criteria	Does Not Meet Criteria	Does Not Meet Criteria
Detection Task Accuracy > Destination Entry	NA	Meets Criteria	Meets Criteria	Meets Criteria

Test outcomes for the various metrics can be assessed by comparing them to the consensus position that destination entry is generally not acceptable for performance by the driver while driving, while radio tuning is acceptable for performance by the driver while driving. Three metrics found destination entry to be unacceptable and radio tuning acceptable, including TEORT obtained in the driving simulator, lane position variability (SDLP), and detection task accuracy. Results from TSOT obtained in the occlusion protocol revealed an acceptance criteria pattern consistent with this pattern; however the specific criterion was inappropriate for supporting these acceptance criteria outcomes.

Conclusions

1. The proposed 12-second criterion for total eyes-off-road time (TEORT) achieved results consistent with the expectation concerning radio tuning and destination entry.
2. The proposed 9-second total shutter open time (TSOT) criterion for occlusion trials did not provide test results consistent with the expectation that radio tuning would meet the acceptance criteria and destination entry would not meet the acceptance criteria.
3. The predicted 4/3 relation between TEORT and TSOT was not found. The observed relation is closer to 1/1; however among individual tasks, the correlations between the two metrics were weak, suggesting no systematic relation between the two metrics. The heterogeneity associated with the wide age range of participants is a possible reason for this result.
4. The proposed criterion that defines the acceptable proportion of long glances can better ensure that tasks are performed without long glances than the criterion based on mean glance duration. The overall mean glance duration was 0.97 seconds and none of the participant means in any condition was greater than 2.0 seconds. This, together with the finding that most participants had some proportion of glances longer than 2.0 seconds, supports the conclusion that a criterion directly related to long glance proportion will be more effective in limiting long glances than one based on the mean glance duration.
5. Strong and consistent differences were observed among age groups. Generally, younger participants (18-25) were more likely to provide outcomes meeting the acceptance criteria, while older participants (60-75) were more likely to provide outcomes not meeting the acceptance criteria.
6. The effects of several test procedural details, including the method of defining task boundaries, the selection of outliers, and the use of medians versus means as a measure of central tendency appear to be strong enough to influence test outcomes. Additional effort is warranted to assess the implications of these procedural details and recommend specifications.

1.0 Introduction

1.1 Background

On February 24, 2012, NHTSA released a proposed set of Driver Distraction Guidelines covering original equipment in-vehicle device secondary tasks (where secondary tasks are defined as communications, entertainment, information gathering, and navigation tasks not required to drive) performed by the driver through visual-manual means (meaning the driver looking at a device, manipulating a device-related control with the driver's hand, and watching for visual feedback). These Guidelines will be used to identify secondary tasks that interfere with a driver's ability to safely control a vehicle (National Highway Traffic Safety Administration [NHTSA], 2012). The NHTSA Guidelines specify multiple test methods with acceptance criteria. While including a total of seven test protocols in its Notice of Proposed Federal Guidelines (NHTSA, 2012), NHTSA indicates that it prefers the following two test protocols and their associated acceptance criteria:

- Option EGDS: Eye Glance Testing Using a Driving Simulator, and
- Option OCC: Occlusion Testing.

Following an approach set forth by the Alliance of Automobile Manufacturers' (Alliance) and the growing body of research indicating that eye glance behavior is a prominent indicator of driver distraction, these preferred test protocols focus on eye glance behavior as a method for determining which visual or visual-manual secondary tasks are not acceptable for performance by a driver while operating a vehicle.

The Alliance's 2006 "Statement of Principles, Criteria and Verification Procedures on Driver Interactions with Advanced In-Vehicle Information and Communication Systems" (or "Alliance Guidelines") also contains a test option (Alternative 2.1B) that is based on driving performance metrics. NHTSA's Guidelines similarly include a candidate test protocol based on driving performance; however, this protocol was not amongst those preferred for ultimate use in NHTSA's finalized Guidelines for visual-manual secondary tasks.

1.2 Objectives

The purpose of this research was to support NHTSA's driver distraction guideline development efforts. The specific objectives centered around evaluating several aspects of the various test protocols and task acceptance criteria being considered.

One objective of the current research was to evaluate the proposed criteria using glance data obtained directly from a protocol in which drivers perform secondary tasks while driving a simulator. One option in the NHTSA Guidelines involves the use of Alliance metrics with slightly different criteria. For example, using a test protocol that involves simulated driving and concurrent secondary task performance, NHTSA proposed the use of a criterion based on the proportion of longer glances exhibited during a task to ensure that most glances away from the forward roadway view are less than 2 seconds in duration. Specifically, it was proposed that no more than 15% of glances away from the forward view be longer than 2 seconds while the secondary task is being performed. To limit the incidence of long-duration glances, NHTSA also proposed to retain the Alliance criterion that requires mean duration of all glances away from the forward roadway view to be less than 2 seconds. NHTSA also proposed a criterion of no more than 12 seconds of total eyes-off-road time (TEORT). This criterion is based on recently collected eye glance data associated with the Alliance radio tuning benchmark, reflecting the Alliance assertion that concurrent performance of radio tuning represents the limit of acceptable distraction in everyday driving.

A second objective of the current research was to evaluate the occlusion protocol and the proposed criterion. Following the Alliance protocol, NHTSA also proposed allowing occlusion to be used as a surrogate means of obtaining information comparable to glance metrics obtained directly from a driving protocol. ISO Standard 16673 (ISO, 2007) presents a standardized protocol for occlusion, which simulates the switching of attention between the primary (driving) and secondary task by periodic masking of the secondary task interface. Intervals during which the display interface is masked are intended to simulate the intervals during which a driver must look at the forward roadway view to ensure safe vehicle operation. The primary benefits of occlusion are that it does not require extensive resources and can be applied to evaluate a functioning prototype early in the design cycle. NHTSA has proposed a 9-second total shutter open time (TSOT) criterion for use with the occlusion protocol.

The provision of allowing different methods to be used to assess the visual demand associated with secondary tasks requires establishing the degree of correspondence between the respective metrics. In particular, it is essential to determine whether the total eyes-off-road time (TEORT) required to perform secondary tasks in a driving protocol corresponds directly to the total shutter open time (TSOT) required to perform the same tasks in the occlusion protocol. Although published empirical support is not strong, the assumed relation between TEORT and TSOT is 4/3. This assertion is the basis for the proposed NHTSA criteria of 12 and 9 seconds, respectively. In an attempt to provide stronger empirical support, the third objective of the current research was to determine the relation between these two metrics and in particular, whether this relation remains constant across a variety of secondary tasks.

Direct measurement of glance behavior during driving can be obtained using the "Dynamic Following and Detection" (DFD) driving simulator protocol, which combines car following and visual target detection. In this protocol, test participants perform secondary tasks while driving a fixed-base driving simulator. In addition to the glance metrics, the DFD protocol provides driving performance measures of lateral and longitudinal vehicle control and visual target detection, which have demonstrated sensitivity for detecting the distraction potential associated with both visual-manual and voice-based tasks, independent of the glance-based measures. NHTSA is considering two approaches using driving performance metrics obtained in the DFD protocol, one based on fixed acceptance criteria and one based on performance relative to a benchmark task. The proposed fixed acceptance criteria were derived from previous DFD research results and reflect the values associated with the radio tuning task. The proposed benchmark task is the destination entry task, reflecting the emerging consensus that performing this task is not acceptable while driving. For the benchmark option, acceptable levels of distraction would require tasks to be demonstrably better than the levels of distraction-related driving performance degradation associated with destination entry. Therefore, the fourth objective of this research was to use specified driving performance metrics to determine the test outcomes of different secondary tasks with both fixed acceptance and benchmark criteria. A second part of this objective was to compare the test outcomes derived using driving performance metrics with those derived using the glance-based metrics.

Test outcomes are determined in part by the composition of test samples used to assess the secondary tasks. Larger, more homogeneous samples tend to provide more consistent test results than smaller samples with significant heterogeneity. Thus, while it is important to make sure the test results generalize to the population segment of likely users, it is also important to select a sample size that allows for efficient testing of secondary tasks. Accordingly, the fifth objective of the research was to determine how the test outcomes differ when using different sample sizes and whether repeated testing using independent samples provides consistent results.

2.0 Experiment

2.1 Approach

A single experiment was conducted using two protocols. One protocol involved visual occlusion as defined by ISO 16673; the second protocol was the Dynamic Following and Detection (DFD) protocol developed at NHTSA's Vehicle Research and Test Center (VRTC) (Ranney, et al., 2011), where simulator drives consisted of close car following on a straight road with minimal other traffic present. Data was collected in August and September of 2011. Both protocols were completed using the integrated in-vehicle information system (IVIS) of a 2010 Toyota Prius V. Metrics obtained from the respective protocols were used together with proposed acceptance criteria to assess a range of secondary tasks performed with a single integrated in-vehicle interface.

2.2 Experimental Design

The experiment used a within-participants' repeated-measures design, in which all participants completed all task conditions. The main design factors were the test protocol (occlusion, DFD) and the visual-manual secondary task (4 levels), which included:

- Destination entry by address,
- Radio tuning,
- 10-digit phone number dialing, and
- 7-digit phone number dialing.

The four visual-manual secondary tasks were performed using the original equipment radio, navigation system, and phoning interface of two 2010 Toyota Prius V test vehicles. For each secondary task, each participant completed one simulator drive of approximately 3.5 minutes (3-minute data collection interval) and 5 occlusion trials, plus training and practice. Simulator drives required participants to follow, at a specified distance, a lead vehicle with varying speed on a straight road with minimal other traffic present.

2.3 Participants and Sample Size

Participants were 72 members of the general public meeting the following general criteria:

- In good general health,
- Active driver with a valid driver's license,
- Drive a minimum of 7,000 miles per year,
- Have experience using a wireless phone while driving,
- Be comfortable communicating via text messages, and
- Not have a substantial amount of experience with systems/devices being tested.

Test participant ages were equally distributed across the following three age ranges:

- 18 through 25 years old, inclusive, and
- 26 through 59 years old, inclusive, and
- 60 through 75 years old, inclusive.

Test participant gender was balanced within each of the three age ranges. To facilitate repeated testing, the 72 participants were divided into three independent samples of 24 participants, each having age distributions consistent with the overall sample construction (eight participants from each specified age group, with balanced gender in each group). Since this research was designed and implemented while the

NHTSA Guidelines specifications were evolving, certain procedural details, such as these age ranges, are not entirely consistent with the details specified in the recently-released draft NHTSA Guidelines, in which four age groups are specified.

2.4 Recruitment

Recruiting participants was done using ads placed on the website craigslist.org, as well as in local newspapers, including the Columbus Dispatch and smaller local papers, including the papers in Marysville and Bellefontaine. The ad is shown in Appendix B. Respondents were offered several methods of responding, including via phone, email, and internet application. People who responded via the internet were directed to a part of the Transportation Research Center Inc. website, which contained a slightly longer description of the study and a link to a protected website. At this website, respondents were asked to provide basic information, including age and gender to determine eligibility, residence zip code to determine mileage reimbursement, and email contact information. Both the expanded description of the study and the specific information obtained from these respondents are taken directly from the participant screening questions (see Appendix B), which were used during call back for respondents meeting the selection criteria. During this brief call, respondents were asked a series of questions to ensure that they were licensed drivers with no vision problems or health problems, active text message users, and regular users of a cell phone while driving.

As part of the recruitment process, basic information was collected to track the success rate of obtaining participants for research, especially since the research involved such a broad age range of candidates. The collected recruitment statistics can be found in Appendix J, showing the quantities of candidates screened and scheduled by age range and gender.

2.5 Participant Compensation

Participants were paid an hourly rate of $42.00 for a single session lasting approximately 4 hours. Based on previous experience, some difficulty obtaining a sufficient number of suitable participants was expected, particularly since the target population is more likely to live closer to the population center in Columbus. Because prospective participants were recruited from areas as far as 40 miles away from the test site, mileage-based pay was provided in addition to the hourly base rate to encourage study participation.

2.6 Task Performance Feedback

In the DFD protocol, participants were given feedback regarding their performance on the primary driving task (car following and target detection) and secondary tasks. Feedback was used to motivate performance, but no incentive pay was provided.

Task performance priorities were outlined to all participants. Driving was described as the highest priority for all participants. Table 65 of Appendix E shows the task performance incentive criteria that was presented to each participant, and used for determining task performance feedback.

3.0 Procedure

The experimental procedure consisted of several components, including: (1) introduction, general instructions and informed consent; (2) DFD task trials; (3) occlusion task trials; and (4) participant debriefing. Each component is discussed in detail.

3.1 Introduction, General Instructions and Informed Consent

Participants selected following screening were scheduled individually for a single session of approximately 4 hours. Upon arrival at the data collection site, the participant was asked to read the Participant Information Summary, which included the informed consent statement (Appendix C). After all questions were answered, the participant signed the document, thereby consenting to participate in the study. If an individual declined to participate, he or she would be given the mileage reimbursement plus a nominal payment commensurate with time spent at the facility, and permitted to leave. However, no participants declined to participate.

3.2 General Protocol and Task Presentation Order

After consenting to the terms of participation, the participant was escorted to the experimental vehicle. Half of the participants were given DFD trials first; the other half received occlusion trials first. Task presentation order of the secondary tasks was balanced across age groups. Training, practice, and testing were completed for each block in the order assigned to a particular participant.

3.3 DFD Task Trials

When seated in the vehicle, the participant was given an overview of the controls and displays, and shown how to adjust the seat. This was followed by familiarization with the driving simulator (Appendix E) and a description of task performance feedback (Table 65). Next, the eye tracker was calibrated to the test participant (Appendix F). During this procedure, the participant was asked to affix latex stickers to his or her face in specific locations. The experimenter instructed the participant concerning head position and point of gaze. Following the eye tracker calibration, the participant was given instructions for the specific driving tasks that comprise the DFD protocol, which include car following and visual target detection (Appendix E). This was followed by practice drives allowing the participant to become familiar with the feel of simulator driving and the driving task components (car following plus visual target detection).

The participant then began the 5 DFD trials, which included one baseline trial and 4 secondary task trials. Before each trial, the experimenter described the secondary task (Appendix D) and the participant was given practice in performing the task alone, followed by a practice drive involving concurrent secondary task performance, and then the main trial. Each main trial lasted approximately 3.5 minutes and included a 3-minute data collection interval, during which the participant performed the specified secondary task continuously. The beginning of each task instance was preceded by an audio file that presented specific task instructions followed by the instruction to begin. Task instructions were also visible on the task screen. Upon hearing the instruction to begin, the participant began the secondary task, when driving conditions permitted. When done with each task instance, the participant said “done” aloud and pressed the task screen to initiate presentation of instructions that directed the participant to move the system from the task end state to the task begin state. Once the system was in the task begin state, the participant touched the task screen to initiate the audio file that contained new task instructions and the instruction to begin. This sequence continued until the data collection interval ended. Secondary task conditions changed after each main trial.

The participant was offered periodic breaks and encouraged to ask questions at any time. During task trials, the experimenters were positioned at a control station behind the vehicle. Communication with the participant was accomplished via two-way radio. Data collection intervals were initiated automatically with an audio file that presented specific task instructions followed by the instruction to begin.

After the DFD test trials, the participant completed a simulator sickness questionnaire (Appendix G) to determine if additional rest was required before performing the occlusion test trials or driving home (depending on the order of protocol presentation).

3.4 Occlusion Task Trials

Once the participant was seated in the vehicle, the experimenter described and demonstrated the occlusion procedure (Appendix H) using a simple visual inspection (Circles) task, which required the participant to look for a larger (target) circle in a pattern of smaller circles. When comfortable with the occlusion procedure, the participant completed 4 secondary task trial sets, one for each secondary task. Before each trial set, the experimenter described the secondary task (Appendix H) and the participant was given practice in performing the task alone and with occlusion. Once practice was complete, the participant performed the main trial component of the task.

At the beginning of each trial, the experimenter initiated presentation of an audio file that contained specific task instructions. Task instructions were also visible on the task screen. When ready, the participant was instructed to press a virtual go button on the task screen. Pressing the go button would cause the occlusion goggles to switch to the occluded state for several seconds, ending with the instruction to begin, at which time the alternating 1.5-second open and 1.5-second closed sequence began. The participant began work on the secondary task during the first open interval. Upon task completion, the participant was instructed to say “done” aloud and touch the task screen. Then, the participant was instructed to move the system from the task end state to the task begin state to be ready for the next task. Once the participant was ready, the experimenter initiated another audio file containing the next set of specific instructions to start the process again. As specified in the ISO occlusion protocol, the participant completed 5 main trials of each secondary task.

An experimenter was positioned outside the vehicle at a control station behind the driver during data collection. Communication with the participant was accomplished directly through an open window.

3.5 Participant Debriefing

At the completion of data collection of all DFD and occlusion task trials, the participant exited the vehicle and proceeded to the designated debriefing room. The pay amount was explained and given to the participant (Appendix I). The experimenter answered questions posed by the participant and returned the participant to his or her personal vehicle.

4.0 Results

The experiment provided three data sets, including: (1) Occlusion; (2) DFD eye glance data; and (3) DFD driving performance data. Specific details of the way each set was prepared for analysis are presented in the respective sections below. In general, the proposed acceptance criteria of the NHTSA Guidelines were assessed relative to the expectations that radio tuning is acceptable and destination entry is not acceptable for performance by the driver while driving. Acceptance criteria providing results consistent with these expectations were considered successful. Results from the full sample (N = 72) were interpreted to represent a "ground truth." Accordingly, results from small sample testing were assessed relative to the full sample results.

4.1 Evaluation of NHTSA Proposed Criteria Based on Occlusion and Glance Durations

The first two study objectives focused on evaluating NHTSA's proposed acceptance criteria using data obtained from occlusion trials and eye glance data obtained from the DFD driving trials. Occlusion results are presented first, followed by the total eyes-off-road time (TEORT) results and a comparison of the relation between TSOT and TEORT, which addresses the third study objective.

4.1.1 Occlusion

TSOT refers to the total shutter open time associated with a single occlusion trial. Following ISO 16673, each participant completed 5 trials for each secondary task while wearing occlusion goggles that switched every 1.5 seconds between open and closed (occluded) states. ISO 16673 recommends removal of outliers, which are defined as values that appear to be outside the overall pattern of a distribution. To identify outliers, the entire distribution of TSOT values, computed using the touch screen press as the end of the trial, was separated by secondary task. Visual inspection of the distributions led to the following definition of outliers, as shown in Table 2.

Table 2. Occlusion Outlier Criteria by Secondary Task

Secondary Task	Outlier Criterion
Destination entry	TSOT > 45 s
Dialing 10-digit	TSOT > 20 s
Dialing 7-digit	TSOT > 15 s
Radio tuning	TSOT > 20 s

There were 1440 total trials, of which 1425 (99%) provided usable data. Of these, 18 (1.3%) were identified as outliers and removed. Table 3 provides summary statistics for the respective TSOT distributions by secondary task. The last column presents the proportion of the distribution with values less than or equal to the 9-second criterion.

Table 3. TSOT Summary Statistics by Secondary Task (Touch Screen End Time)

Secondary Task	N	Mean TSOT (s)	SD	P ≤ 9 s
Destination entry	345	21.96	4.94	0.00
Dialing 10-digit	347	10.34	2.62	0.37
Dialing 7-digit	355	7.87	1.99	0.78
Radio tuning	358	7.81	2.32	0.81

Proposed NHTSA Guidelines (D.6.a) state that, for at least 21 of 24 participants, the task must be successfully completed during 6 or fewer unoccluded intervals (i.e., a maximum of 9 seconds of shutter open time). Acceptance criteria decisions relative to the 9-second criterion are based on group means rather than the entire distribution of TSOT values; however, the characteristics of the overall distribution (following removal of outliers) provide an indication of overall compliance with the 9-second criterion. From this perspective, because the overall proportions of trials with TSOT ≤ 9 seconds were all less than 0.875 (which corresponds to 21/24 participants), it is expected that none of the secondary tasks would provide test outcomes consistent with satisfactory test compliance.

Following ISO 16673, mean TSOT values were computed for each combination of participant and secondary task. Each mean summarized performance over 5 trials. The mean TSOT values were compared against NHTSA's proposed 9-second TSOT criterion. The results are presented in Table 4 for the entire sample (N = 72).

Table 4. Test Outcome Frequency for Mean TSOT Values Relative to NHTSA 9-Second Criterion

	Meets Criteria		Does Not Meet Criteria		Total	
Secondary Task	**N**	**p**	**N**	**p**	**N**	**p**
Destination entry	1	0.01	71	0.99	72	1.00
Dialing 10-digit	21	0.29	51	0.71	72	1.00
Dialing 7-digit	51	0.71	21	0.29	72	1.00
Radio tuning	56	0.78	16	0.22	72	1.00

According to the proposed acceptance criterion, a test is acceptable for performance by the driver while driving if at least 21 of 24 (87.5%) participants performed the task with no more than 9 seconds TSOT. The proportions presented in Table 4 indicate that none of the 4 secondary tasks attained the 87.5% acceptance criterion, supporting the conclusion that none of the secondary tasks would be acceptable for performance by the driver while driving, given the proposed TSOT acceptance criteria.

The sample was constructed to include three independent groups of 24 participants, each having age distributions consistent with the overall sample construction, which included equal numbers of participants from three specified age groups. The fifth analysis objective was to determine the consistency of the test results (i.e., test-retest reliability) across the three smaller samples. Table 5 presents the frequencies and proportions of trials meeting the acceptance criterion for the three groups of 24 participants.

Table 5. TSOT Acceptance Criterion Frequencies and Proportions by 24-Participant Group

	Group 1		Group 2		Group 3	
Secondary Task	**N**	**p**	**N**	**p**	**N**	**P**
Destination entry	0	0.00	0	0.00	0	0.00
Dialing 10-digit	8	0.33	7	0.29	6	0.25
Dialing 7-digit	17	0.71	15	0.63	19	0.79
Radio tuning	18	0.75	18	0.75	20	0.83

Although the frequencies differed slightly across the three groups, the outcomes relative to the (21 of 24, p = 0.875) acceptance criterion were consistent; none of the tasks had proportions meeting the acceptance criterion. As a result, the comparison among groups provided a relatively weak test of the test-retest reliability. A statistical comparison of acceptance criteria proportions was conducted to provide a more sensitive test of the stability of test outcome results. Two way chi-square tests were conducted for each secondary task. The results are presented in Table 6.

Table 6. Statistical Test Results of TSOT Group Differences

Secondary Task	Group Differences χ^2	df	p
Destination entry	0	2	1.00
Dialing 10-digit	0.40	2	0.82
Dialing 7-digit	1.61	2	0.44
Radio tuning	0.64	2	0.72

None of the effects were statistically significant, which indicates that the test outcomes for all secondary tasks were consistent across the 3 independent groups (N = 24).

The sample construction also allowed for the creation of 24-participant samples for each of the three age groups used in the study. The data were separated by age group to assess the effects of participant age group on the test outcome. Results are presented in Table 7.

Table 7. TSOT Acceptance Criterion Frequencies and Proportions by Participant Age Group

	Younger (18 – 25)		Middle (26 – 59)		Older (60 – 75)	
Secondary Task	N	p	N	p	N	p
Destination entry	0	0.00	0	0.01	0	0.00
Dialing 10-digit	11	0.46	8	0.33	2	0.08
Dialing 7-digit	20	0.83	18	0.75	13	0.54
Radio tuning	22	0.92	22	0.92	12	0.50

Age effects are apparent. For each secondary task type, the proportion of older participants associated with TSOT values meeting the acceptance criterion was considerably smaller than that associated with the two younger groups. The overall effect, collapsed across secondary tasks, was statistically significant (χ^2 (3) = 113.4, $p < .0001$). Statistical tests were conducted separately for each secondary task and the results are presented in Table 8.

Table 8. Statistical Test Results of TSOT Differences among Age Groups

Secondary Task	Age Group Differences χ^2	df	P
Destination entry	0	2	1.00
Dialing 10-digit	8.47	2	0.01
Dialing 7-digit	5.24	2	0.07
Radio tuning	16.07	2	0.003

Among the three age groups, significant differences were observed for all tasks except destination entry. Although not tested explicitly, the pattern of results shown in Table 7 indicates that the differences were primarily between the older and the two younger age groups.

In addition to age effects, there are several factors that may have contributed to the test outcome. First, the use of the touch screen press time to define the end of each trial may have added time to the TSOT that was not actually required to complete the task. Participants were instructed to say "done" aloud at the completion of the task, and then touch the screen. The use of the touch screen times for defining the end of each trial is considerably easier for data reduction than the alternatives; however, the possibility exists that some participants were either late or inaccurate in their touch, particularly if the screen touch was attempted during an occluded interval. Touch screen times could therefore include an unwanted delay that was not related to task performance, or included task performance plus a simple operator response time. This potential problem is mitigated to some extent by the likelihood that participants would verbally indicate task completion at the end of a shutter-open interval. If so, then the time between the utterance and touch screen press would occur during the shutter-closed interval and would thus not add to the TSOT value. Moreover, any additional time associated with the touch screen press would likely correspond to an experimenter's response time if a stop watch were being used to record the TSOT times.

In an attempt to increase the precision of the end time, software was developed to identify the precise time at which the "done" utterance was made by the participants. TSOT values were computed based on these times. Table 9 presents the characteristics of the overall distribution.

Table 9. TSOT Summary Statistics by Secondary Task ("Done" Utterance End Time)

Secondary Task	N	Mean TSOT (s)	SD	P ≤ 9 s
Destination entry	297	21.48	4.93	0.00
Dialing 10-digit	297	9.77	2.46	0.51
Dialing 7-digit	302	7.25	1.94	0.85
Radio tuning	320	7.38	2.29	0.86

The numbers of trials used to compute the summary statistics were smaller than those shown in Table 3, which was based on screen touch. This reflects the fact that the software was unable to identify "done" utterances on approximately 13% of the trials. Relative to the values shown in Table 3, the mean TSOT values based on the "done" utterance (shown in Table 9) decreased for all secondary tasks, on average by 0.51 seconds. This reflects the fact that the "done" utterance generally occurred before the touch screen press. This resulted in an increase in the proportions of individual trials meeting the acceptance criterion. Specifically, of 1405 total trials, 189 (13.0%) had no discernible utterance. Of the remaining 1216 trials, 75 (6.2%) moved from 'Does Not Meet Acceptance Criteria' to 'Meets Acceptance Criteria' when the end time was tied to the utterance. The difference in acceptance criteria categorizations between the two end time definitions was statistically significant (χ^2 (1) = 948.9, $p < .0001$).

4.1.1.1 Alternate Metrics

The ISO occlusion protocol (ISO, 2007) does not specify whether an outlier should be identified within the context of the 5 trials obtained from a single participant or from the larger distribution of TSOT values obtained from all participants for a given secondary task. Both approaches have potential problems. For example, 5 trials are too few to sufficiently define a distribution for determining if one or more of the values may appear "outside the overall pattern." In contrast, using the entire distribution of TSOT can be difficult to interpret if the sample contains significant heterogeneity, as might be expected with the use of a wide range of ages, as in the present study. As noted above, the TSOT values identified as outliers in the aggregated distribution were primarily trials of older participants. This suggests that at least some may have been real data, which may not have appeared to be outside the distribution if data only from the older group had been examined separately.

The problem of determining the best estimator of central tendency among a sample of measures is a common problem in statistics. Interpreting the mean as a measure of central tendency typically assumes a Gaussian (normal) distribution; however, many empirical distributions do not satisfy this assumption. This is often true for task completion times, especially for short-duration tasks. Rosenberger and Gasko (1983) considered alternatives to the mean, including the median and metrics derived by removing the extreme values from a sample. For groups of n = 5, which corresponds to the present situation, they demonstrated that the median and a broadened median (BMED) provided better efficiency in estimating the central tendency of a distribution than the mean for a range of situations, including distributions of different shapes. While the median uses only the midpoint of the sample of 5, the BMED is a compromise between the mean and median, constructed by eliminating the high and low value of each set of 5 and taking the mean of the remaining 3 values. The use of either the median or BMED has the added advantage of eliminating the ambiguity associated with identifying outliers. The effects of these alternate metrics were explored in the following analyses.

Table 10 presents the frequencies and proportions of participants with TSOT values less than or equal to 9 seconds based on different ways of computing TSOT. The ISO approach with outliers removed is compared to three approaches that do not involve removal of outliers explicitly. Both the median and BMED systematically eliminate high and low values without consideration of whether or not they may be outliers.

Table 10. Frequency and Proportion of Participants with TSOT ≤ 9 Seconds (N = 72)

	Outliers Removed		**No Outliers Removed**					
	Mean		**Mean**		**Median**		**BMED**	
Secondary Task	**N**	**p**	**N**	**p**	**N**	**p**	**N**	**p**
Destination entry	0	0.00	1	0.01	1	0.01	2	0.03
Dialing 10-digit	21	0.29	21	0.29	26	0.36	25	0.35
Dialing 7-digit	51	0.71	50	0.69	58	0.81	53	0.74
Radio tuning	56	0.78	56	0.78	62	0.86	60	0.83

Two observations are noteworthy. First, the removal of outliers had only a minor impact on test outcomes based on mean values. Second, both the median and BMED computation approaches were consistently associated with an increased number of scores meeting the acceptance criterion. With the exception of destination entry, median values generally had the highest number of scores meeting the acceptance criterion, with BMED most often being intermediate between the mean and median.

Table 11 presents the mean TSOT values across all 72 participants. TSOT values for individual participants computed using the Mean (of 5 trials, with outliers removed), the Median and BMED metrics (with upper and lower values systematically removed).

Table 11. Mean TSOT Values Based on 3 Computational Methods

	Computed with Mean		**Computed with Median**		**Computed with BMED**	
Secondary Task	**Mean (s)**	**SD**	**Mean (s)**	**SD**	**Mean (s)**	**SD**
Destination entry	21.91	3.99	21.41	4.24	21.59	4.39
Dialing 10-digit	10.50	2.40	10.43	2.83	10.43	2.76
Dialing 7-digit	7.89	1.65	7.76	1.67	7.77	1.68
Radio tuning	7.84	1.71	7.55	1.61	7.64	1.69

The trend indicates that the mean TSOT values decreased slightly when the Median or BMED approach is used to compute TSOT for a group of 5 trials. This suggests that even with the outliers removed from the mean computation, the remaining highest values in sets of 5 trials have slightly greater influence than the lowest values. (If the influence were equivalent, there would be no discernible differences between means computed using TSOT means and those computed using medians or BMED values.)

4.1.2 Computation of Total Eyes-Off-Road Time from DFD Eye Tracker Data

DFD driving protocol metrics included total eyes-off-road time (TEORT). The TEORT required to perform the secondary task was estimated using an algorithm that searched aggregated eye position data for each secondary task instance to identify the two regions with the highest density of samples. The regions were defined to be circles of 8 degree radius, following work done by Victor, Harbluk and Engstrom (2005). The regions represent the point of gaze (1) when the driver is looking forward and engaged in car following and (2) when the driver is looking down and toward the secondary task display. The regions are referred to as the primary and secondary regions, respectively.

Primary and secondary regions were defined for each secondary task instance, rather than once for each 3-minute driving trial, due to changes in absolute gaze position values that can occur with the eye tracking system, which may be due to small changes in drivers' seated position or head position that occur over time. Previously, this approach has shown that the primary region is robust and generally contains almost all gaze position samples associated with looking forward. The secondary region is somewhat more variable for several reasons, including the fact that the eye tracker accuracy is more likely to deteriorate as the driver's point of gaze moves farther away from the view straight ahead, which is the location best covered by the eye tracking cameras.

In addition, the actual secondary region is likely to vary in size and location, particularly for secondary tasks performed with hand-held devices with no fixed position. There is also a considerable amount of variability in the proportion of unassigned gaze time, defined as the time the gaze position is located outside of the primary and secondary regions. The mean proportion of unassigned gaze time was 0.18 (SD = 0.08), with the range between 0.01 and 0.60. It was therefore concluded that the values associated with the secondary region were not sufficiently robust to allow their direct use in estimating the visual cost of performing the secondary task. These findings, together with the fact that any time spent looking away from the forward view can compromise safety, led to the decision to use the total time outside the primary region as an estimate of TEORT. This estimate is referred to as TEORT1.

Recent examination of video and data files revealed potential shortcomings of this approach. Specifically, it was observed that for some participants a modest proportion of the eye position samples outside the primary region had no apparent relation to secondary task performance. Some of these samples occurred when the eye tracker temporarily lost track of the driver's eye position and the eye position was recorded erroneously. In an attempt to improve the validity of the TEORT estimate, the search algorithm was modified.

Specifically, the primary and secondary regions were computed as defined above, requiring further that the primary region be associated with looking forward. Next, each eye position sample that was neither in the primary nor secondary region was examined to determine whether or not it was in the vicinity of the secondary region. This was done by defining a set of quadrants using the position of the center of the primary region as the origin (see Figure 1). The algorithm determined the quadrant within which the secondary region resided and classified each unassigned eye position sample as being either in the same or in a different quadrant. Unassigned samples in the same quadrant (Quadrant 3 in Figure 1) were considered to be in the vicinity of the secondary task and thus assumed to be related to secondary task performance. The time associated with the samples within the vicinity of the secondary task was added to the time associated with the secondary region to obtain a revised estimate of the total eyes-off-road time associated with secondary task performance. This revised estimate is the basis for the measure referred to as TEORT2.

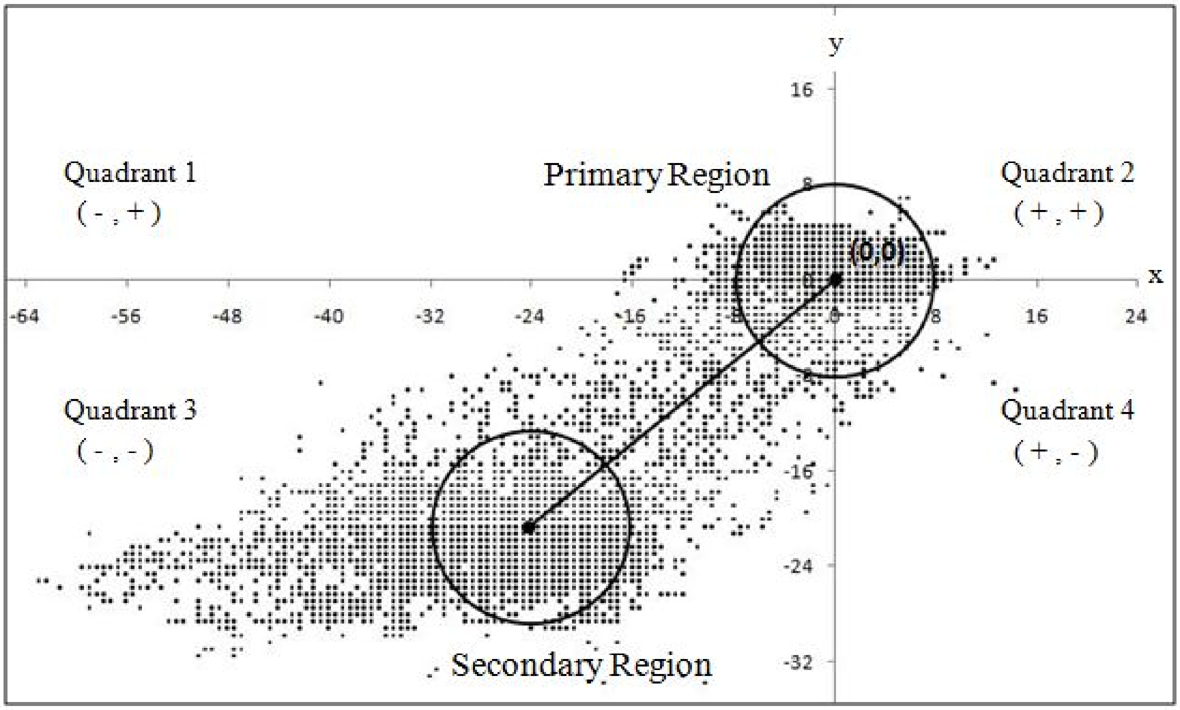

Figure 1. Regions for Classification of Eye Position Data

An additional adjustment was made to both the TEORT1 and TEORT2 metrics to eliminate any time associated with glances made in the direction of the touch screen during the interval between the participants' "done" utterance at the end of a secondary task instance and their (mostly) subsequent touch screen presses. The use of the touch screen is necessary in the DFD protocol to control and record timing of the stimulus presentation sequence during continuous secondary task performance. Using the touch screen press for segmenting the trial into task instances during data reduction is also considerably easier than determining exactly when the "done" utterance was made, particularly if manual reduction is necessary to determine the time of the utterance. It is, however, possible that the use of the touch screen approach served to introduce some amount of unwanted glance time into the TEORT estimates.

Software was developed for use with the audio record to determine the "done" utterance time. The software was generally successful, however approximately 13% of the secondary task instances did not provide a usable utterance time. This is a significant proportion and was considered to be an unacceptable proportion of data loss. To reduce the amount of loss, the median time difference between the touch screen press and "done" utterance (0.53 seconds) was used to estimate the time of utterance, when the software did not provide this information. Within the intervals between "done" utterance and touch screen press, the proportion of eye tracker samples outside of the primary region was converted to time and these times were subtracted from the estimates of TEORT. The adjustment therefore involved elimination of the time between the "done" utterance and the touch screen press during which the participant's gaze location was outside the primary visual area. The resulting metrics are AdjTEORT1 and AdjTEORT2.

While this adjustment provides a more accurate estimation of the total eyes-off-road time required to perform the secondary task, it is unknown whether this level of precision will be available to others who might adopt this approach to assessing TEORT. In particular, the times derived from automated extraction of the utterance timing are not likely to be consistent with times obtained from approaches that involve an experimenter's manual entry of "done" utterance timing. In this regard, the touch screen press time method used in the DFD protocol may actually provide time estimates closer to the expected times obtained via manual recording because both methods include a component of human response time following the "done" utterance. Thus, the adjustments made here may provide more precise estimations of the actual data collection interval, but they may also yield slightly shorter times than would be expected in practice, if "done" utterance times are entered manually. The combined effects of the two methods of TEORT estimation described above and the adjustments incorporated based on the "done" utterance timing provides four measures of TEORT, which are based on combinations of two parameters as summarized in Table 12.

Table 12. Total Eyes-Off-Road Time Measure Definitions Based on Two Parameters

Measure	Eye Positions Included	Time Interval
TEORT1	All positions outside primary region	Auditory begin to touch screen press
AdjTEORT1	All positions outside primary region	Auditory begin to "done" utterance
TEORT2	All positions outside primary region either in secondary region or in vicinity of secondary region	Auditory begin to touch screen press
AdjTEORT2	All positions outside primary region either in secondary region or in vicinity of secondary region	Auditory begin to "done" utterance

Table 13 presents the means for each of the 4 measures, separated by secondary task.

Table 13. Glance Time Means by Secondary Task for 4 Metrics

Secondary Task	Measure	N	Mean (s)	SD	P ≤ 12 s
Destination entry	TEORT1	172	26.84	8.54	0.01
	AdjTEORT1	172	26.14	8.52	0.02
	TEORT2	172	24.04	7.38	0.02
	AdjTEORT2	172	23.35	7.37	0.02
Dialing 10-digit	TEORT1	338	11.49	3.99	0.63
	AdjTEORT1	338	10.89	4.00	0.70
	TEORT2	338	9.95	3.67	0.76
	AdjTEORT2	338	9.35	3.69	0.82
Dialing 7-digit	TEORT1	427	8.24	2.72	0.92
	AdjTEORT1	427	7.60	2.69	0.94
	TEORT2	427	7.19	2.51	0.95
	AdjTEORT2	427	6.55	2.51	0.97
Radio tuning	TEORT1	450	8.31	3.33	0.89
	AdjTEORT1	450	7.76	3.26	0.91
	TEORT2	450	7.30	2.85	0.93
	AdjTEORT2	450	6.75	2.80	0.95

The overall effect of the "done" utterance adjustment (AdjTEORT1 or 2 vs. TEORT1 or 2) was a 5.5% reduction in TEORT. The overall effect of the use of secondary plus vicinity measures (TEORT2 vs. TEORT1) was a 12% reduction in TEORT. The combined effect of these two adjustments was an approximately 17.5% reduction in TEORT. Considering all completed instances of each secondary task, the final column in Table 13 shows the effects of the two computation methods and adjustments on the proportion of instances that meet the 12-second criterion proposed in the NHTSA guidelines.

Generally, both adjustments were associated with smaller estimates of TEORT; this effect is seen both in the pattern of mean values and in the proportions of secondary task instances that satisfy the proposed 12-second criterion. (Note, however that these proportions are based on the aggregated data set; test outcomes based on means computed for individual participants may differ slightly.) With respect to the specific secondary tasks, the results suggest that both 7-digit dialing and radio tuning would likely meet the acceptance criteria based on a 12-second criterion, since both tasks had more than 87.5% of the samples with TEORT values not greater than 12 seconds. This outcome was not affected by the method of TEORT computation or by the adjustment for the "done" utterance. In contrast, the results suggested that 10-digit dialing would likely not meet the acceptance criteria because the proportions were less than 0.875 for all computation methods. The same outcome was found with destination entry, which had almost no task instances with TEORT values less than 12 seconds.

Although not apparent for this set of secondary tasks, the pattern of proportions shown in Table 13 for 10-digit dialing suggests that the specific computation method adopted could alter the outcome for a task with results slightly closer to the acceptance criterion borderline. Specifically, for 10-digit dialing the proportion of trials with TEORT values less than or equal to 12 seconds increased from 0.63 to 0.82 depending on the computation method, which represents a 30% increase in projected test outcome. It is important to note that this presentation is based on the data considered in the aggregate. This approach differs from the test protocol, which requires mean values to be computed for each participant. However, given the relatively large number of samples, these values provide a good indication of the likely test outcomes, which are considered below.

Participants performed secondary tasks continuously during each 3-minute driving trial. The total number of instances completed and the average number per participant (N = 72) are presented for each secondary task in Table 14.

Table 14. Average Number of Task Instances Completed in Driving Protocol per Participant

Secondary Task	**N**	**Mean**	**SD**
Destination entry	172	2.46	0.87
Dialing 10-digit	338	4.69	1.11
Dialing 7-digit	427	5.93	1.20
Radio tuning	450	6.25	1.68

Unlike the occlusion protocol, the number of completed secondary task instances differed by task and by participant. Radio tuning had the most completions (450), while destination entry had the fewest (172). On average, participants were able to complete many more instances of radio tuning (450/72 = 6.25) and 7-digit dialing (427/72 = 5.93) than destination entry (172/72 = 2.46). Continuous secondary task performance during a 3-minute drive required some amount of "overhead" between the end of one task instance and the beginning of the next. During this interval, the participant was required to move the system from the end state of the just completed task instance to the beginning state for the next instance. The data collection process was designed such that touch screen presses defined the beginning and end of this overhead interval. The button press following the "done" utterance signaled the end of one task instance, while the subsequent button press signaled the beginning of the next task instance. Accordingly, these non-task-related intervals were eliminated from the data, and glances made to the secondary task screen during these intervals were not included in the computation of TEORT.

4.1.2.1 Assessment of Glance Metrics Relative to Proposed Guidelines Criteria

A secondary task meets the proposed Guidelines acceptance criteria if it meets all three criteria presented in Table 15.

Table 15. Glance-Based Guidelines Acceptance Criteria

Metric	Time Interval	Fixed Criterion
Percent of glances away from forward road scene lasting more than 2.0 seconds	One instance of secondary task performance	No more than 15 percent (rounded up) for at least 21of 24 test participants
Mean duration of all glances away from forward road scene	One instance of secondary task performance	Less than 2.0 seconds for at least 21of 24 test participants
Sum of the glance durations away from forward road scene	One instance of secondary task performance	No more than 12.0 seconds for at least 21of 24 test participants

For each participant, the mean TEORT was computed using all completed instances of each secondary task. These participant mean values were then compared with the proposed 12-second TEORT criterion. Table 16 and Table 17 present the test outcome results for the 72 participants, using the AdjTEORT1 and AdjTEORT2 metrics, respectively.

Table 16. Number of Participants Meeting 12-Second TEORT Criterion by Secondary Task (Based on AdjTEORT1 Metric)

	Meets Criteria		Does Not Meet Criteria		Total	
Secondary Task	**N**	**p**	**N**	**p**	**N**	**p**
Destination entry	0	0.00	72	1.00	72	1.00
Dialing 10-digit	48	0.67	24	0.33	72	1.00
Dialing 7-digit	68	0.94	4	0.06	72	1.00
Radio tuning	64	0.89	8	0.11	72	1.00

Table 17. Number of Participants Meeting 12-Second TEORT Criterion by Secondary Task (Based on AdjTEORT2 Metric)

	Meets Criteria		Does Not Meet Criteria		Total	
Secondary Task	**N**	**p**	**N**	**p**	**N**	**p**
Destination entry	0	0.00	72	1.00	72	1.00
Dialing 10-digit	57	0.79	15	0.21	72	1.00
Dialing 7-digit	69	0.96	3	0.04	72	1.00
Radio tuning	69	0.96	3	0.04	72	1.00

For both metrics, radio tuning and 7-digit dialing had a sufficient number of participants with values less than 12 seconds to satisfy the 0.875 (21/24 or 63/72) acceptance criteria. While test outcomes were not affected, the AdjTEORT2 metric was associated with higher proportions of outcomes meeting the acceptance criteria, due to its elimination of glances that occurred following the "done" utterance and were thus presumed not to be relevant to secondary task performance. The AdjTEORT2 metric, which uses the combination of time spent looking within and in the vicinity of the secondary region, is used for the next set of analyses.

Table 18 presents the test outcome summary for the 3 independent groups of 24 participants constructed with the same age category composition. The test outcome results for the three independent (N = 24) groups matched the test outcome results for all 72 participants; both radio tuning and 7-digit dialing satisfied the 0.875 acceptance criterion in all three groups of 24 participants. The test outcome results for the other two tasks (10-digit dialing and destination entry) were also consistent across the three groups.

Table 18. Number of Participants Meeting 12-Second TEORT Criterion by Secondary Task and 24-Participant Group

	Group 1		Group 2		Group 3	
Secondary Task	**N**	**p**	**N**	**p**	**N**	**P**
Destination entry	0	0.00	0	0.00	0	0.00
Dialing 10-digit	20	0.83	18	0.75	19	0.79
Dialing 7-digit	24	1.00	22	0.92	23	0.96
Radio tuning	22	0.92	23	0.96	24	1.00

A statistical comparison of the respective group frequencies provides a slightly stronger test of test-retest reliability. Accordingly, each secondary task was tested individually. The results are presented in Table 19.

Table 19. Results of Statistical Tests of TEORT Differences among Groups

	Group Differences		
Secondary Task	χ^2	**df**	**p**
Destination entry	0	2	1.00
Dialing 10-digit	0.51	2	0.78
Dialing 7-digit	2.09	2	0.35
Radio tuning	2.09	2	0.35

None of the secondary tasks exhibited differences among the three groups that approached statistical significance. Acceptance criterion outcome results are presented for the three age groups in Table 20.

Table 20. Number of Participants Meeting 12-Second TEORT Criterion by Secondary Task and Age Group

	Younger (18 – 25)		Middle (26 – 59)		Older (60 – 75)	
Secondary Task	**N**	**p**	**N**	**p**	**N**	**p**
Destination entry	0	0.00	0	0.00	0	0.00
Dialing 10-digit	23	0.96	19	0.79	15	0.63
Dialing 7-digit	24	1.00	24	1.00	21	0.88
Radio tuning	24	1.00	24	1.00	21	0.88

The results indicate that 7-digit dialing and radio tuning were consistent with the acceptance criterion for participants in all three age groups. Note that among older participants, the minimum level of participants (i.e., 21 of 24) met the 12-second criterion. For the 10-digit dialing task, the younger age group met the criterion, while the other two age groups did not.

Differences among age groups were tested statistically. Each secondary task was tested individually. The results are presented in Table 21.

Table 21. Results of Statistical Tests of TEORT Differences among Age Groups

Secondary Task	Age Group Differences		
	χ^2	df	p
Destination entry	0	2	1.00
Dialing 10-digit	8.08	2	0.02
Dialing 7-digit	6.26	2	0.04
Radio tuning	6.26	2	0.02

Differences among the age groups are statistically significant for three secondary tasks. Although not tested explicitly, the outcomes for the older group were generally different from those for the younger and middle aged groups. In particular, older participants were significantly less likely to comply with the 12-second TEORT criterion than participants in the other groups.

4.1.3 Proportion of Long Glances

Analyses of TEORT presented above utilized eye position data in the aggregate for each task instance and did not consider the characteristics of individual glances. Analyses that characterize glances required a different analysis approach, in which each glance away from the forward roadway was defined and characterized in terms of location and duration. Glances away from the forward roadway view recorded by the eye tracker were analyzed first in the aggregate and then collapsed across trial and segment (each instance of a secondary task performance is referred to as a segment). For this presentation, glances with duration less than 0.1 second or greater than 5 seconds were eliminated as erroneous. In addition, only glances deemed to be relevant to secondary task performance were included in these analyses. Relevant glances were defined as those that involved some portion of time either in the secondary region or in the vicinity of the secondary region, as defined previously. Table 22 shows the glance duration summary statistics by secondary task for all eligible glances.

Table 22. Relevant Glance Duration Summary Statistics by Secondary Task

Secondary Task	N	Mean (s)	SD	P > 2 s
Destination entry	4733	1.00	0.48	0.04
Dialing 10-digit	4401	0.95	0.47	0.03
Dialing 7-digit	3987	0.95	0.47	0.03
Radio tuning	4206	0.92	0.42	0.02

In the aggregate, all tasks had mean glance durations of one second or less and all tasks had no more than 4 percent of glances with durations longer than 2 seconds.

4.1.3.1 Glance Duration Summarized over Each Secondary Task Instance

Table 23 presents the proportions of drivers that had fewer than 15% of glances away from the forward view longer than 2 seconds for each secondary task. The data represent multiple instances of each secondary task, which differed by participant and by secondary task. For example, destination entry was the most time consuming of the tasks and many participants did not finish more than one destination during the 3-minute drive. It is for this reason that the denominator for the BMED computation is so much smaller than for the other computation methods. The BMED computation required a minimum of 3 secondary task instances to allow the two extreme values to be trimmed.

Table 23. Proportion of Drivers with Less Than 15% of Glances Longer Than 2 Seconds: Three Computation Methods to Summarize over Multiple Task Instances

	Mean		Median		BMED	
Secondary Task	**N**	**p**	**N**	**p**	**N**	**p**
Destination entry	67/72	0.93	67/72	0.93	38/40	0.95
Dialing 10-digit	64/72	0.89	67/72	0.93	65/70	0.86
Dialing 7-digit	68/72	0.94	68/72	0.94	66/70	0.94
Radio tuning	71/72	0.99	72/72	1.00	69/69	1.00

Of interest here is that all secondary tasks had at least some participants with more than 15% of the glances away from the forward view lasting longer than 2 seconds. The BMED approach appears inappropriate for use with longer tasks (destination entry) due to the relatively small number of participants who completed three task instances during the drive. Otherwise, as shown previously, metrics computed using the median of proportions associated with different task instances were slightly higher than those computed using the mean values, although for this measure the differences were minimal.

A more direct means of determining the proportion of glances with duration longer than 2 seconds involved combining all relevant glances during the 3-minute drive, during which the participants worked on the secondary task continuously. (Glances made during reset intervals were eliminated such that all glances pertained to secondary task performance.) Accordingly, the summary presented in Table 24 is based on all relevant glances that occurred during the 3-minute drive without consideration of the individual task instances (segments).

Table 24. Glance Duration Summarized over the 3-Minute Drive

	Meets Criteria		Does Not Meet Criteria		Total	
Secondary Task	**N**	**p**	**N**	**p**	**N**	**p**
Destination entry	65	0.93	5	0.07	70	1.00
Dialing 10-digit	64	0.89	8	0.11	72	1.00
Dialing 7-digit	68	0.94	4	0.06	72	1.00
Radio tuning	72	1.00	0	0.00	72	1.00

These proportions are consistent with those obtained when segmenting by task instance.

4.1.4 Mean Glance Duration

Table 25 presents the proportions of drivers for which mean glance duration was less than 2 seconds.

Table 25. Proportion of Drivers Mean Glance Duration Less than 2 Seconds: 3 Computation Methods to Summarize over Multiple Task Instances

	Mean		Median		BMED	
Secondary Task	**N**	**p**	**N**	**p**	**N**	**p**
Destination entry	72/72	1.00	72/72	1.00	42/42	1.00
Dialing 10-digit	72/72	1.00	72/72	1.00	70/70	1.00
Dialing 7-digit	72/72	1.00	72/72	1.00	72/72	1.00
Radio tuning	72/72	1.00	72/72	1.00	70/70	1.00

The results indicate that none of the participants had mean glance duration values greater than 2 seconds. This finding is consistent with the finding that when considered in the aggregate, less than 5 percent of glances had durations longer than 2 seconds.

4.1.5 Relation between TSOT and TEORT

The use of the same participants for occlusion and DFD testing allowed a direct comparison of the TSOT and TEORT distributions associated with the various secondary tasks. Across all secondary tasks, TSOT values were highly correlated with TEORT values (R square = 0.81). Two regression equations were used to explore the relation between the two metrics. Results are presented in Table 26.

Table 26. Relation between TSOT and TEORT Based on Regression Models

	TEORT when TSOT = 9 s		TSOT when TEORT = 12 s		R-squared	
Secondary Task	**TSOT**	**TEORT**	**TSOT**	**TEORT**	**N**	**R^2**
All Task Conditions	9	8.66	11.75	12	280	0.81

The results indicated that the relation between the TSOT and TEORT values obtained by the same people performing the same secondary tasks was close to equivalence. Specifically, when predicting TEORT from TSOT, it was found that TEORT = 8.66 seconds when TSOT = 9 seconds; when predicting TSOT from TEORT, it was found that TSOT = 11.75 seconds when TEORT = 12 seconds. The resulting ratio values (TEORT/TSOT) were 0.96 and 1.02, respectively, which are considerably smaller than the predicted ratio (4/3 = 1.33). Within the individual secondary tasks, correlations between TSOT and TEORT were very low, indicating weak relations between the two metrics. Therefore, mean values were compared for individual tasks rather than regression parameters. Table 27 presents the respective mean values for the TSOT and TEORT metrics and the ratio of TEORT/TSOT computed from these overall means.

Table 27. Mean Values for TSOT and TEORT Metrics

	TEORT (s)		TSOT (s)		TEORT/TSOT Ratio
Secondary Task	**Mean**	**SD**	**Mean**	**SD**	
Destination entry	24.09	5.50	21.46	3.80	1.12
Dialing 10-digit	9.26	2.57	10.25	2.36	0.90
Dialing 7-digit	6.78	2.23	7.75	1.65	0.87
Radio tuning	7.12	2.91	7.54	1.59	0.94

The relation between the mean values is consistent with the regression results. Only destination entry had more TEORT than TSOT time. SD values were generally smaller with TSOT, reflecting the constraints of the occlusion versus unconstrained driving methods.

4.2 DFD Metrics

DFD driving performance metrics included: car following delay, standard deviation of lane position (SDLP), target detection accuracy (proportion correct), and target detection response time. The NHTSA Guidelines propose two sets of criteria for these DFD metrics, one based on fixed criteria and one based on values associated with a specified benchmark task. Analyses presented in the following sections consider both proposed approaches.

4.2.1 Comparison with Fixed Criteria Values

The specific fixed criteria for DFD metrics are presented in Table 28.

Table 28. DFD Metric Criteria

DFD Metric	Criterion Value	Test: Meets Acceptance Criterion
Car Following Delay	4.6 seconds	Sample mean not greater than criterion
Standard Deviation of Lane Position	1.0 foot	Sample mean not greater than criterion
Target Detection Proportion Correct	0.80	Sample mean greater than criterion
Target Detection Response Time	1.0 second	Sample mean not greater than criterion

To establish acceptance, a secondary task must meet at least three of the four criteria shown in Table 28. DFD metrics summarize performance over a 3-minute drive. Although not shown in the table, the proposed Guidelines specify that all criteria be established with a 95 percent confidence level, which is interpreted to mean that for positive differences, the probability of a Type I error (alpha) is 0.05.

Three of the tests are defined such that the sample mean must be demonstrably *not* greater than the respective criterion value. In the context of the traditional model for hypothesis testing in the experimental method, this test requirement is akin to demonstrating that the null hypothesis is true, which is formally not possible. In practice, while this demonstration is possible, its interpretation can be confounded by the effects of inadequate statistical power due to small sample sizes. In this context, the 95 percent confidence level is interpreted to mean that the test fails to demonstrate a statistically significant difference using alpha = .05[3] The comparisons with specific fixed criteria were done using one-sample, one-sided *t* tests, in which the sample means were compared with the criterion values shown in Table 28. The one-sided *t* tests provide a modest improvement in statistical power over the two-sided *t* tests.

4.2.1.1 Car Following Delay

Delay is a generalized measure of response time in car following. Higher values indicate that drivers are slower in adapting to the momentary speed changes of the lead vehicle, reflecting increasing performance degradation due to concurrent secondary task performance. The validity of delay requires relatively accurate car following, which is measured by car following similarity or coherence. Degradation of car following performance typically involves two stages. In the first stage, drivers follow the lead vehicle accurately and delay values increase with the combined demands of primary and secondary tasks. Specifically, in lower demand situations, drivers follow accurately and delay values are relatively small. In moderate demand situations, drivers typically follow accurately and delay values increase. However, as task demands continue to increase, car following performance eventually breaks down, which is reflected in very low similarity (or coherence) values. The resulting delay values are typically not valid. Traditionally, researchers have set a criterion value for similarity, such that when similarity values dip below this criterion, the delay is not considered valid.

The disposition of trials with invalid delay values in research protocols can include either elimination or replacement of participants if car following performance does not meet a predefined car following

---

[3] However, one could interpret this proposed requirement to mean that the probability of a Type II error be no more than 0.05. A Type II error occurs when the test results indicate that there is no difference, when in reality there is a difference. To ensure that the probability of a Type II error is no more than 0.05, it is necessary to design a test that has power of 1 – beta = 0.95. This is considerably more difficult than setting a criterion alpha level of 0.05 and accepting as true the finding of no difference, with a sample of a given size.

similarity criterion consistently across task conditions. However in the context of compliance testing, which could involve time-constrained testing of relatively small samples, neither approach may be practical. Accordingly, because invalid delay values are virtually always associated with extremely degraded car following performance, an alternate disposition was used here. Specifically, the invalid delay values were replaced with 8.5 seconds, which represents the 99th percentile of the distribution of delay values presumed to be valid (8.5 = mean + 3 times the standard deviation (SD) or 3.4 + (3 x 1.7)) and reflects the poorest car following performance consistent with the overall distribution. For purposes of comparison, the results are presented first with invalid trials removed and second with invalid trials replaced by the estimated 99th percentile values.

To evaluate the secondary tasks in the context of the specific criteria, one-sample one-tail *t* tests were performed using the specific criteria as H_0 values. The probability values shown in the rightmost column reflect the likelihood that the sample mean value is statistically greater than the criterion (4.6 seconds).

Table 29 and Table 30 show the results of the one-sided *t* tests for the car following delay metric.

Table 29. One-Sided *t*-Test Results for Car Following Delay: Invalid Data Removed

Condition	N	Mean	Std Dev	Std Err	95% CL Mean	df	t Value	Pr > t
Baseline	72	1.8995	0.7352	0.0866	1.7551	71	-31.17	1
Destination entry	69	4.1256	1.5866	0.191	3.8071	68	-2.48	0.9923
Dialing 10-digit	70	3.6214	1.6335	0.1952	3.2959	69	-5.01	1
Dialing 7-digit	70	3.6381	1.4434	0.1725	3.3505	69	-5.58	1
Radio tuning	69	4.0498	1.9129	0.2303	3.6657	68	-2.39	0.9902

Table 30. One-Sided *t*-Test Results for Car Following Delay: Invalid Data Replaced

Condition	N	Mean	Std Dev	Std Err	95% CL Mean	df	t Value	Pr > t
Baseline	72	1.8995	0.7352	0.0866	1.7551	71	-31.17	1
Destination entry	72	4.3079	1.7848	0.2103	3.9573	71	-1.39	0.9154
Dialing 10-digit	72	3.7569	1.8014	0.2123	3.4031	71	-3.97	0.9999
Dialing 7-digit	72	3.7731	1.6346	0.1926	3.4521	71	-4.29	1
Radio tuning	72	4.2352	2.0752	0.2446	3.8276	71	-1.49	0.9299

The column labeled '95% CL Mean' is the lower boundary of the 95% confidence interval (CI) around the mean. This value represents the critical value for statistical testing. Specifically, for a criterion of 4.6 seconds, the statistical significance is determined by the relation of the 95th percentile CL mean value to the criterion value. In the tables above, none of the CL values are greater than 4.6 seconds, which indicates that none of the mean values are deemed to be statistically greater than 4.6 seconds.

With respect to the statistical testing outcomes, there were no differences between the two approaches for handling invalid delay values; however the descriptive statistics associated with the replacement approach were consistently greater than those associated with the elimination of data from invalid trials, as would be expected.

4.2.1.1.1 Demonstration of Car Following Test Results Using a Hypothetical More Restrictive Criterion

For demonstration purposes, Table 31 and Table 32 were created using a hypothetical criterion of 3.8 seconds, which was selected based on the emerging consensus that destination entry is not acceptable for driving. Extreme trials were removed for the first analysis.

Table 31. One-Sided *t*-Test Results for Car Following Delay with Hypothetical Restrictive Criterion (3.8 seconds): Invalid Data Removed

Condition	N	Mean	Std Dev	Std Err	95% CL Mean	df	t Value	Pr > t
Baseline	72	1.8995	0.7352	0.0866	1.7551	71	-21.93	1
Destination entry	69	4.1256	1.5866	0.191	3.8071	68	1.7	0.0464
Dialing 10-digit	70	3.6214	1.6335	0.1952	3.2959	69	-0.91	0.8182
Dialing 7-digit	70	3.6381	1.4434	0.1725	3.3505	69	-0.94	0.8244
Radio tuning	69	4.0498	1.9129	0.2303	3.6657	68	1.08	0.141

Here, the statistical test results indicate that the mean value for the destination entry task is significantly greater than the hypothetical restrictive criterion (p = 0.0464). The marginal statistical significance reflects the fact that the lower CI value (3.8071) is marginally greater than the hypothetical criterion. In Table 32, the same demonstration is presented using data for which the invalid values have been replaced with the 99th percentile values, derived as described above.

Table 32. One-Sided *t*-Test Results for Car Following Delay with Hypothetical Restrictive Criterion (3.8 seconds): Invalid Data Replaced

Condition	N	Mean	Std Dev	Std Err	95% CL Mean	df	t Value	Pr > t
Baseline	72	1.8995	0.7352	0.0866	1.7551	71	-21.93	1
Destination entry	72	4.3079	1.7848	0.2103	3.9573	71	2.41	0.0092
Dialing 10-digit	72	3.7569	1.8014	0.2123	3.4031	71	-0.2	0.5801
Dialing 7-digit	72	3.7731	1.6346	0.1926	3.4521	71	-0.14	0.5552
Radio tuning	72	4.2352	2.0752	0.2446	3.8276	71	1.78	0.0397

Two secondary tasks (destination entry and radio tuning) had mean delay values that were statistically greater than the hypothetical criterion of 3.8 seconds. Comparing this result to the previous table in which the invalid values were removed, it is apparent that the statistical significance of one result, namely that for radio tuning, was affected by the differential treatment of the outliers. When outliers were removed, the radio tuning task was not statistically greater than 3.8 seconds; however, when the outliers were changed to 99th percentile values, the test mean value increased from 4.05 seconds to 4.24 seconds. The lower CI value increased from 3.67 seconds to 3.83 seconds and because the latter value is greater than the hypothetical criterion (3.8 seconds), the test result was statistically significant. Both the selection of test criterion and the treatment of suspected outliers influenced statistical test outcome.

4.2.1.1.2 Effects of Sample Size, Using the Car Following Delay Metric

The results presented immediately above derive from use of the full data sample (N = 72). In routine testing, using samples of N = 24, the relation between the sample mean and the lower CI value will be slightly different than for the larger sample. Specifically, the smaller sample size will be associated with a slightly wider confidence interval, the result of which is that the 95th percentile CL (lower CI) value will be smaller than for the larger sample. It is the relation of this CL value to the criterion that determines statistical significance. In particular, the (lower) CL value must be greater than the specified test criterion (4.6 seconds). A wider CI for smaller samples, characterized by a smaller CL value, implies that the probability of a statistically significant outcome (CL > criterion) decreases with decreasing sample size.

The following tables (Table 33 through Table 35) present the one-sided *t* test outcomes based on the 4.6 second criterion. Tests were conducted separately on the three independent samples of similar demographic construction.

Table 33. One-Sided *t*-Test Results for Car Following Delay with Invalid Data Replaced: N = 24, Group 1

Condition	N	Mean	Std Dev	Std Err	95% CL Mean	df	t Value	Pr > t
Baseline	24	2.2097	0.7112	0.1452	1.9609	23	-16.47	1
Destination entry	24	4.6250	1.892	0.3862	3.9631	23	0.06	0.4745
Dialing 10-digit	24	3.8972	1.5373	0.3138	3.3594	23	-2.24	0.9825
Dialing 7-digit	24	4.2736	1.8924	0.3863	3.6116	23	-0.84	0.7966
Radio tuning	24	4.5986	1.8069	0.3688	3.9665	23	0	0.5015

Table 34. One-Sided *t*-Test Results for Car Following Delay with Invalid Data Replaced: N = 24, Group 2

Condition	N	Mean	Std Dev	Std Err	95% CL Mean	df	t Value	Pr > t
Baseline	24	1.7736	0.8121	0.1658	1.4895	23	-17.05	1
Destination entry	24	4.1375	1.7946	0.3663	3.5097	23	-1.26	0.8903
Dialing 10-digit	24	4.1458	2.1613	0.4412	3.3897	23	-1.03	0.843
Dialing 7-digit	24	3.7194	1.7319	0.3535	3.1135	23	-2.49	0.9898
Radio tuning	24	4.4361	2.4324	0.4965	3.5852	23	-0.33	0.6278

Table 35. One-Sided *t*-Test Results for Car Following Delay with Invalid Data Replaced: N = 24, Group 3

Condition	N	Mean	Std Dev	Std Err	95% CL Mean	df	t Value	Pr > t
Baseline	24	1.7153	0.5923	0.1209	1.5081	23	-23.86	1
Destination entry	24	4.1611	1.6959	0.3462	3.5678	23	-1.27	0.8912
Dialing 10-digit	24	3.2278	1.5848	0.3235	2.6733	23	-4.24	0.9998
Dialing 7-digit	24	3.3264	1.0931	0.2231	2.944	23	-5.71	1
Radio tuning	24	3.6708	1.8974	0.3873	3.007	23	-2.4	0.9875

Relative to the 4.6 second criterion, the test results are all consistent; none of the tasks provided delay values that are significantly greater than the criterion values. For a slightly stronger determination of test-retest reliability, the results in these three tables can be compared with the hypothetical 3.8 second criterion, described previously. Specifically, the 95% CL values can be compared with 3.8 seconds, such that any value greater than 3.8 seconds would indicate a statistically greater delay value. Among the tests represented in the three tables, the results indicate that two of the secondary tasks, namely destination entry and radio tuning, had statistically greater delay values relative to the hypothetical 3.8 second criterion for Group 1. Respective tests in Groups 2 and 3 were not different. This demonstrates that for borderline secondary tasks, it is likely that test outcomes may vary from sample to sample when the sample size is 24 subjects.

4.2.1.2 Standard Deviation of Lane Position (SDLP)

SDLP is a measure of lateral vehicle control, specifically of how well the driver keeps the vehicle in the center of the travel lane. Increasing SDLP values reflect degraded lateral control performance and typically indicate that a driver is not paying adequate attention to steering the vehicle. The proposed NHTSA criterion for SDLP is 1 foot. Table 36 through 39 present the results of the one-sided *t*-tests for the SDLP metric, first for the full sample and then for each of the 3 independent samples of similar demographic construction.

Table 36. One-Sided *t*-Test Results for SDLP: Full Sample (N = 72)

Condition	N	Mean	Std Dev	Std Err	95% CL Mean	df	t Value	Pr > t
Baseline	72	0.7669	0.2262	0.0267	0.7224	71	-8.75	1
Destination entry	72	1.1659	0.5996	0.0707	1.0481	71	2.35	0.0108
Dialing 10-digit	72	1.0229	0.2951	0.0348	0.9649	71	0.66	0.2565
Dialing 7-digit	72	1.0548	0.3682	0.0434	0.9825	71	1.26	0.1053
Radio tuning	72	1.057	0.3485	0.0411	0.9886	71	1.39	0.0846

Statistical test results indicate that the destination entry task was associated with SDLP values significantly greater than 1.0 foot. It is noteworthy that all secondary task conditions were associated with mean values greater than 1.0 foot; however only destination entry had the lower CI values greater than 1.0 foot.

Table 37. One-Sided *t*-Test Results for SDLP: Group 1 (N = 24)

Condition	N	Mean	Std Dev	Std Err	95% CL Mean	df	t Value	Pr > t
Baseline	24	0.7953	0.2456	0.0501	0.7093	23	-4.08	0.9998
Destination entry	24	1.1461	0.6412	0.1309	0.9218	23	1.12	0.1379
Dialing 10-digit	24	0.9649	0.25	0.051	0.8774	23	-0.69	0.7509
Dialing 7-digit	24	1.0071	0.3941	0.0804	0.8692	23	0.09	0.4654
Radio tuning	24	0.9715	0.297	0.0606	0.8676	23	-0.47	0.6788

Table 38. One-Sided *t*-Test Results for SDLP: Group 2 (N = 24)

Condition	N	Mean	Std Dev	Std Err	95% CL Mean	df	t Value	Pr > t
Baseline	24	0.7018	0.211	0.0431	0.628	23	-6.92	1
Destination entry	24	1.0969	0.6484	0.1324	0.8701	23	0.73	0.2356
Dialing 10-digit	24	1.0113	0.3231	0.066	0.8983	23	0.17	0.4327
Dialing 7-digit	24	0.9973	0.3276	0.0669	0.8827	23	-0.04	0.516
Radio tuning	24	1.0603	0.3656	0.0746	0.9324	23	0.81	0.2135

Table 39. One-Sided *t*-Test Results for SDLP: Group 3 (N = 24)

Condition	N	Mean	Std Dev	Std Err	95% CL Mean	df	t Value	Pr > t
Baseline	24	0.8035	0.2154	0.044	0.7282	23	-4.47	0.9999
Destination entry	24	1.2546	0.5145	0.105	1.0746	23	2.42	0.0118
Dialing 10-digit	24	1.0924	0.3053	0.0623	0.9856	23	1.48	0.0758
Dialing 7-digit	24	1.1601	0.3722	0.076	1.0299	23	2.11	0.0231
Radio tuning	24	1.1393	0.3718	0.0759	1.0092	23	1.84	0.0397

If the large sample results are taken to represent the 'ground truth' with respect to test outcomes, and as a result we believe the destination entry task to have SDLP values statistically greater than the proposed criterion, then it is apparent that the results from tests using the three smaller samples provide inconsistent and potentially contradictory outcomes. For the destination entry task, only Group 3 had results consistent with the larger sample. However, Group 3 results also revealed differences that were not apparent in the larger sample. This pattern of differences raises questions about how to interpret the test results. Specifically, destination entry test results for this sample were consistent with our 'ground truth' while the differences observed for the other three tasks were not consistent with the 'ground truth' associated with the larger sample. Similarly, destination entry results for Groups 1 and 2 were not consistent with the larger sample results, but Group1 and 2 results for the other three tasks were consistent. In other words, none of the outcomes represented in the smaller groups were consistent with the results of the larger group. It should be noted that no adjustments were made here for the effects of multiple comparisons. As such, in the interim it may be prudent to adopt a more stringent criterion (e.g., $p < .01$) for interpreting differences in this context.

4.2.1.3 Detection Task Response Time

One-sided t tests were computed using $H_0 = 1$ second. Tests were computed both for the entire sample (N = 72) and for each of the groups that comprise the overall sample (N = 24). Results of these tests are presented in the following tables (Table 40 through Table 43).

Table 40. One-Sided *t*-Test Results for Detection Task Response Time: Full Sample (N = 72)

Condition	N	Mean	Std Dev	Std Err	95% CL Mean	df	t Value	Pr > t
Baseline	72	0.638	0.0652	0.00768	0.6252	71	-47.12	1
Destination entry	72	1.0076	0.2159	0.0254	0.9652	71	0.3	0.3828
Dialing 10-digit	72	0.9647	0.1558	0.0184	0.9341	71	-1.92	0.9706
Dialing 7-digit	72	0.9877	0.1622	0.0191	0.9558	71	-0.64	0.7389
Radio tuning	72	0.9817	0.159	0.0187	0.9505	71	-0.97	0.8335

None of the secondary task conditions were associated with detection task response times significantly greater than 1.0 second. The mean response time for the destination entry task was nominally higher than 1.0 second, but the difference was not statistically significant. A similar result was observed for two secondary task conditions in Group 1, in which the mean detection task response times were nominally greater than 1.0 second but not statistically greater.

Table 41. One-Sided *t*-Test Results for Detection Task Response Time: Group 1 (N = 24)

Condition	N	Mean	Std Dev	Std Err	95% CL Mean	df	t Value	Pr > t
Baseline	24	0.65	0.0718	0.0147	0.6249	23	-23.87	1
Destination entry	24	1.0462	0.1723	0.0352	0.986	23	1.31	0.1008
Dialing 10-digit	24	0.9955	0.1673	0.0342	0.9369	23	-0.13	0.5523
Dialing 7-digit	24	0.995	0.1076	0.022	0.9573	23	-0.23	0.5893
Radio tuning	24	1.0003	0.1495	0.0305	0.948	23	0.01	0.4956

Table 42. One-Sided *t*-Test Results for Detection Task Response Time: Group 2 (N = 24)

Condition	N	Mean	Std Dev	Std Err	95% CL Mean	df	t Value	Pr > t
Baseline	24	0.6254	0.0495	0.0101	0.6081	23	-37.08	1
Destination entry	24	0.9912	0.2573	0.0525	0.9011	23	-0.17	0.5662
Dialing 10-digit	24	0.9514	0.1669	0.0341	0.893	23	-1.43	0.9163
Dialing 7-digit	24	0.9701	0.1851	0.0378	0.9054	23	-0.79	0.7812
Radio tuning	24	0.9503	0.1682	0.0343	0.8915	23	-1.45	0.9193

Table 43. One-Sided *t*-Test Results for Detection Task Response Time: Group 3 (N = 24)

Condition	N	Mean	Std Dev	Std Err	95% CL Mean	df	t Value	Pr > t
Baseline	24	0.6387	0.072	0.0147	0.6135	23	-24.57	1
Destination entry	24	0.9855	0.2137	0.0436	0.9107	23	-0.33	0.629
Dialing 10-digit	24	0.9473	0.1328	0.0271	0.9008	23	-1.95	0.968
Dialing 7-digit	24	0.998	0.1868	0.0381	0.9326	23	-0.05	0.5209
Radio tuning	24	0.9945	0.1607	0.0328	0.9383	23	-0.17	0.5653

Overall, the test outcome results were consistent for detection task response time. None of the secondary tasks had response times that were statistically greater than 1.0 second.

4.2.1.4 Detection Task Proportion Correctly Detected

Table 44 through Table 47 present the results of the one-sided *t* tests for the detection task proportion correctly detected metric, first for the full sample and then for each of the 3 independent samples of similar demographic construction.

Table 44. One-Sided *t*-Test Results for Detection Task Proportion Correctly Detected: Full Sample (N = 72)

Condition	N	Mean	Std Dev	Std Err	95% CL Mean	df	t Value	Pr > t
Baseline	72	0.9909	0.024	0.00283	0.9862	71	67.47	<.0001
Destination entry	72	0.7967	0.1708	0.0201	0.7632	71	-0.16	0.5641
Dialing 10-digit	72	0.8541	0.1104	0.013	0.8325	71	4.16	<.0001
Dialing 7-digit	72	0.8364	0.1411	0.0166	0.8087	71	2.19	0.0159
Radio tuning	72	0.8374	0.1299	0.0153	0.8118	71	2.44	0.0086

Table 45. One-Sided *t*-Test Results for Detection Task Proportion Correctly Detected: Group 1

Condition	N	Mean	Std Dev	Std Err	95% CL Mean	df	t Value	Pr > t
Baseline	24	0.9878	0.0297	0.00606	0.9774	23	30.96	<.0001
Destination entry	24	0.7901	0.1336	0.0273	0.7433	23	-0.36	0.6406
Dialing 10-digit	24	0.8472	0.1148	0.0234	0.807	23	2.01	0.0279
Dialing 7-digit	24	0.8416	0.1037	0.0212	0.8053	23	1.97	0.0308
Radio tuning	24	0.8468	0.1235	0.0252	0.8036	23	1.86	0.0381

Table 46. One-Sided *t*-Test Results for Detection Task Proportion Correctly Detected: Group 2

Condition	N	Mean	Std Dev	Std Err	95% CL Mean	df	t Value	Pr > t
Baseline	24	0.9957	0.0125	0.00256	0.9913	23	76.54	<.0001
Destination entry	24	0.8187	0.1866	0.0381	0.7534	23	0.49	0.3145
Dialing 10-digit	24	0.8662	0.1011	0.0206	0.8309	23	3.21	0.0019
Dialing 7-digit	24	0.8613	0.1284	0.0262	0.8163	23	2.34	0.0142
Radio tuning	24	0.862	0.108	0.022	0.8242	23	2.81	0.0049

Table 47. One-Sided *t*-Test Results for Detection Task Proportion Correctly Detected: Group 3

Condition	N	Mean	Std Dev	Std Err	95% CL Mean	df	t Value	Pr > t
Baseline	24	0.9892	0.0265	0.00541	0.98	23	34.98	<.0001
Destination entry	24	0.7815	0.1912	0.039	0.7146	23	-0.47	0.6799
Dialing 10-digit	24	0.849	0.1183	0.0241	0.8076	23	2.03	0.0271
Dialing 7-digit	24	0.8064	0.1805	0.0368	0.7432	23	0.17	0.432
Radio tuning	24	0.8033	0.1524	0.0311	0.7499	23	0.1	0.4588

Interpretation of these results differs from the other DFD metrics because the acceptance criterion for this metric requires performance to be better than the specified criterion, rather than performance no worse than the criterion, which reflects the construction of the other three DFD metrics used with fixed values (see Table 28). The results for the large sample (N = 72) in Table 44 reveal that all secondary tasks except destination entry had target detection proportions that were statistically greater than 0.80, reflecting the conclusion that all tasks except destination entry would be acceptable relative to this criterion, given this sample size. The results for the smaller samples were not entirely consistent with this pattern. Results for Groups 1 and 2 were consistent with the larger sample results, but the results for Group 3 were different.

4.2.2 DFD Metrics Compared with Benchmark

One NHTSA Guidelines option proposes the use of destination entry as a benchmark value, representing a task that is not acceptable for performance by the driver while driving. Accordingly, for a task to be acceptable, it would require metric values that are statistically better in terms of driving performance degradation than those associated with destination entry. Tasks that are not statistically better are presumed to have approximately the same effect on driving performance, which is unacceptable. The results for the 4 DFD metrics are presented in the following sections.

4.2.2.1 Standard Deviation of Lane Position (SDLP)

Results of planned comparisons for SDLP are presented in Table 48. Probability values are adjusted for familywise error using the Hochberg procedure. Group means (± standard error) are presented in Figure 2.

Table 48. SDLP Planned Comparison Results

Comparison	df	F Value	PR > F
Destination Entry vs. Dialing 10-Digit	1,207	9.11	0.0086
Destination Entry vs. Dialing 7-Digit	1,207	5.5	0.0226
Destination Entry vs. Radio Tuning	1,207	5.28	0.0226

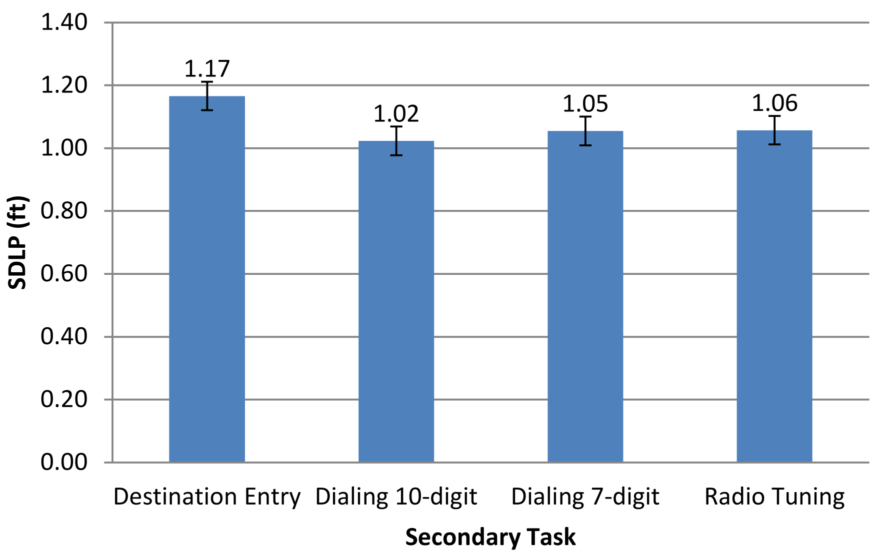

Figure 2. Mean SDLP by Secondary Task (N = 72)

Statistical test results indicate that all three tasks are statistically better than destination entry for the SDLP metric using this approach. Among the mean values, 10-digit dialing was associated with the least amount of performance degradation.

4.2.2.2 Car Following Delay

Results of the statistical tests for car following delay comparing the two dialing and radio tuning tasks to the proposed benchmark destination entry task are presented in Table 49. Group means (± standard error) are presented in Figure 3.

Table 49. Car Following Delay Planned Comparison Results

Comparison	df	F Value	PR > F
Destination Entry vs. Dialing 10-Digit	1,207	4.78	0.0702
Destination Entry vs. Dialing 7-Digit	1,207	4.5	0.0702
Destination Entry vs. Radio Tuning	1,207	0.08	0.7734

None of the three tests demonstrated significantly less performance degradation than that associated with destination entry. However, it is noteworthy that two of the three comparisons exhibited differences that were marginally significant and would likely have been significant had the tests been conducted in isolation, thus eliminating the need for multiple comparison adjustments.

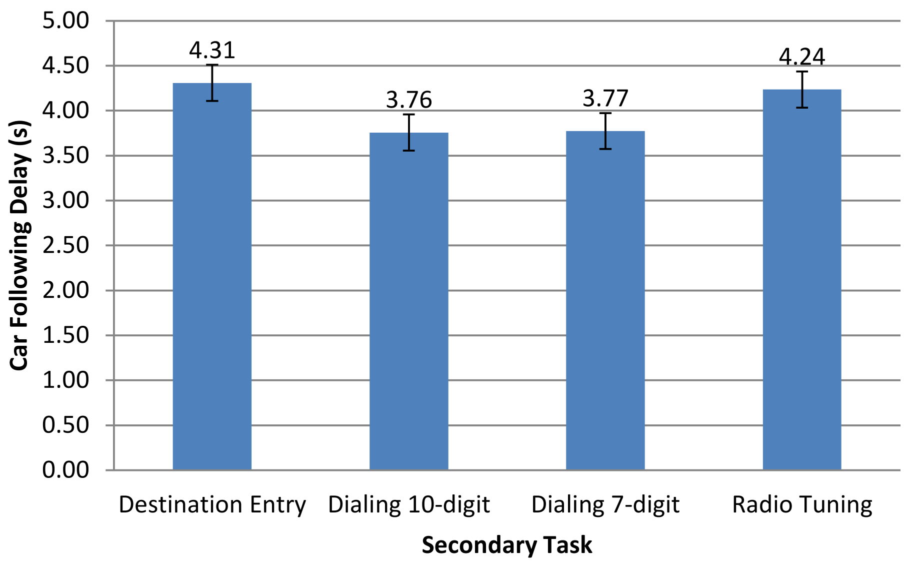

Figure 3. Mean Car Following Delay by Secondary Task (N = 72)

4.2.2.3 Detection Task Response Time

Statistical test results for target detection mean response time are presented in Table 50. Group means (± standard error) are presented in Figure 4.

Table 50. Target Detection Response Time Planned Comparison Results

Comparison	df	F Value	PR > F
Destination Entry vs. Dialing 10-Digit	1,207	7.45	0.0207
Destination Entry vs. Dialing 7-Digit	1,207	1.61	0.2065
Destination Entry vs. Radio Tuning	1,207	2.71	0.2022

Based on these tests, 10-digit dialing was the only of the three tasks to have been associated with significantly faster target detection response times than destination entry.

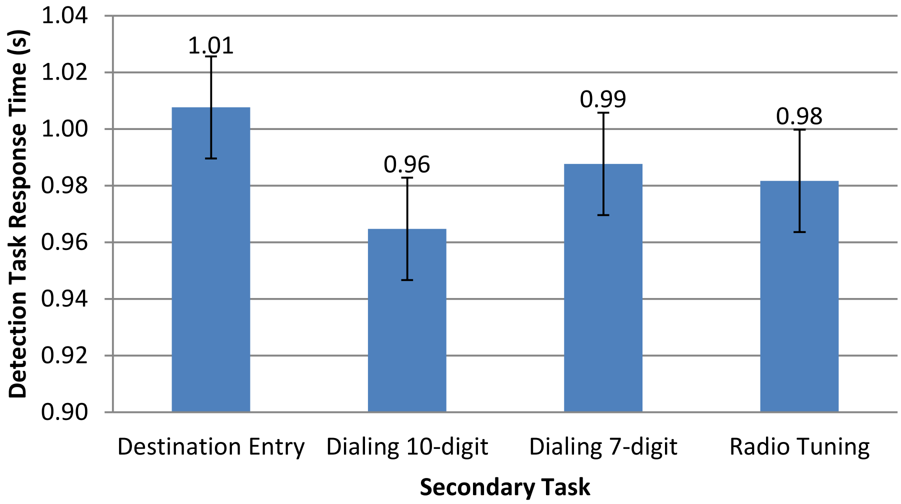

Figure 4. Mean Target Detection Response Time by Secondary Task (N = 72)

4.2.2.4 Detection Task Proportion Correctly Detected

Statistical test results for proportion of targets detected are presented in Table 51. Group means (± standard error) are presented in Figure 5.

Table 51. Proportion of Targets Correctly Detected Planned Comparison Results

Comparison	df	F Value	PR > F
Destination Entry vs. Dialing 10-Digit	1,207	25.04	<.0001
Destination Entry vs. Dialing 7-Digit	1,207	11.96	0.0007
Destination Entry vs. Radio Tuning	1,207	12.54	0.0007

For this metric, all three tasks were associated with better target detection accuracy than the destination entry task.

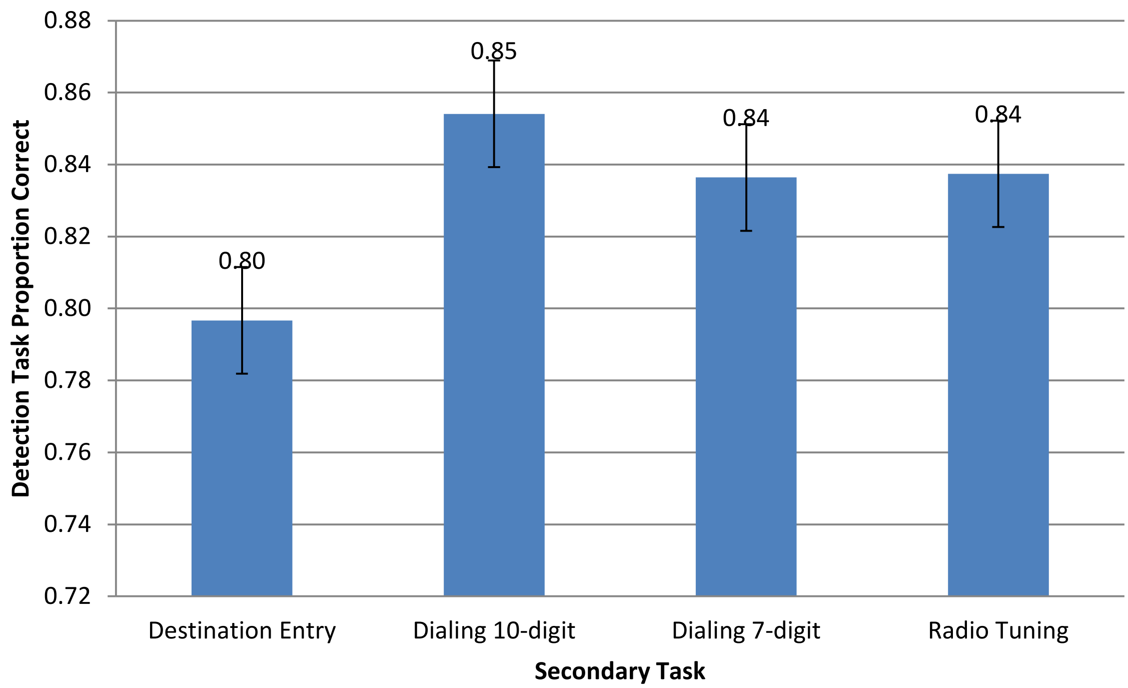

Figure 5. Mean Proportion of Targets Detected by Secondary Task (N = 72)

4.2.3 Age Effects

Statistical test results of age effect for the 4 DFD metrics are presented in Table 52. Group means (± standard error) are presented in Figure 6 through Figure 9.

Table 52. Age Effects on DFD Performance Metrics

	Age Main Effect			**Age X Secondary Task Interaction Effect**		
Metric	**df**	**F**	**Pr > F**	**df**	**F**	**Pr > F**
SDLP	2,69	12.21	< .0001	6,207	0.38	0.89
Car Following Delay	2,69	13.49	< .0001	6,207	1.74	0.11
Detection Task Response Time	2,69	15.90	< .0001	6,207	1.49	0.18
Detection Task Proportion Correct	2,69	11.66	< .0001	6,207	2.60	0.02

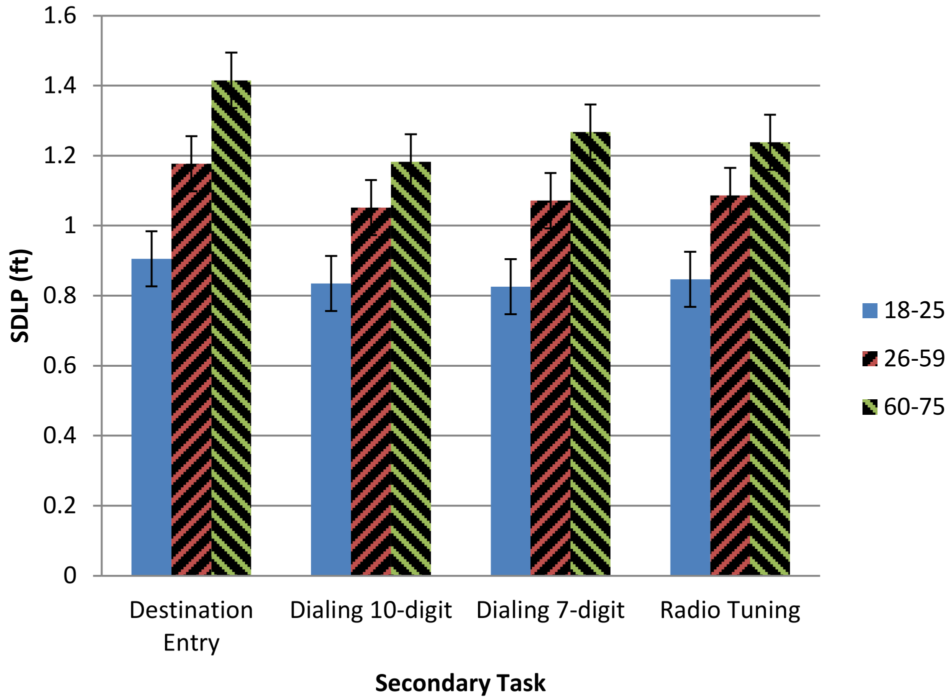

Figure 6. Mean SDLP by Age and Secondary Task (N = 72)

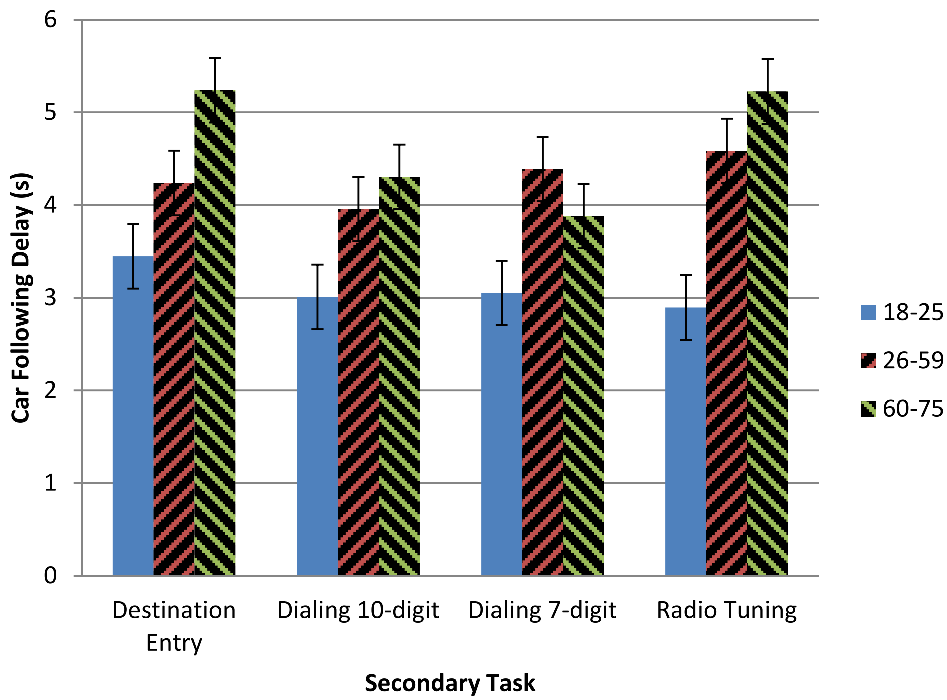

Figure 7. Mean Car Following Delay by Age and Secondary Task (N = 72)

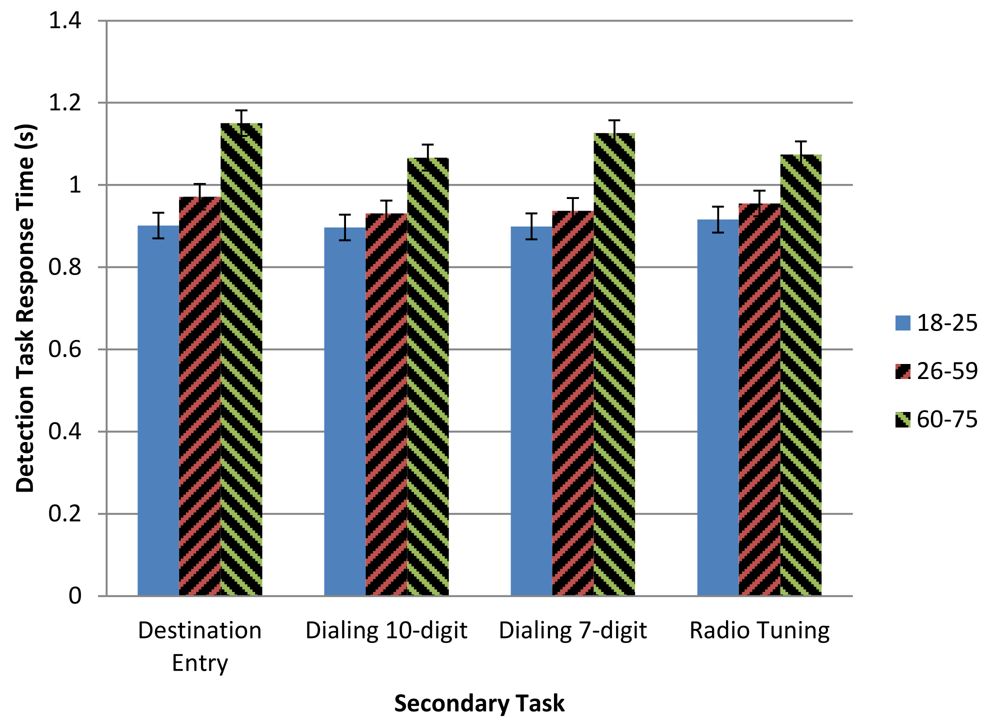

Figure 8. Mean Detection Task Response Time by Age and Secondary Task (N = 72)

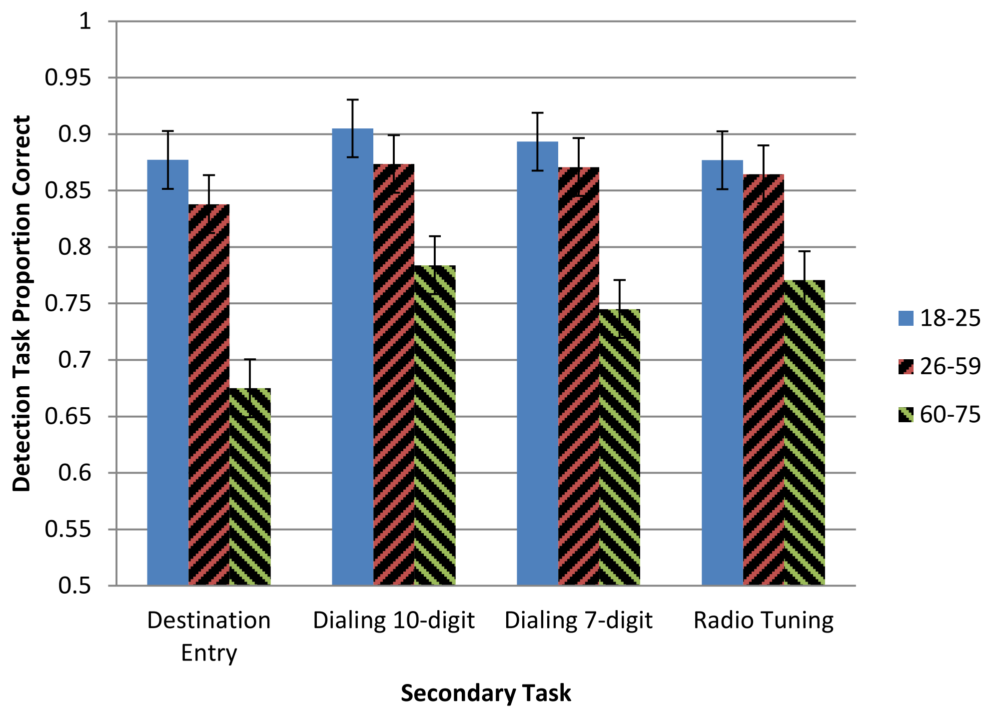

Figure 9. Mean Detection Task Proportion Correct by Age and Secondary Task (N = 72)

Age main effects were statistically significant for all metrics, reflecting increasing levels of performance degradation with increasing age. Results presented previously, particularly for the glance-based metrics, suggested that the younger and middle-aged groups were similar in performance, both being different from the older group. Examination of the means in the figures suggests a less consistent pattern for these metrics. For example, the differences among age groups for SDLP exist among all three groups. A similar pattern is apparent for car following delay. In contrast, differences for detection task response time suggest that the older group was generally slower than the younger and middle-aged groups. Differences between the older and other groups were also more pronounced for the proportion of targets detected. Among the four metrics, only the metric for proportion of targets detected had a significant Age x Secondary Task interaction effect, due most likely to the apparently much lower value for destination entry for older drivers. Also noteworthy, is the observation that the differences among the three age groups appear to be larger and more robust than differences among the various secondary task conditions. Implications of this pattern for the sensitivity of the test are discussed in the following section.

5.0 Discussion

5.1 Summary of Results

Proposed test criteria were assessed relative to the expectations that radio tuning would be an acceptable task and destination entry would be unacceptable. Expectations for the two dialing tasks were less clear; for time-based criteria based on radio tuning values, 10-digit dialing was considered unlikely to be acceptable, while 7-digit dialing was more likely than 10-digit dialing to be acceptable due to its shorter duration. The full sample (N = 72) results were intended to have sufficient statistical power to represent the "ground truth" with respect to the performance of the test criteria. Table 53 presents a summary of the main results based on the proposed test criteria and data from the full sample.

Table 53. Summary of Test Outcomes for Metrics using the Full Sample (N = 72)

Metrics	Destination Entry	Radio Tuning	7-Digit Dialing	10-Digit Dialing
TSOT < 9 seconds	Does Not Meet Criteria	Does Not Meet Criteria	Does Not Meet Criteria	Does Not Meet Criteria
* TEORT < 12 seconds	Does Not Meet Criteria	Meets Criteria	Meets Criteria	Does Not Meet Criteria
PLG < 15% glances longer than 2 seconds	Meets Criteria	Meets Criteria	Meets Criteria	Meets Criteria
MGD < 2 seconds	Meets Criteria	Meets Criteria	Meets Criteria	Meets Criteria
Delay < 4.6 seconds	Meets Criteria	Meets Criteria	Meets Criteria	Meets Criteria
* SDLP < 1 foot	Does Not Meet Criteria	Meets Criteria	Meets Criteria	Meets Criteria
DT Response Time < 1.0 second	Meets Criteria	Meets Criteria	Meets Criteria	Meets Criteria
* DT Accuracy > 0.8	Does Not Meet Criteria	Meets Criteria	Meets Criteria	Meets Criteria
Delay < Dest. Entry	NA	Does Not Meet Criteria	Does Not Meet Criteria	Does Not Meet Criteria
* SDLP < Dest. Entry	NA	Meets Criteria	Meets Criteria	Meets Criteria
DT Response Time < Dest. Entry	NA	Does Not Meet Criteria	Does Not Meet Criteria	Does Not Meet Criteria
* DT Accuracy > Dest. Entry	NA	Meets Criteria	Meets Criteria	Meets Criteria

* Test outcome results consistent with predictions.

Results presented in the first section of the table pertain to the visual performance metrics associated with the occlusion protocol and the glance-based results from the DFD driving simulator protocol. Of these metrics, only the TEORT metric results were consistent with expectations. The second section of the table presents results for the DFD driving performance metrics with fixed acceptance criteria. Among these metrics, SDLP and DT Accuracy provided results consistent with expectations. The third section of the table presents results for DFD driving performance metrics using the destination entry as a benchmark unacceptable task. For this set of tests, meeting the criterion required performance significantly better than the benchmark. Because destination entry was used as a benchmark task, only the defined expectations for radio tuning could be evaluated. The expectation was that radio tuning would be considered acceptable. For these metrics both SDLP and DT Accuracy found radio tuning to be significantly better than destination entry, and thus acceptable.

The full sample was constructed based on the proposed Guideline age group definitions that were under consideration at the time the experiment was designed. As released, the proposed guidelines included slightly different age group specifications for test samples. A re-analysis of data based on these newer age grouping specifications is presented later in Section 5.4.5.

The full sample was constructed to provide three subsets of 24 participants, such that each subset had the same age/gender balance required by the aforementioned earlier version of the draft guidelines. The construction of these subsets allowed analyses to be conducted to address two questions. The first question was whether the smaller samples provided consistent results across the three samples. The second question was whether the smaller sample's test results were consistent with those obtained using the full sample.

The following tables (Table 54 and Table 55) present the frequencies of participants with outcomes meeting the proposed NHTSA Guidelines acceptance criteria in each of the 3 groups of 24 participants. Shaded cells represent outcomes consistent with test performance meeting the acceptance criteria. For each metric, the three columns represent the outcomes for the three smaller samples. Recall that 21 is the minimum frequency necessary for acceptance.

Table 54. Frequency of Participants with Outcomes Meeting the Acceptance Criteria (Visual Metrics) by Test Group (N = 24)

	Number of Participants With:											
	TSOT < 9.0 s			**Mean TEORT < 12 s**			**(PLG) < 15% of Glances > 2 s**			**Mean Glance Duration < 2 s**		
Secondary Task	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**
Destination entry	0	0	0	0	0	0	23	22	22	24	24	24
Dialing 10-digit	8	7	6	20	18	19	23	22	19	24	24	24
Dialing 7-digit	17	15	19	24	22	23	23	23	22	24	24	24
Radio tuning	18	18	20	22	23	24	23	24	24	24	24	24

Three of the metrics provided outcomes that were consistent across the three smaller samples. These included the 9.0-second TSOT criterion, the 12-second TEORT criterion, and the Mean Glance Duration. Only the Proportion Long Glances had one small sample outcome (10-digit dialing) that was not consistent across the three smaller samples.

A similar presentation of acceptance criteria outcomes for the DFD metrics is presented in Table 55. Individual outcome frequencies are not appropriate here as the tests were based on statistical comparisons using data from the entire sample, rather than individually for each participant as required for the visual metrics.

Table 55. Secondary Tasks with Outcomes Meeting the Acceptance Criteria (DFD Performance Metrics) by Test Group (N = 24)

	Test Outcome: Meets Acceptance Criteria = M, Does Not Meet Acceptance Criteria = X											
	CF Delay < = 4.6 s			**SDLP < = 1 ft**			**DTRT < = 1 s**			**DT P Correct > .80**		
Secondary Task	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**
Destination entry	M	M	M	M	M	X	M	M	M	X	X	X
Dialing 10-digit	M	M	M	M	M	M	M	M	M	M	M	M
Dialing 7-digit	M	M	M	M	M	X	M	M	M	M	M	X
Radio tuning	M	M	M	M	M	X	M	M	M	M	M	X

The DFD summary results indicate some amount of variation among the 3 samples for SDLP and Detection Task Proportion Correct. The latter metric criterion (DT P Correct > .80) provides test outcomes more aligned with expectations for Groups 1 and 2, in which the radio tuning task provided acceptable outcomes while the destination entry task did not.

With respect to the question of whether the small-sample results were consistent with the large-sample results, the occlusion and glance-based metrics provided more consistency than the DFD driving performance results. Specifically, among the former, three of four metrics had small-sample outcomes entirely consistent with the large-sample outcomes. Only the proportion of long glances exhibited variation and that was for a single comparison. The DFD small-sample results were slightly more inconsistent with the large-sample results. Two DFD metrics (car following delay and detection task response time) had consistent outcomes among the smaller and larger samples. The other two metrics (SDLP and detection task accuracy) exhibited some variability. For SDLP, all small sample findings were at odds with the larger sample results. Group 3 participants had different outcomes from those in Groups 1 and 2 and it is clear that the large-sample result for destination entry (unacceptable) reflected the Group 3 result (unacceptable) but not those for Groups 1 and 2 (both acceptable). This pattern is significant in that it suggests that a consensus result from three small samples (2 of 3 groups found destination entry acceptable) may not always agree with the large sample outcome (destination entry not acceptable).

The level of agreement among the small sample outcomes may be misleading in the present context because for some metrics the large-sample outcome was not in accordance with expectations. This was true for three of the visual metrics (TSOT, proportion of long glances and mean glance duration), none of which differentiated among destination entry (unacceptable) and radio tuning (acceptable). This was also true for car following delay and detection task response time. It would appear most important to identify criterion values for metrics that provide the sensitivity necessary to differentiate between these two tasks before making conclusions about the consistency among the smaller-sample results.

5.2 Effects of Sample Size

In the context of the proposed guidelines testing in which acceptance criteria are specified (e.g., 21/24 = 0.875) and comparisons are made on an individual absolute basis, the main effect of using larger sample sizes is to reduce the size of the confidence interval (CI) around the observed proportions of outcomes meeting the acceptance criteria. This effect is shown in Table 56.

Table 56. Effects of Sample Size on the 95% Confidence Interval Associated with a Single Acceptance Criterion

Sample Size N	Sample Acceptance Criterion	Proportion	Lower Bound 95% CI	Upper Bound 95% CI
24	21	.875	.69	.96
48	42	.875	.75	.94
72	63	.875	.78	.93
96	84	.875	.79	.93

For purposes of comparison, the observed proportion of outcomes meeting the acceptance criterion is assumed to be 0.875 for all sample sizes (21/24, 42/48, etc.). As shown in Table 56, when the sample size is 24 there is a 95% chance that the true expected rate of compliance among real-world drivers with the same general characteristics as those who comprise the test sample, will range between 0.69 and 0.96. The size of the confidence interval is 0.27 (.96 - .69). This means that at least 69 percent of real-world drivers will be able to perform in compliance with the established acceptance criterion. When the sample size is 72, one can conclude with 95% confidence that the true value in the population will be between 0.78 and 0.93. The size of the CI for this sample size is 0.15 (.93 - .78), slightly more than half the size of the CI associated with N = 24. The increased sample size allows the conclusion that there is a 95% chance that real-world compliance will attain a minimum of 78%, versus 69% for the smaller sample.

5.3 Implications of Using 85% Sample Criterion versus 85th Percentile

Using the 85th percentile for design purposes is intended to ensure that 85 percent of potential users have no difficulties performing tasks with a given system. However, to ensure that this will be true, it is not sufficient to apply this criterion directly to a test sample, particularly when sample sizes are small. Setting the sample acceptance criterion at 85% (thus allowing 15% of the sample to experience a problem consistent with not meeting test acceptance criterion) only allows us to conclude with 95% confidence that the proportion of real-world drivers that will experience difficulties lies between the lower and upper CI boundaries surrounding the sample acceptance criterion. Therefore, to ensure (with 95% confidence) that 85 percent of the population of drivers will not have problems with a device, it is necessary to set the test sample acceptance criterion high enough to ensure that the lower CI boundary is not less than 0.85.

Statisticians recommend use of the Score CI with relatively small sample sizes (N < 80-100) and with expected probabilities near to the ends of the distribution, such as those used in the current testing (0.85 and 0.15) (Agresti & Coull, 1998; Devore, 2008). According to these authors, the Score method is more accurate and performs better than the traditional Wald method over the entire range of parameter values and sample sizes. Equations for computing Score CIs are given below:

$$\text{lower confidence limit} = \frac{\hat{p} + \frac{z_{\alpha/2}^2}{2n} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n} + \frac{z_{\alpha/2}^2}{4n^2}}}{1 + (z_{\alpha/2}^2)/n}$$

and

$$\text{upper confidence limit} = \frac{\hat{p} + \frac{z_{\alpha/2}^2}{2n} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n} + \frac{z_{\alpha/2}^2}{4n^2}}}{1 + (z_{\alpha/2}^2)/n}$$

where $\hat{p}$ is the probability of success and $\hat{q}$ = 1-$\hat{p}$, and $z_{\alpha/2}$ is the 97.5th standard normal percentile, which corresponds to a 95th percent CI.

The following tables present the Score CI boundary values for different acceptance criterion values and sample sizes for metrics that involve comparison of individual values with criterion values. Table 57 presents values for the proposed criterion level of 0.875 (21/24). Table 58 presents the same information for a criterion value of 0.917 (22/24). Table 59 presents values for the criterion value of 0.958 (23/24).

Table 57. Sample Confidence Interval Boundaries (21/24 = 0.875)

Sample Size N	Sample Acceptance Criterion	Probability	Lower Bound 95% CI	Upper Bound 95% CI
24	21	.875	.69	.96
48	42	.875	.75	.94
72	63	.875	.78	.93
96	84	.875	.79	.93

Table 58. Sample Confidence Interval Boundaries (22/24 = 0.917)

Sample Size N	Sample Acceptance Criterion	Probability	Lower Bound 95% CI	Upper Bound 95% CI
24	22	.917	.74	.98
48	44	.917	.80	.97
72	66	.917	.83	.96
96	88	.917	.84	.96

Table 59. Sample Confidence Interval Boundaries (23/24 = 0.958)

Sample Size N	Sample Acceptance Criterion	Probability	Lower Bound 95% CI	Upper Bound 95% CI
24	23	.958	.80	.99
48	46	.958	.86	.99
72	69	.958	.88	.99
96	92	.958	.90	.98

A summary of the lower CI boundary values for different combinations of acceptance criterion and sample size is presented in Table 60.

Table 60. Lower Confidence Interval Boundary by Sample Size and Acceptance Criteria Outcome

Sample Size N	.875 (21/24)	.917 (22/24)	.958 (23/24)
24	.69	.74	.80
48	.75	.80	.86
72	.78	.83	.88
96	.79	.84	.90

These results indicate that the combination of a minimum sample size of 48 and a 0.958 acceptance criterion (46/48) is required to ensure that real-world compliance reflects the test outcome for at least 85 percent of the population represented by the sample.

5.4 Test Procedure Considerations

The development and use of guidelines for assessing the distraction potential of secondary tasks requires procedures that differ considerably from those involved in empirical studies designed to test specific hypotheses. The present implementation and evaluation of proposed guidelines, including detailed procedures and specified criteria, raised a number of conceptual and procedural issues that merit additional consideration. Among these are the selection and disposition of outliers, sample size and composition, metric construction, test construction and the interpretation of test results. Each will be discussed briefly.

5.4.1 Treatment of Outliers

Identifying outliers is inherently ambiguous. Defining an outlier as a data point that is outside of the overall distribution does not provide enough guidance to ensure consistent treatment of potentially spurious values. In contrast, a requirement that outliers be defined as being more than 3 standard deviation values away from the mean would be less ambiguous, but may raise questions about what to do if the data are not normally distributed. Moreover, if the challenge is to identify outliers among groups of 5 data points, as in the occlusion protocol, the information provided by such a small number of data points is generally not sufficient to accurately define an underlying distribution or standard deviation value. The use of alternate metrics, including the median and broadened median, eliminates ambiguity concerning the definition of outliers. In samples of 5 trials, as specified for occlusion, the median would use the middle value, while the broadened median would eliminate the high and low values and use the mean of the three central values of each sample. In addition to eliminating the need to identify outliers specifically, these approaches have been shown to provide consistently better estimates of central tendency than mean values for extremely small samples such as those used in the occlusion procedure (Rosenberger & Gasko, 1983).

5.4.2 Sample Age Requirements

Sample composition, particularly age, is another factor that could influence test outcome. To explore the potential effects of different sample constructions based on participant age, two additional analyses were done for selected metrics. The first analysis explored the effects of eliminating the older category in the present study. For this purpose, the 24 participants in the oldest age group (over the age of 59) were removed. The resulting sample size was 48 participants.

Table 61 presents the outcome frequencies for the TSOT and TEORT metrics for this group. Based on the proposed 21/24 criterion, a proportion of 0.875 is required for acceptance.

Table 61. TSOT and TEORT Acceptance Outcome Frequencies for Younger and Middle Age Groups (N = 48)

	TSOT Test Outcomes		TEORT Test Outcomes	
Secondary Task	**Freq.**	**Prop.**	**Freq.**	**Prop.**
Destination entry	0	0.00	0	0.00
Dialing 10-digit	19	0.40	42	0.88
Dialing 7-digit	38	0.79	48	1.00
Radio tuning	44	0.92	48	1.00

The results indicate that radio tuning would be acceptable for both TSOT and TEORT criterion values and the two dialing tasks would be acceptable with TEORT but not with TSOT criterion values. These outcomes differ from the outcomes obtained using the original sample.

The second analysis involved a subset of the present study that was created to reflect the most recent sample construction specifications in the proposed NHTSA Guidelines, namely 6 participants (3 of each gender) in each of the following age groups: 18-24, 25-39, 40-54, and 55+. The overall sample was suitable for construction of one sample with these new specifications. The test outcome frequencies for TSOT are presented by age group in Table 62.

Table 62. TSOT Acceptance Outcome Frequencies by Age Group for More Recent NHTSA Guidelines Specifications (N = 24)

Secondary Task	18-24	25-39	40-54	55+	Total	Prop.
Destination entry	0	0	0	0	0	0.00
Dialing 10-digit	1	2	1	1	5	0.24
Dialing 7-digit	3	3	5	3	14	0.58
Radio tuning	5	5	6	5	21	0.88

For this subset, based on the proposed 9-second TSOT criterion, radio tuning has 21 of 24 participants with acceptable outcomes, which corresponds to the minimum acceptable value. For this single sample, using data from the same participants that revealed unacceptable outcomes under the original age criteria, the radio tuning task has become acceptable. Of interest is the apparently large difference in positive outcomes between radio tuning and 7-digit dialing, which was smaller in the analysis conducted using the earlier age specifications.

6.0 Conclusions

1. The proposed 12-second criterion for total eyes-off-road time (TEORT) achieved results consistent with the expectation concerning radio tuning and destination entry.
2. The proposed 9-second total shutter open time (TSOT) criterion for occlusion trials did not provide test results consistent with the expectation that radio tuning would meet the acceptance criteria and destination entry would not meet the acceptance criteria.
3. The predicted 4/3 relation between TEORT and TSOT was not found. The observed relation is closer to 1/1; however among individual tasks, the correlations between the two metrics were weak, suggesting no systematic relation between the two metrics. The heterogeneity associated with the wide age range of participants is a possible reason for this result.
4. The proposed criterion that defines the acceptable proportion of long glances can better ensure that tasks are performed without long glances than the criterion based on mean glance duration. The overall mean glance duration was 0.97 seconds and none of the participant means in any condition was greater than 2.0 seconds. This, together with the finding that most participants had some proportion of glances longer than 2.0 seconds, supports the conclusion that a criterion directly related to long glance proportion will be more effective in limiting long glances than one based on the mean glance duration.
5. Strong and consistent differences were observed among age groups. Generally, younger participants (18-25) were more likely to provide outcomes meeting the acceptance criteria, while older participants (60-75) were more likely to provide outcomes not meeting the acceptance criteria.
6. The effects of several test procedural details, including the method of defining task boundaries, the selection of outliers, and the use of medians versus means as a measure of central tendency appear to be strong enough to influence test outcomes. Additional effort is warranted to assess the implications of these procedural details and recommend specifications.

7.0 References

Agresti, A. & Coull, B.R. (1998). Approximate is better than “exact” for interval estimation of binomial proportions. *The American Statistician*, Vol. 52, No. 2, 119-126.

Bischoff, D. (2007) Developing guidelines for managing driver workload and distraction associated with telematic devices, SAE Paper Number 07-0082

Devore, J.L (2008). *Probability and Statistics for Engineering and the Sciences* (7^{th} Ed.). Belmont CA: Thomson Higher Education.

Foley, J. P. (2009). Now you see it, now you don’t: Visual occlusion as a surrogate distraction measurement technique. In M. A. Regan, J. D. Lee, & K. L. Young. *Driver Distraction: Theory, Effects, and Mitigation*. Boca Raton: CRC Press.

Hashimoto, K & Atsumi, B. (2001). Study of Occlusion Technique for Making the Static Evaluation Method of Visual Distraction, Japanese Report on Occlusion Workshop, Nov. 15, 2001.

International Standards Organization, *ISO 16673 Road Vehicles – Ergonomic Aspects of Transport Information and Control Systems – Occlusion Method to Assess Visual Demand Due to the Use of In-Vehicle Systems*, Geneva, Switzerland, 2007.

National Highway Traffic Safety Administration (2000). Traffic Safety Facts 2000: Older Population. DOT HS 809 328, National Center for Statistics & Analysis, National Highway Traffic Safety Administration, Washington D.C.

National Highway Traffic Safety Administration (2012). Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices, Notice of Proposed Federal Guidelines. *Federal Register* 77:37 (24 February 2012) p. 11200.

Ranney, T.A., Baldwin, G.H.S., Parmer, E., Martin, J., and Mazzae, E.N. (2011). Distraction Effects of Manual Number and Text Entry While Driving (Rep. No. DOT HS 811 510). Washington, D.C.: U.S. Department of Transportation: National Highway Traffic Safety Administration.

Rosenberger, J.L. & Gasko, M. (1983). Comparing location estimators: Trimmed means, medians and trimean. In D. C. Hoaglin, F. Mosteller, & J.W. Tukey. *Understanding Robust and Exploratory Data Analysis*. New York: John Wiley & Sons.

Stevens, A., Bygrave, S., Brook-Carter, N., & Luke, T. (2004). Occlusion as a technique for measuring In-Vehicle Information System (IVIS) visual distraction: a research literature review (Rep. No. TRL609). Berkshire, United Kingdom: TRL Limited.

Appendix A: Setting Test Criteria Based on Glance Time Required for Radio Tuning

In the rationale for Principle 2.1A, the Alliance cites two older research studies to support their selection of parameter values. Rockwell's (1988) work presents a distribution of individual glance durations associated with manually tuning a radio. Since this distribution was constructed using multiple glances from each research participant, it does not formally satisfy the independence requirement for statistical inference. Nevertheless, this data has been cited repeatedly as the basis for concern with long glances. From this data, the mean glance duration (MGD) during radio tuning is 1.44 (S = .5) seconds. The 85^{th} percentile glance duration value is approximately 2.0 seconds. Thus when considered in the aggregate, approximately 15% of glances involved in radio tuning were longer than 2.0 seconds. No information is provided on the number of glances required for radio tuning.

The second research study cited in support of Principle 2.1A is an on-road study conducted by Dingus and colleagues in 1987. The results of this study indicate that radio tuning required on average 6.91 (S = 2.39) glances and that the mean glance duration for radio tuning was 1.10 seconds. The mean and standard deviation were used together to estimate the 85^{th} percentile number of glances required for radio tuning. The estimate was 9.4 glances, which the Alliance increased to 10 glances for convenience.

These estimates provide information that can be used to compute the expected total eyes-off-road time required for radio tuning. The Dingus data are more appropriate for this purpose because they provide both glance duration and frequency. Accordingly, based on the Dingus study, the total eyes-off-road time (TEORT) required for radio tuning can be estimated by multiplying the MGD by the mean number of glances required. Doing so provides an estimate of 7.6 seconds TEORT for radio tuning.

Parameter estimates from these two studies were combined by the Alliance to create a maximum acceptable TEORT value of 20 seconds. Specifically,

1. Rockwell's data indicates that $85^{\%}$ of glances to the radio are less than 2.0 seconds.
2. Dingus's data indicates that $85^{\%}$ of radio tuning trials required no more than 9.4 glances, which they round to 10 glances.

The Alliance proceeds to multiply these 85^{th} percentile estimates to obtain a criterion of 20 seconds of TEORT for radio tuning. The justification for and interpretation of this multiplication are less clear. Even if the two underlying distributions are normal, which is typically not true for these variables, the resulting distribution of the products will not be normal. The resulting value, a product of two 85^{th} percentile estimates does not correspond to an 85^{th} percentile value for the resulting hypothetical TEORT distribution.

One way to demonstrate the problem of interpreting the product of these two variables is to consider the probability of the joint occurrence of the two events, assuming independence. Specifically, these estimates represent the 85^{th} percentiles of their respective distributions. Accordingly, if

A is defined as P (Single Glance Duration > 2.0 s) = 0.15 and

B is defined as P (Number Glances > 10) = 0.15,

then the probability of their joint occurrence is represented as:

$P (A \cap B) = P (A) * P (B) = 0.15 * 0 .15 = 0.0225$

This value represents the probability that these two events will occur together. Thus, the probability of a driver needing 10 glances to tune the radio and the probability that the duration of a single glance will be at least 2.0 seconds is estimated to be 0.0225. But this joint probability statement only involves the likelihood of <u>one</u> glance being at least 2 seconds in duration. The Alliance computation is based on the assumption that 10 successive glances are of this duration. If one assumes mutual independence among the 10 successive glance durations, the probability that all 10 glances are at least 2 seconds in duration is represented as:

$P(G_1 > 2.0) \cdot P(G_2 > 2.0) \cdot P(G_3 > 2.0) \cdot P(G_4 > 2.0) \cdot P(G_5 > 2.0) \cdot P(G_6 > 2.0) \cdot P(G_7 > 2.0) \cdot P(G_8 > 2.0) \cdot P(G_9 > 2.0) \cdot P(G_{10} > 2.0) = .15^{10} \approx 0$

Mutual independence appears to hold in this situation because the duration of each glance does not depend on the duration of any other glance in the sequence; rather, they are all assumed to be 2.0 seconds in duration. Thus, the probability of all ten glances being at least 2 seconds in duration is essentially zero. When this estimated joint probability is multiplied by the probability that 10 glances are required for radio tuning, the resulting probability remains essentially zero.

If the product of the two Alliance 85^{th} percentile values were intended to represent an 85^{th} percentile TEORT, then one would expect the probability of their joint occurrence to be close to 0.15. The fact that the joint probability is essentially zero suggests a problem with the rationale underlying the Alliance construction. One implication of the zero joint probability value is that the Alliance 20 second criterion can be expected to include the entire distribution of TEORT values.

Next, we consider the less extreme possibility that only some of the glances are longer than 2 seconds in duration. Thus, if we accept the Alliance's 85^{th} percentile estimates, we can use the binomial probability theorem to estimate the likelihood of different proportions of glances being longer than 2 seconds in duration.

Specifically, in a series of n = 10 total glances, the probability of getting k glances of at least 2 seconds in duration is the binomial probability $\binom{n}{k} p^k (1-p)^{n-k},\ where\ \binom{n}{k} = \frac{n!}{k!(n-k)!}$.

Table 63 presents the binomial probabilities for a series of 10 glances based on the probability (p) of a single glance being longer than 2 seconds being equal to 0.15. The binomial probabilities are then combined with the probability that 10 glances are required for radio tuning, as assumed by the Alliance.

Table 63. Binomial Probabilities for Different Numbers of Long Glances in a Sequence of 10 Glances

N	k	p	Binomial Probability	Probability 10 Glances Required	Joint Probability
10	0	0.15	0.20	0.15	0.03
10	1	0.15	0.35	0.15	0.05
10	2	0.15	0.28	0.15	0.04
10	3	0.15	0.13	0.15	0.02
10	4	0.15	0.04	0.15	0.01
10	5	0.15	0.01	0.15	0.00
10	6	0.15	0.00	0.15	0.00
10	7	0.15	0.00	0.15	0.00
10	8	0.15	0.00	0.15	0.00
10	9	0.15	0.00	0.15	0.00
10	10	0.15	0.00	0.15	0.00

The binomial probabilities add to 1.00 since all of the possible combinations of glance duration among sequences of 10 glances are included. Sequences involving more than 5 long-duration glances are predicted never to occur. The most likely outcome (binomial probability = 0.35) is that 1 glance in a series of 10 glances will be longer than 2 seconds. The proportion of trials that will have 0, 1, or 2 glances longer than 2 seconds in duration is 0.83 (0.20 + 0.35 + 0.28). This cumulative probability can be interpreted to suggest that if radio tuning requires exactly 10 glances and the probability of a single glance being longer than 2 seconds is 0.15, then on 83 % of trials no more than 2 of 10 glances will be longer in duration than 2 seconds.

Table 64 presents two estimates of TEORT, assuming a series of 10 glances, based on different frequencies of long (> 2.0 seconds) and shorter glances. The TEORT was estimated to be the sum of the durations of the long glances (2 seconds each) plus the durations of the remaining glances, assuming that the mean of the remaining glances is the same as the mean of the distribution of glance durations. The two TEORT estimates are based on the two mean glance duration (MGD) values cited by the Alliance (1.10 seconds from Dingus and 1.44 seconds from Rockwell). Thus, with 10 total glances and k long glances, the TEORT is estimated to be equal to the sum of the time required for k long glances plus 10-k glances of average duration. Specifically,

TEORT = k x 2.0 + (10-k) x MGD

The joint probabilities of these occurrences are taken from Table 63. Note that the sum of all of the joint probabilities shown in Table 64 is 0.15, which is the probability that radio tuning will require 10 glances.

Table 64. TEORT Estimates as a Function of Long Glance Frequency and MGD

Total Number of Glances	Number of Long Glances (k)	Joint Probability	Long Glance Duration (s)	Remaining Glance Duration (MGT = 1.10 s)	Remaining Glance Duration (MGT = 1.44 s)	TEORT (MGT = 1.10 s)	TEORT (MGT = 1.44 s)
10	0	0.03	0.00	11.00	14.40	11.00	14.40
10	1	0.05	2.00	9.90	12.96	11.90	14.96
10	2	0.04	4.00	8.80	11.52	12.80	15.52
10	3	0.02	6.00	7.70	10.08	13.70	16.08
10	4	0.01	8.00	6.60	8.64	14.60	16.64
10	5	0.00	10.00	5.50	7.20	15.50	17.20
10	6	0.00	12.00	4.40	5.76	16.40	17.76
10	7	0.00	14.00	3.30	4.32	17.30	18.32
10	8	0.00	16.00	2.20	2.88	18.20	18.88
10	9	0.00	18.00	1.10	1.44	19.10	19.44
10	10	0.00	20.00	0.00	0.00	20.00	20.00

Depending on which MGT estimate is used, the results of this analysis suggest that the likelihood of radio tuning requiring more than 4 long glances and a resulting TEORT greater than 14.60 to 16.64 seconds is zero. This suggests that the Alliance 20 second criterion is likely to be approximately 5 seconds beyond the 100th percentile of the distribution of TEORT values based on their assumed values for MGD and number of glances required for radio tuning.

Conclusions:

1. Setting a TEORT criterion by multiplying two 85th percentile estimates for glance duration and number of glances required is inappropriate; the resultant value can be expected to include the entire distribution of TEORT. Using the 85th percentile of a TEORT distribution appears most appropriate since the criterion is set directly from a distribution, rather than indirectly, as was done by the Alliance.

2. Adopting a criterion of no glances longer than 2.0 seconds is consistent with the 85th percentile of the distribution of proportion of long glances. Allowing 14% of glances to be longer than 2.0 seconds represents the 90th percentile of that distribution.

3. In the context of the Alliance criteria, which emphasize the total amount of glance time required to perform a task and the proportion of long glances, the need for a separate criterion based on mean glance duration is unclear.

Glance duration distributions can be expected to differ among different test venues. Different primary task demands may influence the distribution of glance durations. The intended test setting should be taken into consideration when setting criterion values.

Appendix B: Recruitment Materials

Advertisement 1:

Receive $42 per hour, plus mileage allowance, for approximately 3 hours of participation

We are seeking participants for a study of driving performance

The study will be conducted by:

Transportation Research Center Inc. for the National Highway Traffic Safety Administration (NHTSA) of the U.S.D.O.T.
At the proving ground in East Liberty, Ohio
Morning, Afternoon or Evening Sessions
MUST BE:

Licensed driver 18-75 years old
Good general health
7,000+ miles driven per year
Cell phone user while driving

PLEASE REPLY ONLINE: http://www.trcpg.com/researchstudy.asp
OR EMAIL: studyinfo@dot.gov
OR CALL: 1-800-262-8309 ext [ext] or 937-666-3[direct line #]

Project Summary (to appear on TRC website):

The Transportation Research Center Inc. is conducting a research study for the United States Department of Transportation's National Highway Traffic Safety Administration (NHTSA). The study will evaluate different tools that researchers use to measure distraction caused by in-vehicle technologies and portable devices including cell phones. Participants will drive in a driving simulator and perform in-vehicle tasks like using an in-vehicle navigation system or cell phone. If selected, you will be required to come to a laboratory facility located on the Transportation Research Center Proving Grounds in East Liberty, Ohio. Participation involves one morning, afternoon, or evening session of approximately 3 hours. Participants will receive $42 per hour for participating, as well as a monetary travel allowance for mileage to and from the test site for participation.

Advertisement 2:

PARTICIPANTS NEEDED FOR RESEARCH STUDY

We are seeking participants for a study of driving performance.

Receive $42 per hour plus mileage allowance
for approximately 3 hours of participation.

The study will be conducted by the
Transportation Research Center Inc. (TRC) for the National Highway Traffic Safety Administration
(NHTSA) of the United States Department of Transportation.

LOCATION: TRC Proving Ground, East Liberty, Ohio
TIMES: Morning, Afternoon, or Evening Sessions available

REQUIREMENTS:

Licensed driver 18-75 years old
Good general health
7,000+ miles driven per year
Cell phone user while driving

PLEASE REPLY ONLINE: http://www.trcpg.com/researchstudy.asp
OR EMAIL: studyinfo@dot.gov
OR CALL: 1-800-262-8309 ext. xxx or 937-666-3xxx

Participant Screening Questions:

Introduction	Thanks for expressing interest in participating in our research study!
Research Study Purpose	The study is being conducted by Transportation Research Center Inc., for the United States Department of Transportation's National Highway Traffic Safety Administration (NHTSA), to evaluate different tools that researchers use to measure distraction caused by in-vehicle technologies.
Purpose of Screening	During this call, I will describe the study and gather information that can be used by the principal investigator to determine if you qualify for participation. This call will take about 10 minutes. Is now a good time?
Participation Commitment	Participants will drive a driving simulator and perform in-vehicle tasks like using an in-vehicle navigation system or dialing a cell phone. Participation involves one morning or afternoon session of approximately 3 hours. If selected, you will be required to come to a laboratory facility located on the Transportation Research Center Proving Grounds in East Liberty, Ohio.
Participation Compensation	If selected, you may receive up to approximately $126 for participating in the study ($42/hour if asked). You will also receive a travel allowance for mileage to and from your home to our facility.
Information Being Requested & Confidentiality	I would now like to ask you a series of questions to determine your eligibility. Questions will cover: (1) personal information, (2) driving experience, (3) wireless phone usage, and (4) medical history. Note that we (NHTSA and TRC Inc.) will not release any personal identifying information that you provide during this call. The information gathered will be kept confidential, and stored in a password protected database on a local computer. Responses to health related questions will not be retained, they are merely being asked to determine your eligibility for participation. If you do not qualify for the study, the information collected will be destroyed unless you give us permission to retain the information for consideration in a future study. You do not have to answer any question that you do not want to answer and you may end this phone call at any time. At this time, are you willing to proceed with the questions? (If YES, then proceed. If NO, then ask if person would like us to keep his/her contact information in our database and contact him/her for consideration for participation in future studies? If NO, then make note to delete information and inform caller that his/her information will be deleted, or if using paper survey – the survey will be shredded.)
NOTE:(Office Use Only)	Exclusion Criteria are on Subject Info Sheet.

Questions for Subject Recruitment Phone Interview
Subject Number (selection by principal investigator)
Caller Number
Date Interviewed
Interviewers initials
Date Scheduled
NAME (first M.I. last)
PHONE – DAY
PHONE – ALTERNATE
GENDER (M / F)

BIRTHDATE (mm/dd/yyyy)
AGE (office note: must be 18-75 years old to participate)
How did you hear about our research study? (Craigslist Ad, Newspaper Ad, Friend, Relative, Other? If newspaper ad, which newspaper? If Other, specify
HEIGHT (needed to pre-adjust in-vehicle equipment before participant arrival)
OCCUPATION
Does your job involve any type of driving? (If unemployed, did your most recent job involve any type of driving?)
Do you have a valid US driver's license? (Y / N), Are there any restrictions on that license? Are you able to drive without the use of assistive devices?
Number of years of driving experience (office note: need at least **2** years)?
How many miles do you drive per year (office note: **> 7,000**)?
What kind of vehicle do you normally drive (year, make, model)? (Office note: We cannot use participants experienced with the study vehicle or models that have identical in-vehicle interface, i.e.: same radio, navigation and Bluetooth phone capabilities.)
End call here if: driver's license not valid, there are license restrictions other than corrective eyewear, driving experience is less than 2 yrs, the # of miles driven per year is less than 7,000.
Do you wear prescription eye glasses or contacts while driving?
(IF YES to eye glasses) Do you own contact lenses that you can wear if you are selected to participate in the study?
How comfortable (on a scale of zero to ten, with zero being least comfortable) are you at multi-tasking while driving (e.g., eating, drinking, changing radio stations, talking on a cell phone, talking with passengers)?
(IF YES to phone use) What type of cell phone do you use? (indicate make/model, or none)
Do you use your cell phone while driving (Y / N)?
(IF YES to phone use while driving) HOW LONG HAVE YOU USED A CELL PHONE WHILE DRIVING?
(IF YES to phone use) On a scale from zero to one hundred, approximately what percentage of your normal driving time is spent using the cell phone?
Do you regularly communicate using text messages?
(IF YES to text messages) Do you at times send text messages while driving?
Do you use a navigation system, computer, or any other similar devices in your car? If yes, what are they specifically?
Have you ever used a navigation system to obtain route guidance directions while driving?
Are you familiar with the normal computer keyboard (QWERTY)?
ANY CRASHES IN LAST 5 YEARS?
(If YES to crash) Briefly, what was the cause?
ANY OTHER RECENT TRAFFIC VIOLATIONS?
(IF YES to traffic violations) WHAT TYPE(S) OF VIOLATIONS?
I am now going to ask you some confidential questions about your medical history and present condition. You can refuse to answer any question. If you choose to answer, please answer YES or NO to the following.
DO YOU HAVE A KNOWN ALLERGY TO LATEX?
Do you have any health problems that affect driving?
Do you suffer from any heart conditions such as: disturbance of the heart rhythm? Have you had a heart attack within the last 6 months? Have you had a pacemaker implant within the last 6 months? (If yes to any, please describe.)

Do you have high blood pressure that is not controlled by medicine?
Do you have a history of seizures or epilepsy?
Do you have any respiratory disorders such as: asthma, chronic bronchitis? (If yes, please describe.)
Are you susceptible to motion sickness?
Do you have any inner ear, dizziness, vertigo, hearing, or balance problems?
Do you have diabetes for which insulin is required??
Have you ever had brain damage as the result of a stroke, tumor, head injury, infection, or other injury? (If yes, please describe.)
Do you have migraine or tension headaches that impair your ability to drive? (If yes, how frequent?)
Do you have any sleep disorders such as apnea, narcolepsy or chronic fatigue? (If asked, Narcolepsy is a neurological disorder, which causes people to feel like sleeping or to fall asleep uncontrollably during the daytime.)
Are you taking any medications that may impact your driving ability?
Are you taking any prescription or over-the-counter medications that have the potential to cause drowsiness?
Do you currently have a mood disorder such as ADHD, depression, anxiety, or claustrophobia? (ADHD or Attention Deficit Hyperactivity Disorder is a mental disorder in which people are restless, extremely distractible, and have difficulty paying attention.)
That ends the health questions. Now we'll discuss your availability.
Availability
What are your normal work hours? What days can you participate in an approximately 3 hour experiment session?
What time of day would you prefer to participate in testing (Morning or Afternoon)?
Are you available on short notice to participate in our study? Could we call you on the same day to schedule if necessary?
Can we use your email address to help with scheduling? If yes, obtain EMAIL ADDRESS:
Can we contact you by text message to help with scheduling?
If we decide to use Twitter to announce our needs for this and for other studies, would you be interested in being added to our distribution list?
How long would you like to be considered for this study?
TO END CALL SAY:
OK, that's all of our questions. Thank you! We will contact you if you have been accepted for participation. We will provide additional information at that time. If you are eligible but not called for the current study, may we keep your contact information for possible participation in future studies?
CALL BACK FOR SCHEDULING APPOINTMENT
The Principal Investigator or his designated associate will determine which subjects are selected for participation.
Office Use Below
WHO CONTACTED THE SUBJECT? (personnel name)
APPOINTMENT CONFIRMATION CALL BACK
Hi this is _________ from TRC Inc. This is a call back to notify you that you have been selected to participate in our driving study discussed in earlier phone conversations with ___________. I have several additional questions and then I will schedule a test session.

While participating in the study, you will be required to wear 5 or 6 small stickers on your face to help our instrumentation determine where you are looking while driving. The stickers contain latex, which is why we asked about latex allergies. The stickers would be put on before you begin driving and cannot be removed or moved during the study until the experimenter informs you that you are finished driving. As a result you could possibly be wearing the stickers for 2.5 hours. The stickers are not painful to remove. Will this cause any problems for you?
Are you currently taking any over-the-counter cold or allergy medications? (Record "Yes" or "No") Are you taking any prescription drugs that may affect your driving (such as those that advise you not to work with heavy machinery or operate a vehicle)? (Record "Yes" or "No")
SCHEDULING PARTICIPANTS
I would like to schedule an appointment with you at this time. The first available openings are: ____________ (calendar of events needed w/ date and time frame of approximately 3 hours - try to utilize previous call input for choices, before calling). Do any of those dates and times work for you? (If YES, schedule. If NO, then offer next available set of times, perhaps by week, until scheduled. If no good dates, find a time when best for them and say we will see what we can do and call back later.)
(Appointment Confirmation) Ok. I have you scheduled for ________. Please try to arrive promptly.
Can I please get your street address so that we can calculate the amount of your mileage reimbursement and have the payment ready to give you at your participation visit? **STREET ADDRESS:** **ZIP CODE:**
Also, please be sure to bring your valid, U.S. driver's license to the appointment for identification purposes. Dress comfortably for driving and weather conditions and wear comfortable driving shoes. Do not bring another guest with you, unless prior arrangements have been made with us. Note that your personal wireless devices must be turned off while you are participating in this study. If your hair hangs in your face, you will be asked to pull it back out of the way. **Please refrain from drinking alcohol or taking non-prescription drugs for at least the 24 hours preceding the session.** Do you understand these requirements? (Record "Yes" or "No")
Note that: Cameras, firearms, and alcoholic beverages are not permitted at the data collection facility.
DIRECTIONS: We will send a map link to you using the email address that you provided.
Do you have any questions at this time?
If you have any questions before your scheduled date, please feel free to call me at 1-800-262-8309 ext xxx. If you need to contact us on the day of your scheduled appointment, please call xxx-xxx-xxxx. Or you can send a text message to this number XXX-XXX-XXXX.
If something comes up and you need to cancel or reschedule your appointment, please try to call at least 24 hours in advance. Otherwise, we look forward to seeing you on (date at time) ___________.

Appendix C: Information Summary and Confidential Information Form

PARTICIPANT INFORMATION SUMMARY AND CONFIDENTIAL INFORMATION FORM

STUDY TITLE:	Driver Eye Glance Behavior During Secondary Task Performance
STUDY INVESTIGATOR:	Thomas A. Ranney, Ph.D.
STUDY SITE:	Transportation Research Center Inc. 10820 State Route 347 East Liberty, OH 43319
TELEPHONE:	800-262-8309
SPONSOR:	National Highway Traffic Safety Administration

You are being asked to participate in a research study. Your participation in this research is strictly voluntary, meaning that you may or may not choose to take part. To decide whether or not you want to be part of this research, the risks and possible benefits of this study are described in this form so that you can make an informed decision. This process is known as informed consent. This consent form describes the purpose, procedures, possible benefits and risks of the study. This form also explains how your information will be used and who may see it. You are being asked to take part in this study because the study investigator feels that you meet the qualifications of the study.

The study investigator or study staff will answer any questions you may have about this form or about the study. Please read this document carefully and do not hesitate to ask anything about this information. This form may contain words that you do not understand. Please ask the study investigator or study staff to explain the words or information that you do not understand. After reading the consent form, if you would like to participate, you will be asked to sign this form. You will be offered a copy the form to take home and keep for your records.

PURPOSE

This research study is being conducted by the National Highway Traffic Safety Administration (NHTSA). The purpose of this study is to evaluate the different tools that researchers use to measure the level of distraction caused by "in-vehicle technologies" and portable devices such as cell phones. The latest in-vehicle technologies and some portable cell phones provide services such as internet access, navigation information (maps and driving directions), as well as the ability to send and receive e-mails and text messages. As new in-vehicle technologies are developed and marketed, there is a concern that these systems may interfere with driving. NHTSA is conducting this research study to determine the best way to collect data (information) on the use and impact of in-vehicle technologies while driving.

STUDY REQUIREMENTS

You are being asked to participate in this research study because:

- You are 18-75 years of age,
- You have a valid, unrestricted U.S. driver's license (except for restrictions concerning corrective eyeglasses and contact lenses),
- You have a minimum of two years driving experience,
- You drive at least 7,000 miles per year, and
- You are in good general health.

NUMBER OF STUDY SITES AND STUDY PARTICIPANTS

This study will take place at one research site (Transportation Research Center Inc.) and will include approximately 72 participants.

STUDY PROCEDURES

Before participating in this research study, you will be asked to read this Participant Informed Consent Form in its entirety. After all of your questions have been answered, you will be asked to sign this form to show that you voluntarily consent to participate in this research study.

Your participation in this research study will consist of one session lasting approximately 3 hours. A member of the study staff will give you detailed instructions and will accompany you at all times during your participation in this research study.

<u>In-Vehicle Tasks:</u>
During this session you will be asked to perform various in-vehicle tasks that involve actions similar to those required to operate in-vehicle systems. The in-vehicle tasks will consist of tasks using the radio, navigation system, and the phoning interface in the study vehicle.

You will be asked to perform the in-vehicle tasks while driving in a driving simulator as well as while wearing visual occlusion glasses that temporarily impede your vision to simulate a driver glancing at the roadway.

<u>Visual Occlusion:</u>
Visual occlusion is a technique used to simulate the visual demands of driving in a stationary setting. It requires that you wear a set of glasses, which have lenses that can be made to be either transparent or opaque. An electrical current can quickly change the glasses between these two states. When they are transparent, you will be able to see normally; however, when they are opaque, you will not be able to see through them. The glasses are connected to a computer, which controls when the lenses change between opaque and transparent. When the glasses are transparent, you will be able to see the in-vehicle task controls and displays. When the glasses are opaque, you will not be able to see the in-vehicle task controls and displays. The opaque condition is intended to simulate the time during which you would need to look at the roadway ahead in order to maintain vehicle control while driving.

<u>Simulated Driving:</u>
During your session you will be asked to drive a fixed-base simulator. A fixed-based simulator is a machine that imitates the conditions of driving in real life, but does not move. The simulator will be connected to a recent model-year passenger vehicle (sedan, minivan, or SUV). While driving the simulator, you will sit in the driver's seat of the study vehicle. You will control the simulator by moving the steering wheel and the gas and brake pedals of the study vehicle. The vehicle will have its engine turned off. The vehicle used with the driving simulator is equipped with sensors to collect information on your steering, braking and gas pedal usage. The sensors are located so that they will not affect your driving. The information collected by these sensors is recorded so that it can be analyzed at a later time. A large screen in front of the vehicle will display a computer-generated image of the virtual road on which you will be driving.

While operating the simulator, you will be asked to perform specific driving tasks. These tasks will involve activities such as following a car at a specified distance and detecting simple visual targets that appear on the computer-generated roadway image.

<u>Eye Movement Recording and Monitoring:</u>
Video cameras will be used to monitor your eye movements while operating the driving simulator and performing the in-vehicle tasks. The video cameras are located so that they will not affect your driving. The information collected using these video cameras is recorded so that it can be analyzed later.

There are certain requirements for accurately recording your eye movements while driving. These requirements are as follows:

- Your entire face must be clearly visible while driving. If your hair hangs in your face, you may be asked to use clips or an elastic band to keep it out of your face.
- If you require corrective lenses and have contact lenses, you will be asked to wear them rather than glasses.
- You will not be permitted to wear sunglasses while driving.
- To help the eye tracking system better identify and track your facial features, you will be required to wear several small stickers on your face. The stickers will be put on before you begin driving and cannot be

removed or moved until a member of the study staff informs you that you are finished driving. As a result you may be wearing the stickers for up to 2 hours.

<u>Summary of Study Procedures:</u>
The following procedures will take place at your session:

- After signing this consent form, you will be provided instructions and training on driving the simulator, occlusion glasses operation, and performing the in-vehicle tasks. You will also be given the opportunity to practice each of these before performing test trials.
- You will then complete 2 sets of trials, including approximately 10 simulator driving trials each lasting 3.5 minutes and approximately 30 occlusion trials each lasting between 30 seconds and 2 minutes.
- After completing both simulator driving and occlusion trials, the session will end and your participation in this research study will be complete.

NEW INFORMATION
We do not anticipate that any changes to procedures will take place during this study. However, any new information developed during the course of the research that may affect your willingness to participate will be provided to you.

RISKS of STUDY PARTICIPATION
Most people enjoy driving in the simulator and do not experience any discomfort. However, a small number of participants experience symptoms of discomfort associated with simulator disorientation. Previous studies with similar driving intensities and simulator setups have produced mild to moderate disorientation effects such as slight uneasiness, warmth, or eyestrain for a small number of participants. These effects typically last for only a short time, usually 10-15 minutes, after leaving the simulator. If you ask to stop driving as a result of discomfort, you will be allowed to stop at once. You will be asked to sit and rest before leaving, while consuming a beverage and a snack. There is no evidence that driving ability is hampered in any way; therefore, if you show minimal or no signs of discomfort, you should be able to drive home. If you experience anything other than slight effects, transportation will be arranged through other means. This outcome is considered unlikely since studies in similar devices have shown only mild effects in recent investigations and evidence shows that symptoms decrease rapidly after simulator exposure is complete.

You will be asked to wear several small latex stickers on your face while driving. These stickers may cause skin irritation in people with an allergy to latex. Allergic reaction may be mild (rash, hives) to severe (difficulty breathing, or a collapse of blood circulation and breathing systems). A severe allergic reaction, which is extremely unlikely, would require immediate medical treatment and could result in permanent disability or death.

There are no known physical or psychological risks associated with participation in this study beyond those described above.

BENEFITS of STUDY PARTICIPATION
This research study will provide data on driver behavior and in-vehicle task performance that will be used by researchers to provide a scientific basis for developing recommendations or standards for performing in-vehicle tasks while driving. Your participation in this study will provide data that may help develop these recommendations or standards.

You are not expected to receive direct benefit from your participation in this research study.

ALTERNATIVES
This study is for research purposes only. Your alternative is to not participate.

CONDITIONS OF PARTICPATION, WITHDRAWAL, AND TERMINATION
Participation in this research is voluntary. By agreeing to participate, you agree to operate the research vehicle in accordance with all instructions provided by the study staff. If you fail to follow instructions, or if you behave in a dangerous manner, you may be terminated from the study. You may withdraw your consent and discontinue participation in the study at any time without penalty.

COSTS TO YOU
Other that the time you contribute, there will be no costs to you.

COMPENSATION
You will receive $42.00 per hour for the time you spend at the data collection facility. You will receive mileage reimbursement for travel to and from the data collection site.

If you voluntarily withdraw or are terminated from this study, you will be paid for the number of hours that you participated in the study.

USE OF INFORMATION COLLECTED
In the course of this study, the following data will be collected:

- Engineering data (such as the information recorded by the study vehicle sensors)
- Video/audio data (such as the information recorded by the video cameras)

Information NHTSA may release:
The **engineering data** collected and recorded in this study will include performance scores based on the data. This data will be analyzed along with data gathered from other participants. NHTSA may publicly release this data in final reports or other publication or media for scientific, education, research or outreach purposes.

The **video/audio data** recorded in this study includes your video-recorded likeness and all in-vehicle audio (including your voice). The video/audio data may include information regarding your driving performance. Video and in-vehicle audio will be used to examine your driving performance and other task performance while driving. NHTSA may publicly release video image data (in continuous video or still formats) and associated audio data, either separately or in association with the appropriate engineering data for scientific, educational, research or outreach purposes.

Information NHTSA may not release:
Any release of **engineering data** or **video/audio data** shall not include release of your name. However, in the event of a court action, NHTSA may not be able to prevent release of your name or other personal identifying information. NHTSA will not release any information collected regarding your health and driving record.

QUESTIONS
Any questions you have about the study can be answered by Thomas Ranney, Ph.D., or the study staff by calling 1-800-262-8309.

If you have any questions regarding your rights as a research participant, or if you have questions, concerns, complaints about the research, would like information, or would like to offer input, you may contact: Rev. Paul E. Gamber, J.D., Chairman of Sterling Institutional Review Board, 6300 Powers Ferry Road, Suite 600-351, Atlanta, Georgia 30339 (mailing address) at telephone number 1-888-636-1062 (toll free).

INFORMED CONSENT
By signing the informed consent statement contained in this document, you agree that your participation is voluntary and that the terms of this agreement have been explained to you. Also, by signing the informed consent statement, you agree to operate the study vehicle in accordance with all instructions provided by the study staff. You may withdraw your consent and discontinue participation in the study at any time without penalty.

NHTSA will retain a signed copy of this Informed Consent form. A copy of this form will also be offered to you.

Informed Consent Statement

I certify that:

- I have a valid, U. S. driver's license.
- All personal and vehicle information, as well as information regarding my normal daily driving habits provided by me to NHTSA, and/or Transportation Research Center Inc. employees associated with this study during the pre-participation screening and the introductory briefing was true and accurate to the best of my knowledge.
- I have been informed about the study in which I am about to participate.
- I have been told how much time and compensation are involved.
- I have been told that the purpose of this study is to evaluate the tools that researchers use to measure driving and in-vehicle task performance.
- I agree to operate the research vehicle in accordance with all instructions provided to me by the study staff.

I have been told that:

- Part of the study will be conducted in a fixed-base driving simulator and that the risk of discomfort associated with simulator disorientation is minimal.
- For scientific, educational, research, or outreach purposes, video images of my driving, which will contain views of my face and accompanying audio data, may be used or disclosed by NHTSA, but my name and any health data or driving record information will not be used or disclosed by NHTSA.
- My participation is voluntary and I may refuse to participate or withdraw my consent and stop taking part at any time without penalty or loss of benefits to which I may be entitled.
- I have the right to ask questions at any time and that I may contact the study investigator, Thomas Ranney, Ph.D., or the study staff at (937) 666-4511 or 800-262-8309 for information about the study and my rights.

I have been given adequate time to read this informed consent form. I hereby consent to take part in this research study.

I, ____, voluntarily consent to participate.
(*Printed Name of Participant*)

____
Signature of Participant Date

INFORMATION DISCLOSURE

By signing the information disclosure statement contained in this document, you agree that the National Highway Traffic Safety Administration (NHTSA) and its authorized contractors and agents will have the right to use the NHTSA engineering data and the NHTSA video and audio data for scientific, educational, research, or outreach purposes, including dissemination or publication of your likeness in video or still photo format, but that neither NHTSA nor its authorized contractors or agents shall release your name; and you have been told that, in the event of court action, NHTSA may not be able to prevent release of your name or other personal identifying information. NHTSA will not release any information collected regarding your health and driving record, either by questionnaire or medical examination. Your permission to disclose this information will not expire on a specific date.

Information Disclosure Statement

I, ____, grant permission to
(Printed Name of Participant)

the National Highway Traffic Safety Administration (NHTSA) to use, publish, or otherwise disseminate NHTSA engineering data and NHTSA video image data, as defined in the Participant Informed Consent Form (including continuous video and still photo formats derived from the video recording), and associated with the appropriate engineering data for scientific, educational, research, or outreach purposes. I have been told that such use may involve widespread distribution to the public and may involve dissemination of my likeness in video or still photo formats, but will not result in release of my name or other identifying personal information by NHTSA or its authorized contractors or agents. I have been told that my permission to disclose this information will not expire on a specific date.

____

Signature of Participant Date

Appendix D: Secondary Task Instructions (Driving Simulator Protocol)

While driving in the simulator, you will be asked to perform a series of tasks, which we call "Secondary Tasks." We refer to driving (car following and target detection) as the "Primary Task" because safe control of the vehicle is more important than performing the secondary tasks. Performing secondary tasks can interfere with car following and target detection but it is important that you don't let primary task performance deteriorate too much while performing the secondary tasks.

Each drive will involve one secondary task. You will perform the specified secondary task repeatedly during the drive when the lead vehicle is present. The specific secondary task details will be presented before each drive. At this time I will explain the way in which the secondary task information is presented to you. Let's take radio tuning as an example. In this task you will be asked to tune the radio to different stations. We will tell you which stations to select; this information will be presented auditorily so that you do not have to look away from the driving task. It will also be displayed on the computer screen located to the right of the center console, in case you forget. We call this the Task Screen. The Task Screen is a "touch screen," which means that you will touch or press it when you complete each task.

When you are driving, the first radio frequency will be presented shortly after the lead vehicle appears. This will be followed by the instruction to "BEGIN." When you hear the word "BEGIN," you should work quickly and accurately to complete the secondary task. When you have entered the first radio frequency, you should say "DONE" out loud. Then, press the "DONE" button on the Task Screen. You will then be prompted to return the system to an initial condition such that the next radio tuning task can be performed using the same sequence. After returning the system to the desired initial condition, you will press the "NEXT" button on the Task Screen. This will cause the next radio frequency to be presented to you followed by the instruction to "BEGIN." At this point, you would select this new frequency on the radio, say "DONE" out loud, and then press the "DONE" button on the Task Screen again. You should continue tuning the radio in this way until the lead vehicle disappears at the end of the trial at which time you can stop working on the secondary task.

I want to say a few words about errors: Mistakes are inevitable. If you make a mistake while performing a secondary task, please try to correct the error before moving on. We will provide specific information about how to recover from errors for each secondary task. It is important that you try to complete each task if possible. It is also important that you work continuously on secondary tasks during the entire drive.

Any questions?

Radio Tuning (Driving Simulator Protocol)

In this task you will tune the radio to a designated frequency by using the tuning knob at the upper right corner of the radio/navigation module. This vehicle has buttons on the steering wheel for this purpose, but we want you to use the knob on the console at all times. During the drive, you will select several different radio frequencies, one at a time. You will be presented the band (AM or FM) and the frequency. The trial will require that you complete as many radio tuning instances as possible.

The first frequency will be presented shortly after the lead vehicle appears. Start the task when instructed to "Begin."

Press the "AUDIO" button at the bottom of the column of buttons to the left of the Prius video screen. The audio display will then appear on the video screen. Note that the Prius video screen is also a touch screen.

Select the frequency band by pressing the AM or FM button located to left of the video screen. (Note, we want you to use the buttons for this task, not the on-screen AM/FM tabs.) The current band is displayed in the upper left of the screen and current frequency in the upper right. If you select the wrong band, press the button for the appropriate band. (After about 20 seconds of inactivity, the display will revert to the MAP display or the 'Phone' screen, depending upon which was most recently used. If this occurs, press the "AUDIO" button again to return to the audio screen.)

Use the tuning knob, located to the upper right of the screen, to adjust the frequency. When you have reached the specified frequency, say "Done" out loud, and then press the "DONE" button on the Task Screen to complete the radio tuning task.

If you make an error, attempt to fix the band and frequency before moving on. If you make an error but have already said "DONE", continue on to the next step on the Task Screen.

When a radio tuning task is complete, there is a bit of work we want you to do before beginning the next radio tuning task. First, press the 'MAP' button to the right of the Prius video screen to take us out of radio mode, such that the next radio tuning task will also commence by pressing the 'AUDIO' button. Once you've pressed the 'MAP' button, then press the 'NEXT' button on the Task Screen such that both screens are ready for the next task. Once you press the 'NEXT' button, you should hear and see the next radio frequency to tune to.

Any questions before we practice this task?

Navigation System Destination Entry by Address (Driving Simulator Protocol)

In this task you will enter destinations into the navigation system. You will be presented the city, street name, and house number. Each trial will require that you complete as many destination entry tasks as possible.

The first destination will be presented shortly after the lead vehicle appears. Start the task when instructed to "Begin."

Four icons will be displayed in the middle of the screen. Note that the Prius video screen is also a touch screen. Press the icon labeled "**Address**". The system will display three options for destination entry.

We will always enter the city first. Press the "**City**" button. A keyboard will appear on the screen. Enter the city name on the on-screen keyboard until the system displays a list. Select the city from the list by pressing the bar on which the city name is displayed.

Two buttons can help you correct **errors**. If you make an error during keyboard entry, pressing the "Delete" button (a left-pointing arrow in the upper right portion of the on-screen keyboard) will erase the most recently entered letter, one at a time. If the system has generated a list which does not contain the specified city or street, pressing the "Back" button (a U-shaped arrow pointing left at the top right portion of the screen) will allow you to go back to the previous screen. This "Back" button is available on every screen.

Once you have selected a city, the **Street Name** screen will appear. Enter the street name on the on-screen keyboard until the system displays a list. Select the correct street name from the list by pressing the bar on which the street name is displayed. If the wrong list appears, use the "Back" and "Delete" buttons to correct any errors.

Once you have selected a street, the **House Number** screen will appear. Enter the house number on the numeric keyboard. Press the "OK" button.

A map screen containing the address will appear. If the displayed address is not correct, use the "Back" and "Delete" buttons to go back and correct any mistakes.

When you have correctly entered the destination, press the red "Go To" icon at the bottom of the touch screen, say "Done" out loud, and then press the "DONE" button on the Task Screen to complete the destination entry task.

If you make an error, attempt to fix the destination before moving on. If you make an error but have already said "DONE", continue on to the next step.

When a destination entry task is complete, there is a bit of work we want you to do before beginning the next destination entry task. First, press the 'DEST' button to the right of the Prius video screen to take us back to the desired initial screen for the next destination entry task. Once you've pressed the 'DEST' button, then press the 'NEXT' button on the Task Screen such that both screens are ready for the next task. Once you press the 'NEXT' button, you should hear and see the next destination to enter.

Any questions before we practice this task?

Ten-Digit Dialing Task (Driving Simulator Protocol)

In this task, you will use the Prius information system to dial 10-digit phone numbers. You will be presented the phone number to call. Each trial will require that you complete as many phone dialing tasks as possible.

The first phone number will be presented shortly after the lead vehicle appears. Start the task when instructed to "Begin."

Press the "Phone" icon on the left of the three icons on the touch screen. A numeric keypad will appear on the screen.

Dial the Ten-digit number using this numeric keypad.

If you make an error, use the "Delete" icon (left-pointing arrow on the upper right of the touch screen) to erase an incorrect digit or digits.

When you have correctly entered the phone number, press the green "Call" icon at the lower right of the touch screen, say "Done" out loud, and then press the "DONE" button on the Task Screen to complete the ten-digit dialing task.

If you make an error, attempt to fix the phone number before moving on. If you make an error but have already said "DONE", continue on to the next step.

When a ten-digit dialing task is complete, there is a bit of work we want you to do before beginning the next ten-digit dialing task. First, press the red "End Call" icon (which has replaced the green "Call" icon) to end the call. Once you've ended the call, press the "Info-Phone" button to the right of the screen to return you to the desired initial screen for the next dialing task. Then press the 'NEXT' button on the Task Screen such that both screens are ready for the next task. Once you press the 'NEXT' button, you should hear and see the next phone number to dial.

Any questions before we practice this task?

Seven-Digit Dialing Task (Driving Simulator Protocol)

In this task, you will use the Prius information system to dial 7-digit phone numbers. You will be presented the phone number to call. Each trial will require that you complete as many phone dialing tasks as possible.

The first phone number will be presented shortly after the lead vehicle appears. Start the task when instructed to "Begin."

Press the "Phone" icon on the left of the three icons on the touch screen. A numeric keypad will appear on the screen.

Dial the Seven-digit number using this numeric keypad.

If you make an error, use the "Delete" icon (left-pointing arrow on the upper right of the touch screen) to erase an incorrect digit or digits.

When you have correctly entered the phone number, press the green "Call" icon at the lower right of the touch screen, say "Done" out loud, and then press the "DONE" button on the Task Screen to complete the seven-digit dialing task.

If you make an error, attempt to fix the phone number before moving on. If you make an error but have already said "DONE", continue on to the next step.

When a seven-digit dialing task is complete, there is a bit of work we want you to do before beginning the next seven-digit dialing task. First, press the red "End Call" icon (which has replaced the green "Call" icon) to end the call. Once you've ended the call, press the "Info-Phone" button to the right of the screen to return you to the desired initial screen for the next dialing task. Then press the 'NEXT' button on the Task Screen such that both screens are ready for the next task. Once you press the 'NEXT' button, you should hear and see the next phone number to dial.

Any questions before we practice this task?

Appendix E: Simulator Orientation and DFD Protocol Instructions

SIMULATOR ORIENTATION

This vehicle is a Toyota Prius, which has been modified to collect driving performance data. You will be sitting in this vehicle today to drive a driving simulator. Please get into the driver's seat and adjust the seat to your comfort. You should also make sure that you can reach the buttons on the center console and the task screen located to your right. The seat controls are under the front and on the lower left side of the seat. There is no need to adjust the mirrors as you will not be using them for this experiment.

We have added sensors to the steering wheel, accelerator and brake pedals. These sensors allow us to run the driving simulator without having the vehicle turned on. Your control inputs are recorded by these sensors and input to the simulator to change the roadway image projected on the screen in front of you.

DRIVING TASK INSTRUCTIONS

The STISIM is a fixed-base driving simulator, meaning that it has no motion. The simulated driving environment will be a 4-lane roadway with a lead vehicle traveling in front of you and occasional oncoming traffic.

When the roadway image first appears, your vehicle will be stopped. When instructed to begin driving, you should accelerate to 55 mph and maintain that speed. Within a few seconds after getting around an initial curve, a lead vehicle will appear ahead of you in your travel lane. We call this the "lead vehicle" because it is leading you in the car following task. Your task is to follow that vehicle, adjusting your speed as necessary to maintain a constant following distance behind the lead vehicle. The initial distance at which the vehicle appears ahead of you is the desired following distance (120 feet). You should try to maintain this following distance throughout the entire drive. Please be sure to note this distance when the lead vehicle first appears on the screen because after several seconds the lead vehicle speed will change.

This task is intended to simulate car following on a moderately congested freeway. The lead vehicle speed will change frequently and you should change your speed as necessary to maintain the same following distance. You should continue following at this distance until the lead vehicle disappears. If your following distance increases beyond an acceptable range, an auditory warning tone will sound to indicate that you should speed up and follow more closely. This alarm will sound every 5 seconds until you get within an acceptable range of the lead vehicle.

While driving in the simulator, you should also try to keep the vehicle centered within the travel lane at all times because lane keeping performance will be measured.

Each drive will last almost 4 minutes. At the end of the drive, the lead vehicle disappears. Just keep driving after the lead vehicle disappears, and the simulator screen will shut off and go blank once you pass the point where the lead vehicle disappeared."

Any questions about the driving simulator or car following tasks?

Visual Target Detection Task Description

While driving the simulator, you will be asked to perform a visual target-detection task, which requires you to respond to a sequence of simple targets that appear one at a time on the roadway display. You will respond to a target by pressing a micro-switch that will be attached to your finger. The micro-switch is attached by wire to a small transmitter box that you will wear on your wrist. This equipment allows us to record the time at which each response is made.

The targets are red dots that appear along the horizon at different distances from the center of the roadway, as seen in the example on the screen. When you see a target appear, you should respond as quickly as possible by pressing the micro switch attached to your finger. A target will appear every 3 to 5 seconds and will remain on until the button is pressed, or remain on for about 1.5 seconds if no response is made. You will be scored based on your speed and accuracy in detecting the targets while driving.

Detection Task Instruction and Practice – Stationary Vehicle

First, please place the response button on your left index finger and attach the transmitter box to your wrist so that it is comfortable and the button can be pressed while you are holding the steering wheel.

(Exp: Make sure transmitter box and wire are positioned correctly.)

Now, please try a few button presses in response to the targets. If you press the button quickly, the target will disappear. If you do not respond quickly, it goes out after 1.5 seconds.

Any questions about this task?

Task Performance Feedback Description

Table 65. Task Performance Incentive Criteria

Task	Good Performance	Acceptable Performance	Poor Performance
Car Following	Maintains close following distance consistently with minor deviations	Maintains close following distance mostly with some noticeable deviations	Generally fails to maintain close following distance
Target Detection	Consistently attentive to target detection, detecting most targets	Moderate number of targets not detected	Fails to detect significant number of targets
In-Vehicle Secondary	Performs secondary task continuously with minimal errors	Performs secondary task either intermittently or with moderate number of errors	Performs secondary task with considerable difficulty, slowly, and with moderate number of errors

Appendix F: Eye Tracker Setup Procedure and Script

We use an EyeTracker to help us know where you are looking during the trials. That is the purpose of the two cameras on the dashboard. Please make sure the seat is adjusted to your comfort; then sit as you will drive and look straight ahead. [Exp 1] will adjust the camera tilt, rotation, and focus.

Now please exit the vehicle while I make some measurements.

Please put these five stickers on your face as shown in this photo (one above each eye, one on each cheek, and one above bridge of nose). If any stickers come off during testing, please notify us but do not try to replace it.

(If wearing glasses: Now please remove your glasses momentarily so that we may take a photo.) With a neutral expression and keeping your head still, please look between the cameras. You may relax (and put your glasses back on).

Now please look straight ahead while I adjust the settings. (Tracking parameters)

Now we will calibrate your gaze. Please look at the center of the lens of the camera on the right while limiting your head movements and facial expressions. Now please look at the center of the lens of the camera on the left. Now please look at the center of the lens of the camera on the right. Now please look at the center of the lens of the camera on the left.

Now we will set up what we call the World Model to tell the Eye Tracker when you are looking straight ahead. Please look straight ahead with a steady gaze. Now please look directly at the navigation screen in the dashboard.

Now we will set up the Scene Camera (point it out.) It gives us a view of what you see during testing. On the projection screen in front of the car are four digits, one through four. I will ask you to look directly at each one for a brief time, starting with one and going in order one-two-three-four. I will tell you when to start and stop looking at each digit. Are you ready? Please look at the 1 while I record the data. (about 3 seconds). Ok. Now look directly at the 2. OK. (Exp repeats for the 3 and 4.) (This entire process may require a couple of repetitions.)

Now we will use the same four points to check the quality of our calibration. I will ask you to look directly at each one for a brief time, starting with one and going in order one-two-three-four. I will tell you when to start and stop looking at each digit. Are you ready? Please look at the 1. Ok. Now look directly at the 2. OK. (Exp repeats for the 3 and 4.) Now please look straight ahead. OK. Now please look at the Prius navigation screen. Ok.

Appendix G: Simulator Sickness Questionnaire

Participant Number: ___

Directions:

Circle one option for each symptom to indicate whether that symptom applies to you right now.

1. General Discomfort.................... None .... Slight .... Moderate......... Severe
2. Fatigue .... None .... Slight .... Moderate......... Severe
3. Headache .... None .... Slight .... Moderate......... Severe
4. Eye Strain .... None .... Slight .... Moderate......... Severe
5. Difficulty Focusing .... None .... Slight .... Moderate......... Severe
6. Salivation Increased .... None .... Slight .... Moderate......... Severe
7. Sweating .... None .... Slight .... Moderate......... Severe
8. Nausea .... None .... Slight .... Moderate......... Severe
9. Difficulty Concentrating .... None .... Slight .... Moderate......... Severe
10. "Fullness of the Head" .... None .... Slight .... Moderate......... Severe
11. Blurred Vision .... None .... Slight .... Moderate......... Severe
12. Dizziness with Eyes Open .... None .... Slight .... Moderate......... Severe
13. Dizziness with Eyes Closed .... None .... Slight .... Moderate......... Severe
14. *Vertigo .... None .... Slight .... Moderate......... Severe
15. **Stomach Awareness .... None .... Slight .... Moderate......... Severe
16. Burping...................................... No..................... Yes.................... If yes, no. of times ______
17. Vomiting.................................... No..................... Yes.................... If yes, no. of times ______
18. Other ______________________________________

* Vertigo is experienced as loss of orientation with respect to vertical upright.

** Stomach awareness is usually used to indicate a feeling of discomfort which is just short of nausea.

Appendix H: Occlusion Protocol and Secondary Task Instructions

The occlusion procedure simulates the visual demands of driving in a stationary setting. While sitting in this vehicle, you will be asked to perform a series of tasks, which we call "Secondary Tasks." While performing these secondary tasks, you will wear a set of glasses with lenses that can be made to be either transparent or opaque. When they are transparent, you will be able to see normally; however, when they are opaque, you will not be able to see through them. When the glasses are transparent, you will be able to see the in-vehicle task controls and displays. When the glasses are opaque, you will not be able to see the task controls and displays but you may continue to work on the task if you are able. The opaque condition is intended to simulate the time during which you would need to look at the roadway ahead to maintain vehicle control in driving.

At the beginning of each trial, after you tell me that you are ready to begin, the task stimulus will be presented auditorily and be displayed on the computer screen located to the right of the center console, for your use in case you forget. We call this the Task Screen; it has a touch screen interface. If you would like to have the task instruction repeated, you may press the REPEAT button on the Task Screen. The stimulus will then be repeated. When you understand the stimulus and are ready to proceed, you should press the GO button on the Task Screen. The glasses will then become opaque for several seconds, during which time you will hear the word "Begin". At this time, the glasses will become transparent, and you should immediately begin, working quickly and accurately to complete the secondary task. When you have completed the task, you should say "DONE" aloud and then press the DONE button on the Task Screen. This is the end of the trial and you should wait for instructions to begin the next trial.

I want to say a few words about errors: Mistakes are inevitable. If you make a mistake while performing a secondary task, please try to correct the error before moving on. We will provide specific information about how to recover from errors for each secondary task. It is important that you try to complete each task if possible.
You will complete 5 trials of each task.

To demonstrate the occlusion procedure, we will use a simple visual search task, called the "Circles" task. In this task, you will look at a series of displays presented one at a time on the small computer screen. Each display consists of a pattern of circles. In some displays there will be one circle that is slightly larger than the others. The larger circle is called the target circle. Your task is to determine if a target circle is present and to reply by using the 'YES' or 'NO' buttons on the screen. When you press the 'YES' or 'NO' button, this will record your response and begin the next trial.

This task is paced by you, so you should try to complete as many trials as you can during this practice session. Remember to look for a larger circle, and if it is present, press 'YES'. If a larger circle is not present, then press 'NO'.

[Run practice file.]

Any questions before we train you on the first task?

Radio Tuning (Occlusion Protocol)

In this task you will tune the radio to a designated frequency by using the tuning knob at the upper right corner of the radio/navigation module. This vehicle has buttons on the steering wheel for this purpose, but we want you to use the knob on the console at all times. During the trial, you will select several different radio frequencies, one at a time. You will be presented the band (AM or FM) and the frequency. The trial will require that you complete up to 5 radio tuning instances.

The occlusion glasses will be transparent. When you are ready, the first radio frequency will be presented both auditorily and visually on the Task Screen. Press the REPEAT button if you want to have it repeated, or press the GO button when you are ready to proceed. This will initiate an occluded interval, during which you will hear the instruction to “Begin.” When the glasses become transparent, you should start the task immediately, working quickly and accurately to complete the secondary task.

Press the “AUDIO” button at the bottom of the column of buttons to the left of the Prius video screen. The audio display will then appear on the video screen. Note that the Prius video screen is a touch screen.

Select the frequency band by pressing the AM or FM button located to left of the video screen. (Note, we want you to use the buttons for this task, not the on-screen AM/FM tabs.) The current band is displayed in the upper left of the screen and current frequency in the upper right. If you select the wrong band, press the button for the appropriate band. (After about 20 seconds of inactivity, the display will revert to the MAP display or the ‘Phone’ screen, depending upon which was most recently used. If this occurs, press the “AUDIO” button again to return to the audio screen.)

Use the tuning knob, located to the upper right of the screen, to adjust the frequency. When you have reached the specified frequency, say “Done” out loud, and then press the “DONE” button on the Task Screen to complete the radio tuning task.

If you make an error, attempt to fix the band and frequency before moving on. If you make an error but have already said “DONE”, continue on to the next step.

When a radio tuning task is complete, there is a bit of work we want you to do before beginning the next radio tuning task. First, press the ‘MAP’ button to the right of the Prius video screen to take us out of radio mode, such that the next radio tuning task will also commence by pressing the ‘AUDIO’ button. Once you’ve pressed the ‘MAP’ button, we will setup the Task Screen such that both screens are ready for the next task.

Any questions before we practice this task?

<u>Navigation System Destination Entry by Address (Occlusion Protocol)</u>

In this task you will enter destinations into the navigation system. You will be presented the city, street name, and house number. Each trial will require that you complete up to 5 destination entry tasks.

The occlusion glasses will be initially transparent. When you are ready, the first destination will be presented both auditorily and visually on the Task Screen. Press the REPEAT button if you want to have it repeated, or press the GO button when you are ready to proceed. This will initiate an occluded interval, during which you will hear the instruction to "Begin." When the glasses become transparent, you should start the task immediately, working quickly and accurately to complete the secondary task.

Four icons will be displayed in the middle of the screen. Note that the Prius video screen is a touch screen. Press the icon labeled "**Address**". The system will display three options for destination entry.

We will always enter the city first. Press the "**City**" button. A keyboard will appear on the screen. Enter the city name on the on-screen keyboard until the system displays a list. Select the city from the list by pressing the bar on which the city name is displayed.

Two buttons can help you correct **errors**. If you make an error during keyboard entry, pressing the "Delete" button (a left-pointing arrow in the upper right portion of the on-screen keyboard) will erase the most recently entered letter, one at a time. If the system has generated a list which does not contain the specified city or street, pressing the "Back" button (a U-shaped arrow pointing left at the top right portion of the screen) will allow you to go back to the previous screen. This "Back" button is available on every screen.

Once you have selected a city, the **Street Name** screen will appear. Enter the street name on the on-screen keyboard until the system displays a list. Select the correct street name from the list by pressing the bar on which the street name is displayed. If the wrong list appears, use the "Back" and "Delete" buttons to correct any errors.

Once you have selected a street, the **House Number** screen will appear. Enter the house number on the numeric keyboard. Press the "OK" button.

A map screen containing the address will appear. If the displayed address is not correct, use the "Back" and "Delete" buttons to go back and correct any mistakes.

When you have correctly entered the destination, press the red "Go To" icon at the bottom of the touch screen, and say "Done" out loud, and then press the "DONE" button on the Task Screen to complete the destination entry task.

If you make an error, attempt to fix the destination before moving on. If you make an error but have already said "DONE", continue on to the next step.

When a destination entry task is complete, there is a bit of work we want you to do before beginning the next destination entry task. First, press the 'DEST' button to the right of the Prius video screen to take us back to the desired initial screen for the next destination entry task. Once you've pressed the 'DEST' button, then we will setup the Task Screen such that both screens are ready for the next task.

Any questions before we practice this task?

Ten-Digit Dialing Task (Occlusion Protocol)

In this task, you will use the Prius information system to dial 10-digit phone numbers. You will be presented the phone number to call. Each trial will require that you complete up to 5 phone dialing tasks.

The occlusion glasses will be transparent. When you are ready, the first phone number to dial will be presented both auditorily and visually on the Task Screen. Press the REPEAT button if you want to have it repeated, or press the GO button when you are ready to proceed. This will initiate an occluded interval, during which you will hear the instruction to “Begin.” When the glasses become transparent, you should start the task immediately, working quickly and accurately to complete the secondary task.

Press the “Phone” icon on the left of the three icons on the touch screen. A numeric keypad will appear on the screen.

Dial the Ten-digit number using this numeric keypad.

If you make an error, use the “Delete” icon (left-pointing arrow on the upper right of the touch screen) to erase an incorrect digit or digits.

When you have correctly entered the phone number, press the green “Call” icon at the lower right of the touch screen, and say “Done” out loud, and then press the “DONE” button on the Task Screen to complete the ten-digit dialing task.

If you make an error, attempt to fix the phone number before moving on. If you make an error but have already said “DONE”, continue on to the next step.

When a ten-digit dialing task is complete, there is a bit of work we want you to do before beginning the next ten-digit dialing task. First, press the red “End Call” icon (which has replaced the green “Call” icon) to end the call. Once you’ve ended the call, press the “Info-Phone” button to the right of the screen to return you to the desired initial screen for the next dialing task. Then, we will setup the Task Screen such that both screens are ready for the next task.

Any questions before we practice this task?

Seven-Digit Dialing Task (Occlusion Protocol)

In this task, you will use the Prius information system to dial 7-digit phone numbers. You will be presented the phone number to call. Each trial will require that you complete up to 5 phone dialing tasks.

The occlusion glasses will be transparent. When you are ready, the first phone number to dial will be presented both auditorily and visually on the Task Screen. Press the REPEAT button if you want to have it repeated, or press the GO button when you are ready to proceed. This will initiate an occluded interval, during which you will hear the instruction to “Begin.” When the glasses become transparent, you should start the task immediately, working quickly and accurately to complete the secondary task.

Press the “Phone” icon on the left of the three icons on the touch screen. A numeric keypad will appear on the screen.

Dial the Seven-digit number using this numeric keypad.

If you make an error, use the “Delete” icon (left-pointing arrow on the upper right of the touch screen) to erase an incorrect digit or digits.

When you have correctly entered the phone number, press the green “Call” icon at the lower right of the touch screen, and say “Done” out loud, and then press the “DONE” button on the Task Screen to complete the seven-digit dialing task.

If you make an error, attempt to fix the phone number before moving on. If you make an error but have already said “DONE”, continue on to the next step.

When a seven-digit dialing task is complete, there is a bit of work we want you to do before beginning the next seven-digit dialing task. First, press the red “End Call” icon (which has replaced the green “Call” icon) to end the call. Once you’ve ended the call, press the “Info-Phone” button to the right of the screen to return you to the desired initial screen for the next dialing task. Then, we will setup the Task Screen such that both screens are ready for the next task.

Any questions before we practice this task?

Appendix I: Debrief and Payment Summary

That brings us to the end of the experiment. Do you have any questions? If not, then I will calculate the amount of money you will receive.

Your total pay will be $______.
This is based on ___ hours at $42 per hour.
Your reimbursement for mileage is $_____.

I would like to thank you for your participation. The data that we have collected is valuable to help us understand distraction and driving behavior. Do you have any other questions regarding your participation today?

Receipt for payment:

	Amount $________________
Description	**Test Participant Payment**
Charged to	**NVS-**
Received by	**Printed Name** *Signature*
Date Received	

Appendix J: Basic Recruitment Information

Table 66 shows the total number of people who responded to the recruitment advertisements (Appendix B) by age and gender.

Table 66. Total Number of Respondents to Recruitment Ads

Age	Female	Male	Total
18-25	70	80	150
26-59	148	146	294
60-75	61	143	204
Total	279	369	648

As can be seen in Table 66, females in the oldest age range were the least likely to respond to such advertisements. Table 67 shows the number of respondents who actually met all of the qualification requirements, again by age and gender.

Table 67. Total Number of Qualified Respondents

Age	Female	Male	Total
18-25	31	41	72
26-59	69	66	135
60-75	35	77	112
Total	135	184	319

As can be seen in comparing these tables, approximately half of the people who respond to the recruitment ads are actually qualified for participation. Table 68 shows the number of participants who were actually scheduled in order to obtain 24 successful test completions per age range.

Table 68. Total Number of Scheduled Participants to Obtain 24 per Age Range

Age	Female	Male	Total
18-25	12	12	24
26-59	12	12	24
60-75	16	15	31
Total	40	39	79

As can be seen in Table 68, extra participants had to be scheduled in the oldest age range in order to successfully complete 24 valid tests. Table 69 makes use of the data from the other tables, and converts the results to percentages (proportions of total respondents and qualified respondents).

Table 69. Qualified and Scheduled Respondents, as Proportions of Total and Qualified Respondents

	% Respondents Qualified			**% Total Respondents Scheduled**			**% Qualified Respondents Scheduled**		
Age	Female	Male	Total	Female	Male	Total	Female	Male	Total
18-25	44.3%	51.3%	48.0%	17.1%	15.0%	16.0%	38.7%	29.3%	33.3%
26-59	46.6%	45.2%	45.9%	8.1%	8.2%	8.2%	17.4%	18.2%	17.8%
60-75	57.4%	53.8%	54.9%	26.2%	10.5%	15.2%	45.7%	19.5%	27.7%
Total	48.4%	49.9%	49.2%	14.3%	10.6%	12.2%	29.6%	21.2%	24.8%

As stated previously, recruitment was done by advertising on Craigslist and in local newspapers. The ads contained a link to a secure online contact information form which asked respondents to list their name, E-mail address, ZIP code, age, gender, telephone number, and how they found out about the study (source). Respondents were also asked whether they had participated in other recent studies, and if so, when. The following paragraphs provide more information about the details of the recruitment process and rate of success.

The application process was in two parts. After submitting the contact information form, each respondent was sent an automatically generated link to a secure online screening form. This was done in part to test the E-mail address submitted, but also save time in screening potential participants and avoid having to ask potentially embarrassing questions over the telephone.

A Craigslist ad was posted first, and then local newspaper ads were published. The first applications were received within minutes of the posting of a Craigslist ad, with the first participant getting scheduled the following day. Participants were subsequently scheduled over an interval spanning 41 days. There were 27 days on which at least one test participant was scheduled, mostly Monday through Friday, but including one Saturday.

Six categories were used for scheduling equal numbers of participants. However, the number of people who submitted the forms and the proportion of "qualified" candidates varied widely by category. Approximately equal numbers of men and women applied in the broad 26 to 59 year old age range, with a similar proportion of candidates deemed to be qualified. In the narrowest age range (18 to 25 years old), marginally more men applied than women (80 versus 70), with a slightly higher percentage of males qualifying on the basis of their screening form responses. In the oldest age range (60 to 75 years old), however, more than twice as many men as women applied, and the percentage of men qualifying for possible participation was more than double that of women.

For the two younger age groups, only the minimum number of 12 participants was scheduled in each of the four categories. However, in the oldest group, four additional women and three more men were scheduled, for various reasons: one was a no show, one cancelled on short notice, a couple suffered from fatigue/discomfort and chose not to complete the testing, and others brought glasses to wear (which adversely affected eye tracker accuracy) despite previously stating they were not required to wear glasses to drive.

Some of the reasons that respondents were removed from the "qualified" list included the following:

(1) Must wear glasses to drive
(2) Drive less than 7000 miles per year
(3) Do not use a cell phone while driving
(4) Participated in the prior distraction study which was similar to this one
(5) Medical issues such as heart problems, susceptibility to motion sickness, or high blood pressure not controlled by medicine
(6) Submitted the contact form but not the screening form
(7) Unavailable at the times that test sessions were being scheduled

Note that not being in the "qualified" group did not *automatically* mean that the respondent could not be scheduled for participation. If needed for participation, some could have been contacted and informed to complete the screening form either online or by phone. Note that each of the 93 respondents who filled out the contact form but not the screening form had been sent two E-mails: one generated automatically with a personalized link to the screening form, and another one later to remind them that if they were still interested in participating, they should submit the screening form.

DOT HS 811 726
March 2013

U.S. Department
of Transportation

**National Highway
Traffic Safety
Administration**

www.nhtsa.gov

9521-032013-v1a