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

People Saving People
http://www.nhtsa.dot.gov

DOT HS 808 467

Final Report Supplement

October 1996

Heavy Vehicle Driver Workload Assessment
Task 5: Workload Assessment Protocol

This research was supported (in part) by the National Highway Traffic Safety Administration (NHTSA), U. S. Department of Transportation, Under Contract No. DTNH22-91-C-07003. The opinions, findings and recommendations contained herein are those of the authors, and do not necessarily represent those of the NHTSA.

Technical Report Documentation Page

1. Report No.

2. Government Accession No.
DOT HS 808 467 (5)

3. Recipient's Catalog No.

4. Title and Subtitle
Final Report Supplement --Task 5: Heavy Vehicle Driver Workload Assessment Protocol

5. Report Date
June, 1995

6. Performing Organization Code

7. Author(s)
Louis Tijerina, Steven Kiger, Thomas Rockwell, and Walter Wierwille

8. Performing Organization Report No.

9. Performing Organization Name and Address
Battelle Memorial Institute
505 King Avenue
Columbus, OH 43201-2693

10. Work Unit No. (TRAIS)

11. Contract or Grant No.
DTNH22-91-C-O7003

12. Sponsoring Agency Name and Address
office of Crash Avoidance Research
National Highway Traffic Safety Administration
400 7th Street, MD-50
Washington, DC 20590

13. Type of Report and Period Covered
Final Report
July, 1991-September, 1995

14. Sponsoring Agency Code

15. Supplementary Notes
Contracting Officer's Technical Representative: Dr. Michael J. Goodman (NRD-50)

16. Abstract

This report presents a description of a prescriptive workload assessment protocol for use in evaluating in-cab devices in heavy vehicles. The primary objective of this heavy vehicle driver workload assessment protocol is to identify the components and processes necessary to conduct an empirical appraisal of the potential of an in-cab device to distract drivers from the driving task. The methodological approaches and experimental design strategies that may be used to conduct a workload assessment are presented in detail. Included in this protocol are sets of workload measurements that can demonstrate to what extent in-cab devices intrude onto the primary driving task. The scientific and theoretical bases for how these various workload measures related to safe vehicle operation are discussed. The sets of workload measures comprise visual allocation measures, driver steering, pedal, and manual activity measures, driver-vehicle performance measures, and driver subjective assessments.

17. Key Words
human factors, heavy vehicle, driver workload measures, safety, assessment protocol

18. Distribution Statement

19. Security Classif. (of this report)

20. Security Classif. (of this Page)

21. No. of Pages
184

22. Price

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

TABLE OF CONTENTS

TABLE OF CONTENTS (CONTINUED)

Page

TABLE OF CONTENTS (CONTINUED)

Page

TABLE OF CONTENTS (CONTINUED)

Page

LIST OF TABLES

LIST OF FIGURES

TABLE OF CONTENTS (CONTINUED)

Page

LIST OF APPENDICES

A Heavy Vehicle Driver Workload Assessment Protocol: In-Vehicle Device Evaluation from a Driver-Oriented Perspective

1.0 PERSPECTIVE

1.1 BACKGROUND

The heavy vehicle driver's primary task is to transport goods and materials efficiently while safely maintaining control of the vehicle at all times. In the past 10 years, a wide variety of driver interface products have been proposed and developed for use in heavy trucks. These systems include the following:

- Satellite tracking, land navigation, and route guidance systems
- Text displays (e.g., pick-up address, package type)
- Vehicle subsystem~monitoring and warning systems (e.g., tire pressure, oil pressure, brake failure, load shifting)
- Computerized trip recorders (e.g., automatic record of speed, RPM, stops; driver entry of fuel purchase; state-line crossings)
- Sophisticated communication links (e.g., cellular phone systems)
- Proximity warning systems (e.g., infra-red and TV systems)
- Changes to existing control and display systems (e.g., head-up displays).

Many of these high technology devices introduce subsidiary tasks which may compete with the primary task of driving. Some of these devices (e.g., anti-lock brakes) can be used concurrently with the primary driving task without interference, but others may not. Of all the contributing factors associated with crashes on the nation's highways, nothing comes close to "driver inattention" in the frequency with which it is called out. Studies of crash statistics and reports suggest that driver inattention is a primary or contributing factor in as many as 50% of all traffic accidents (Treat et al., 1977; Sussman, Bishop, Nadnick, and Walter, 1985). This suggests that there is good reason to be concerned about the potential for a new in-cab device, however, well intentioned its developers and attractive its features, to distract the driver from the primary driving task.

The inventors and manufacturers of high technology in-cab systems intend for these systems to enhance commercial vehicle operations efficiency and effectiveness, to help the driver in doing the job at hand, and to be safe. However, without a driver-oriented assessment of a high technology device, the safety of the system remains largely unknown. What is needed is a set of techniques with which to assess the safety implications of a device from the driver's perspective. In response to this need to assess the safety implications of high technology systems that might be introduced into heavy trucks, the National Highway Traffic Safety Administration (NHTSA) has supported the development of **The Heavy Vehicle Driver Workload Assessment Protocol.** It is hoped that this protocol contributes toward keeping the nation's highways and commercial vehicle operations safe and productive through technology assessments which do not overlook the driver.

1.2 PURPOSE

The primary objective of a workload assessment of an in-cab device or system is to empirically assess the potential of that device to distract the driver from the driving task to the extent that safety may be compromised. Given this primary objective, this document describes a process by which such an assessment may be carried out. It is intended to be applicable to a wide variety of in-cab or in-vehicle devices. In addition, it is intended to support a wide range of individuals who are charged with the responsibility of assessing the distraction potential of new high-technology for use in heavy vehicles.

This wide scope necessitates a general document that provides guidance on the conduct of workload assessments. This document presents a series of stages which, if carried out, will promote a more thorough device evaluation. It does not, in general, provide a single evaluation procedure because variation in technologies and their uses by drivers does not allow it.

This document is targeted to several potential users:

- The protocol document is intended to be of use for new or novice evaluation team personnel and for test engineers who may have little or no experience with driver-oriented data collection and assessment.

- The DOT/NHTSA may use it as a guidance document to manage contractors retained to carry out safety evaluations, especially operational tests. The steps/stages discussed in the document may serve as a set of milestones for a formal evaluation and ensure that all relevant factors have been addressed.

- The document may be of use to researchers in the field of driver workload. Experience has shown that there are special aspects of driver-centered device evaluation that are different from both psychological measurement or engineering assessments. For this reason, there is value in having a guide to the development and conduct of a device evaluation.

While the protocol document is intended to be practical, it describes an idealized process of evaluation, i.e., it is prescriptive or tells what ought to be done. On the other hand, the rigors of realistic evaluation on a specific device or system in a specific circumstance may vary somewhat from what is described here. As Meister (1986) has pointed out, each evaluation has its own unique qualities that may or may not be adequately expressed in general principles.

The protocol aims to assess the extent to which an in-cab device is affecting driver workload and the extent to which it has an adverse impact on safety. The protocol therefore presents guidance for the conduct of an on-the-road field test with an instrumented vehicle. This reflects the view that field observations are the most reliable way of assessing how an in-cab device impacts drivers. However, there are safety concerns that limit the range of workload that might be imposed on the driver in a test situation. Therefore, a driving simulator assessment may be a useful adjunct to assess the impact of device use on factors such as object and event detection, factors that cannot be staged safely or easily on the road.

In the next sections, the motivation and logical foundations behind workload assessment, as presented in this protocol document, are reviewed. This includes a simple model of driving to derive workload measures, the rationale behind omitting certain types of measures from this protocol, approaches to assessing the validity of workload measures, the scientific bases of establishing the safety-relevance of workload measures, and a simple theory of crashes that leads to an emphasis on relative workload assessment

1.3 A SIMPLE MODEL OF DRIVING

A model of driving is needed to point to possible measures of driver behavior and performance that can be sensitive to workload effects associated with in-vehicle device use. The model should also indicate other sources of variation that can overwhelm a workload effect. The model also provides the hypothetical link between workload measures and highway safety. Thus, a simple theory of driving serves as the basis for the workload assessment protocol measures that will be presented in this document.

Figure l-l presents a control-theoretic model of the driving task of lanekeeping taken from Hess (1987). This model assumes that the driver receives inputs about the current lane position (y)

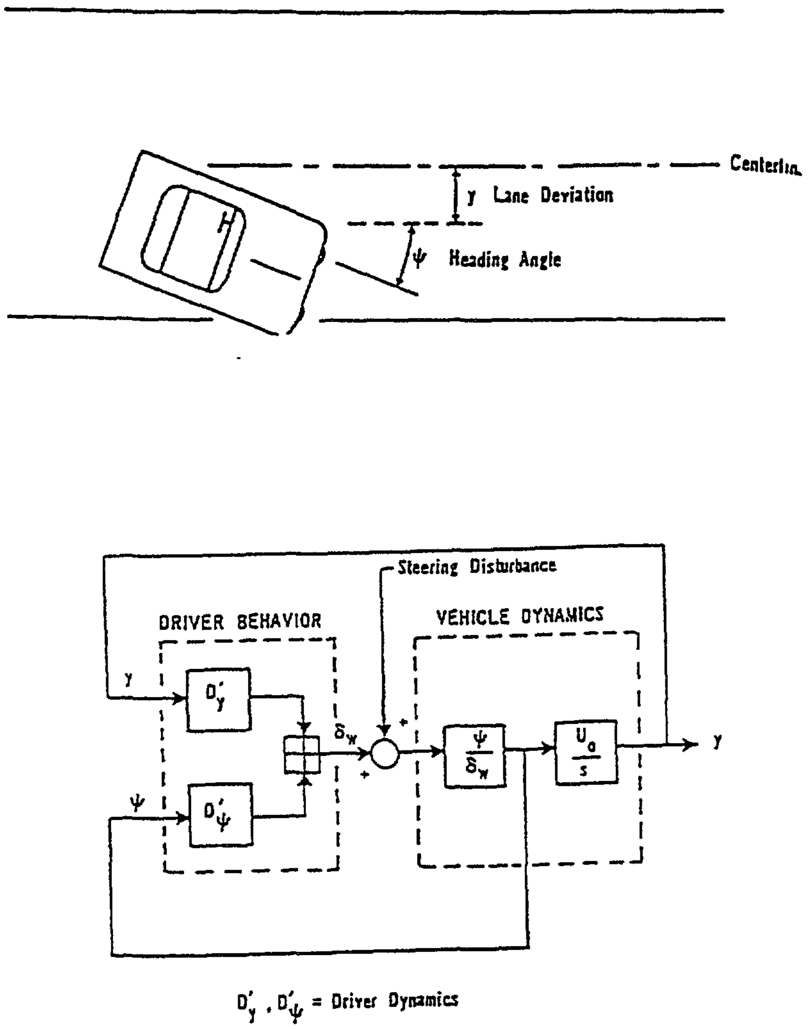


Figure 1-1. A Generic Control-Theoretic Model of Driving. Source: Hess, 1987.

and heading angle (ψ). There is evidence in the human factors literature that such inputs are largely visual in nature. These inputs are considered by the driver in light of his or her particular goals, situational understanding of the current driving conditions, driving style, and vehicle characteristics (this is what is intended to be conveyed by the Driver Behavior block). The driver then changes the steering wheel angle (aw) as appropriate. This steering wheel input, along with steering disturbances that arise from such factors as wind gusts and road surface characteristics, combine with vehicle dynamics to determine the vehicle's heading angle (ψ) moment by moment, and also lane position (y). Note that even if steering disturbances were zero, heading angle is one integration removed from steering angle as a function of time. Furthermore, lane position (on a straightaway, at least) is one time integration removed from heading as a function of time and, thus, two time integrations removed from steering angle inputs by the driver. This suggests that vehicle performance measures may be less sensitive than driver steering or pedal inputs. In turn, these may be less sensitive to workload demand than visual allocation measures because of the many factors that can influence control input and vehicle performance measures.

This model is specific to the driving task of lanekeeping or lateral control. Similar concepts apply to speed maintenance, headway maintenance, and other aspects of longitudinal control. Again, visual inputs are considered by the driver in light of his or her particular goals, situational understanding of the current driving conditions, driving style, and vehicle characteristics. The driver changes the accelerator. pedal angle (ap) or applies brakes, as deemed appropriate. These longitudinal control measures may be influenced by factors other than device workload, factors such as vehicle dynamics (e.g., momentum, braking efficiency, etc.) and disturbance inputs (head- or tail-winds, vertical roadway alignment, etc.). Mechanical relationships also introduce damping or lags in driver inputs to longitudinal control. Therefore, the previous paragraph's comments about relative sensitivity of measures applies equally well here.

This simple approach suggests a set of workload measurements that may shown intrusion of in-vehicle device use onto the primary task of safely controlling the vehicle at all times. These categories of measurement are:

- Visual Allocation Measures. These include measures of how long, how frequently, and how likely it is that a driver looks at a particular location (e.g., in-vehicle device), as well as the sequence of glances. Given that visual attention is the primary input to safe driving, such measures should be relevant. Furthermore, since they are most readily under the driver's control, they are likely to be the most sensitive workload measures as well. Appendix A provides more details about such measures.

- Driver Steering, Pedal, and Manual Activity Measures. These measures capture the inputs drivers make based on the visual information received (or not received), coupled with driver strategies and corrections for various disturbances. The logic behind such measures is described in Appendix B of this report.

- Driver-Vehicle Performance Measures. These are measures of lateral control (i.e., lanekeeping measures), longitudinal control (e.g., measures of speed, headway, accelerations and decelerations), and way finding (measures such as time-to-arrival, missed turns, exit ramps, and entrance ramps). These measures assess the overall quality of driving in terms of meeting goals in a safe and efficient manner. Appendix C presents Driver-Vehicle Performance Measures.

- Driver Subjective Assessments. These are measures that do not fall out of the simple model of driving, per se, but may nonetheless be important. Driver subjective assessments may include workload assessment measures (see Appendix D). They also include driver feedback on debriefing questions about the in-vehicle device under evaluation. The reason for including subjective assessments is that the driver may be in an excellent position to indicate problems or concerns with a new technology, and some means to capture such information is part of a comprehensive workload assessment protocol. Driver subjective assessments can provide an impression of the demands posed by performing the driving task plus the load of in-cab services (i.e., demand-driven workload).

Beyond the classes of response measures just presented, three more classes deserve mention. One class of response measures that might be integrated into driver workload assessment is in-cab device performance. This might include such measures as in-cab task completion time, errors, recovery procedures followed, etc. These are not formally included in this protocol because the performance on the in-cab device is taken as a “given”. That is, the protocol emphasizes the consequences of in-cab device use on the measures introduced above. If these consequences are aggravated with, say, in-cab device use errors, then this is simply taken as a part of that device’s characteristics.

A second class of response measures that might be included is macroscopic driver-vehicle performance. Examples of such measures might include number of missed turns, stop light violations, stop sign violations, or time-to-arrival at a way point or destination. Clearly, these measures may be included in a safety evaluation of driver workload. However, safety dictates of on-the-road safety evaluations may render such measures purely happenstance and may be masked by safety precautions.

In a related vein, Dingus (1995) has recently proposed that safety-relevant evaluation should include measurement of traffic conflicts, a third class of workload measures. The traffic conflict technique has been proposed and tested as a means to improve roadway design, e.g., intersection design (Older and Spicer, 1976). In essence, the traffic conflict technique examines near-crash or potential crash situations to provide information about hazardous conditions. These situations are characterized by human observers interpreting high decelerations (characterizing abrupt stopping maneuvers), skids, or evasive steering maneuvers as evidence of traffic conflicts. There is a purported relationship between traffic conflicts and crashes such that traffic conflicts may be much more numerous (perhaps 1000 traffic conflicts per every one crash that occurs). Thus, it is more likely that one can observe traffic conflicts.

Theory is an appealing approach. However, the traffic conflict technique has been criticized on the grounds that evaluation studies fail to confirm the predictive benefits of the method (Williams, 1981). Areas for improvement include more objective and standardized measurements of what constitute traffic conflicts and methods that do not depend on evasive maneuvers to predict crash occurrence (because many crashes occur without being preceded by an avoidance maneuver). There is also concern by the present authors that conduct of a workload assessment must be carried out so as to minimize the potential for traffic conflicts to be observed. Finally, the likelihood of observing a traffic conflict is expected to be low over the short-run of an on-the-road evaluation. For example, if traffic conflicts occur at a rate of 1000 to 1 with respect to traffic crashes, and a traffic crash occurs once every 10 years, on average, then if these likelihoods are evenly distributed (which they are probably not for the driver participating as a test participant in a controlled evaluation), then one must observe the driver for at least 3-4 days to obtain one traffic conflict. Thus, while traffic conflicts must of course be noted and included in the reporting of a device evaluation, traffic conflicts do not currently play a substantial role in the measurement system presented in this document.

In summary, the average driver is viewed as a basically rational person working within a context of situational understanding and motivations (see Wierwille, Tijerina, Kiger, Rockwell, Lauber, and Bittner, 1992 for further elaboration) to control the vehicle in a safe manner. Visual inputs largely guide driver control activities; device demand that affects visual allocation can be safety-relevant and so visual allocation measures are included in the workload assessment protocol. The driver effects control of the vehicle through manual activities (steering inputs), and pedal activities (accelerator and brake inputs); device workload may disrupt such control activities and so driver in-cab behavior measures are included in the workload assessment system. The trajectory of the vehicle in space and time ultimately determines crash occurrence or non-occurrence; workload measures of such driver-vehicle performance are intrinsically safety-relevant. A crash is considered to always involve undesired contact between the vehicle and other objects (other vehicles, roadside appurtenances, pedestrians, etc.). Finally, subjective assessments from drivers are included in workload assessment to capture important information about driver behaviors and perceptions that may be overlooked or otherwise be difficult to extract from the other measures. Based on a simple model of driving, measures for in-cab device workload assessment can be determined. The relative sensitivity of such measures is considered to form a hierarchy. The rationale for omitting other, alternative measures of workload is discussed next, followed by a discussion of the scientific bases for relating workload measures to highway safety.

1.4 RATIONALE FOR THE OMISSION OF ALTERNATIVE MEASURES OF WORKLOAD

Human factors as a science has repeatedly been confronted with questions about the relevance of its measures to real-world systems. In a very important paper on this topic, Chapanis (1970) illustrated the issue (reproduced in Table l-l). Real-world systems have criteria expressed in terms such as safety, ease of use, and convenience, among others. On the other hand, human factors and ergonomics research often use dependent measures such as heart rate, EEG, reaction times to arbitrary tones, and muscle tension, among others. The obvious question is what dependent measures on the first-half of Table l- 1 have to do with real-world concerns listed in the second-half of the table. Chapanis notes that the answer to such a question is neither obvious nor simple. He goes on to note that any real-world criterion (e.g., safety) is likely to be a complex phenomenon that is affected by many factors. This implies that proper assessment will likely impose a need to measure many different aspects of the human and the situation. It is doubtful that it will ever be possible to assess real-world systems criteria with a single experimental measure. The present authors add that multiple measures must be selected by reference to a model of the real-world system that links measures to criteria more directly rather than less directly. Thus, a model of driving can explain what steering angle has to do with where a vehicle is in space at a given instant; it cannot so readily relate heart rate variability to vehicle location in space and time.

Some researchers may argue that even a less direct measure of workload may be preferred if it is relatively sensitive (i.e., tends to be correlated with workload as determined by other means). This justification must be scrutinized in the context of safety, however, because it compounds substantial problems of causal inference. A simple statistical example will illustrate this point.

Suppose that there is a correlation of 0.7 between heart rate (HR) and a "true” measure of workload (WL) ($r_{HR\,WL}$ = 0.7) and that the correlation between driver workload and a "true” (marginal) measure of highway safety (S) is also 0.7 (i.e., $r_{WL.S}$ = 0.7). (For the moment set aside the difficulties in actually determining such coeffkients). Such correlations in applied human factors research are unusually large, but this only reinforces the point to be made. If the square of a correlation coefficient is computed, the result is the proportion of variability shared between two measures or variables. Accordingly, if r = 0.7, this implies that 0.49 or 49% percent of the variability in one measure (e.g., S) is predictable from variability in the other measure (e.g., WL).

Table 1-1. Common dependent measures (or criteria) used in human factors research (in the column on the left) and typical general systems criteria (in the column on the right). Source: Chapanis, 1970.

Human Factors Criteria or Dependent Measures
Accuracy (or, equivalently, errors)
Cardiovascular responses (e.g., heart rate)
Critical Flicker Fusion Frequency
Electro-Encephalographic Measures (EEG)
Energy Expenditure
Muscle Tension
Psychophysical Thresholds
Ratings (Annoyance, Workload, etc.)
Reaction Time
Respiratory responses
Speed
Trials to Learn
Mental Arithmetic Score
Systems Criteria
Anticipated life of the system
Appearance
comfort
Convenience
Ease of use
Familiarity
Initial Cost
Maintainability
Manpower requirements
Operating cost
Reliability
Safety
Training Requirements

What can be inferred about the degree of association between heart rate (HR) and safety (S)? That is, what is r_{HRS} ? The answer is depicted in the Venn diagrams in Figure l-2. In the figure, overlap of the rectangles signifies the proportion of variation predictable from one measure by another assuming the 0.7 correlations (and 0.49 proportions) introduced earlier. The leftmost diagram shows the case were $r_{HR,S} = 0.0$, i.e., there is no predictive validity at all (see below for further discussion). The Heart Rate variation overlaps with about 50% of the Workload variation and the Workload variation overlaps about 50% with the Safety variation but Heart Rate and Safety do not overlap at all, indicating they are measuring different things. The rightmost diagram shows the case where $r_{HR-S} = 1.0$. Here the Heart Rate variation and Safety variation overlap completely. Furthermore, both overlap about 50% with the Workload variability, as dictated by the individual correlations. This demonstration points out the danger of using measures that have no substantive connection to the real-world system of interest simply because they are reportedly sensitive. This is why the workload measures presented in this document are derived from a model of driving (subjective assessments, excepted).

1.5 VALIDITY OF WORKLOAD MEASURES FOR PREDICTING SAFETY

Workload assessment must be carried out safely for ethical and legal reasons. It is not possible to assess in-vehicle device workload directly in terms of crashes that occur during an evaluation. Instead, safety must be inferred from workload measures such as those presented in this protocol. This naturally leads to questions about the validity of workload measures to infer safety. In this context, validity addresses the question of how appropriate a given workload measure is to answer questions about highway safety.

Criterion-related validity involves the extent to which a response measure is related to a criterion, i.e., a real-world performance of interest. Criterion-related validity is assessed by means of a correlation coefficient (called a validity coefficient). The validity coefficient is a measure of association computed between a response measure (e.g., lane exceedence) and some performance in the real world (e.g., crash incidence or occurrence). Unfortunately, the validity coefficient is unlikely to be a satisfactory approach to assessing the validity of the workload measures presented in this document as will be discussed below.

Validity coefficients between a measure, X, and a real-world performance, y, (designated by the symbol r_{xy}) can take on values ranging from $r = +l.00$ if two variables plot perfectly on a line with positive slope to $r = -1.00$ if two variables plot perfectly on a line with negative slope; $r = 0.00$ indicates no association between the two variables. The coefficient is affected by many factors, including the distributions of x and y. If both variables are standardized (i.e., each x or y value is subtracted from its mean and divided by its standard deviation so that the standardized scores each have a mean of zero and standard deviation of 1.0), then $r = +1.00$ only when each $z_x = z_y$ or $r = -1.00$ when each z_x or $-z_y$. Thus, the validity coefficient can take on its maximum values only when the shapes of the distributions are exactly the same (or exactly the opposite for

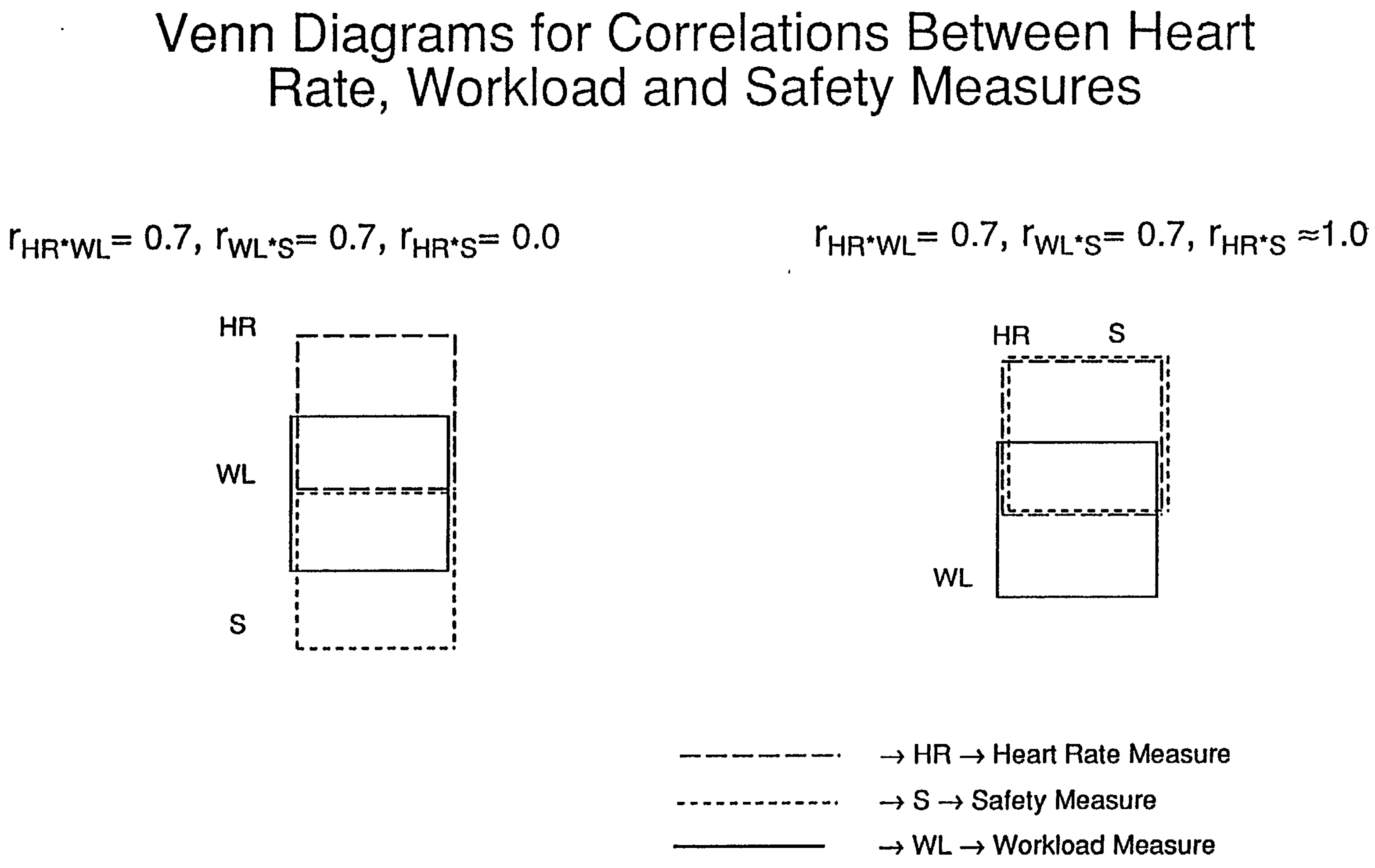


Figure 1-2. Venn Diagrams of Proportions of Variability between an arbitrary Human Factors Measure, Workload, and Safety.

r = -1.00) (Cohen and Cohen, 1983). The greater the departure from distribution similarity, the more severe the restriction. In particular, when the variables are distributed very differently, it is not possible to obtain a large correlation coefficient, including a validity coefficient.

Cohen and Cohen (1983) point out that the correlation between smoking and cancer is only about ***r*** = 0.1 and this is not a statistical artifact. Even though the risk of cancer is about 11 times as high for smokers as for non-smokers, the vast majority of both smokers and non-smokers alike will not contract lung cancer, and the non-association is low because of the non-association in these many cases. The same situation applies to workload measures and highway safety. Crashes are rare events and the population distribution for crash occurrence and non-occurrence is highly skewed in the direction of non-occurrence. On the other hand, phenomena measured in workload assessment, such as lane accedences, occur frequently yet are rarely associated with crashes. Thus, if one attempts to compute a validity coefficient between a workload measure like lane exceedence (LANEX) and, say, roadway departure crash incidence, the validity coefficient will be low or nonsignificant.

A well-established statistical theorem can also be used to show the difficulty in quantitatively establishing the safety relevance of a workload measure. Assume that one wishes to determine the probability of a roadway departure crash (here referred to simply as Crash) given a lane exceedence (here referred to as LANEX). Bayes' Theorem defines this probability as

$$P(Crash/LANEX) = \frac{P(LANEX/Crash)P(Crash)}{P(LANEX/Crash)P(Crash) + P(LANEX/NoCrash)P(NoCrash)}$$

where: P(Crash/LANEX) is the probability of a roadway departure crash given a lane exceedence
P(Crash) is the prior probability (i.e., the base rate or likelihood) of a roadway departure crash
P (LANEX/Crash) is the probability of a given lane exceedence given a roadway departure crash
P(LANEX/No Crash) is the probability of a lane exceedence and no roadway departure crash
P(No Crash) is the prior probability (i.e., the base rate or likelihood) of no crash.

Actual values for all of these terms are not known but plausible hypothetical values can be used to illustrate the point to be made. Let:

P(Crash/LANEX) be the value to be calculated.
P(Crash) = 0.0001 i.e., the base rate or probability of a crash is 1 chance out of a thousand)
P (LANEX/Crash) = 1 .O i.e., by definition, a roadway departure crash was associated with a lane exceedence.

P(LANEX/No Crash) = 0.9999 i.e., Assume 99.99 percent of the time lane exceedence is not associated with a crash.
P(No Crash)= 1 - P(Crash) = 1 - 0.0001 = 0.9999.

Substituting these hypothetical values in to the previous expression yields:

$$P(Crash/LANEX) = \frac{P(LANEX/Crash)P(Crash)}{P(LANEX/Crash)P(Crash) + P(LANEX/No\ Crash)P(No\ Crash)}$$

$$= \frac{(1.0)(0.0001)}{(1.0)(0.0001) + (0.9999)(0.9999)} = 0.0001$$

Essentially, the safety relevance of lane exceedence is calculated to be almost trivial even though physically it is perfectly associated with roadway departure crash incidence! Thus predictive validity coefficients computed between workload measures that can be used in ethical and safe workload evaluations will probably not be of use to establish safety relevance.

Construct validity is the extent to which a measure or test is associated with an abstract concept (like intelligence, motivation, or workload) that cannot be directly observed or measured but is purported to have relevance to real-world performance (job success, highway safety). As explained by Anastasi (1988), a construct is developed to explain and organize observed response consistencies. In the present case, such consistencies include workload-induced driver distraction as related to crash occurrence. Construct validation is determined by the gradual accumulation of information from a variety of sources. These sources included correlations with other measures, and experiments on the effects of certain manipulated or observed variables on particular measures. Evidence for construct validity of various workload measures is provided in each of the appendices of this document that present workload measurement categories. In summary, the safety relevance of workload measures cannot readily be demonstrated by means of traditional validity coefficients. Logical relations between measures and hypothetical constructs must be used instead, and these are derived from a model of driving.

1.6 A SIMPLE THEORY OF CRASHES

The prediction of number of crashes given in-cab device workload demand depends on at least four inputs. First, there must be an index of device-related workload itself, e.g., visual demand. Here the workload assessment protocol and measures provide the necessary indices. Second, there must be a frequency-of-use metric as well as an index of the number of such in-cab devices in the fleet. These metrics help determine the overall level of crash hazard exposure to which the drivers who use a given in-cab device will be put. Frequency-of-use is not part of the workload assessment process and the literature on device frequency-of-use is sparse (see Appendix E for a discussion of this point). However, other sources of frequency-of-use and technology infusion into the fleet may be obtained from a variety of sources over time after the technology has been introduced. Example sources of information might include the following:

- Cellular phone providers can tabulate statistics on the number of times specific vehicles made cellular phone calls;
- Fleet dispatchers can maintain a log of the number (and type) of text messages sent to drivers on the road; and
- Trade organizations can provide data on the numbers of a particular device (or class of device) sold in a given time period and region.

Third, there is also a need to capture information on how a technology is typically used, e.g., when, where, and by whom. These types of performance shaping factors can influence the crash likelihoods substantially. Finally, there is a need for more precise information in crash files that ‘indicates the type of causal factors (e.g., driver inattention, source of distraction, etc.), that can be used to pinpoint crashes that can plausibly be attributed to device workload rather than some other cause or contributing factor. If all such information were available upon which to build models, it may be possible to provide a quantitative estimate of crash incidence given further deployment of the technology or changes in the technology of interest.

Even if all the important factors mentioned above could be characterized, there will still be difficulty in predicting crash occurrence due to the chaotic nature of crashes. Battelle and its subcontractors recently completed a substantial effort to analyze the major types of crashes that occur in the United States (Tijerina, 1995). Analyses were conducted of rear-end crashes, roadway departure crashes, backing crashes, lane-change crashes, various types of intersection crashes, and opposite direction crashes. Based on detailed crash records, the report for each crash type identified putative causal factors and simple kinematic models of crash avoidance requirements. The reports generated from these analyses are intended to support development of crash avoidance systems.

Upon reflection, it appears that while certain causal factors may be attributed to crash incidence as general trends (e.g., driver inattention being a chief causal factor, and hence the motivation behind workload assessment), crash occurrence is in essence a chaotic process. The word ‘chaotic’ is used because the presence of chaos suggests that even if all variables in a non-linear system could be accounted for (the driver/vehicle/driving condition system), general patterns of system behavior (e.g., crash incidence) may be predicted but specific behaviors (e.g., crash occurrence) may not (Barton, 1994).

One general finding of the crash problem studies mentioned above (and other research as well), is that driver inattention is a key contributor to crashes on the highway. Crashes may indeed occur when the driver is not paying attention to the driving scene, but drivers who do not pay attention to their driving do not always have crashes. Crashes occur when a set of circumstances come together in space and time to jointly yield an unfortunate outcome. If drivers are rational within their situational understanding of the driving conditions and their motivations, it is plausible to assume that drivers involved in crashes were inattentive because they expected it to be acceptable to be momentarily inattentive and their expectations were violated by events at the time. If inattentiveness is risky, then other types of risk-taking (e.g., speeding, following too closely, inappropriate lookout) might also reflect expectancy violations or the mistaken belief that such behaviors will have no adverse outcomes. Given that no one is totally attentive to the driving task at all times, the chaotic nature of crash occurrence reinforces the meaning behind the phrase "But for the grace of God, there go I."

What does this theory of crashes have to do with the safety-relevance of workload assessment? Perhaps the best answer is that the possibility of drawing high associations between workload measures and the "ground truth" of highway safety (i.e., crashes) is small. Instead, the chaotic nature of crash occurrence may be taken to imply that new technology that takes the driver's eyes off the road and attention away from the primary task of driving produces a marginal increase in crash hazard exposure. That is, workload assessment can be used to show that one device increases or decreases marginal crash hazard exposure relative to some other in-vehicle transaction. Thus, workload assessment, as described in this document, is largely inferential and relativistic in nature.

1.7 SCIENTIFIC BASES FOR THE SAFETY RELEVANCE OF WORKLOAD MEASURES

Driver workload assessment is intended to uncover predictive evidence that the workload demand of a device may be high enough to degrade safety. This is a difficult problem, as a workshop on safety evaluations for Intelligent Transportation Systems (ITS) recently illustrated (Tijerina, Freedman, and Farber, in press; Dingus, 1995). Several scientific bases that might be used to relate workload assessment measures to safety include: theoretical constructs derived from a model or theory of driving (as was done in the section that dealt with this topic), archival analysis to relate crash incidence to different levels of a workload measure, and principles of physics. While attempts to link workload measures to safety are few, some work has been done in this area. Examples are provided below.

1.7.1 <u>Visual Allocation</u>

In the domain of visual allocation, a theory or model of driver performance indicates a basis for safety relevance. Because vision is the primary means of gathering information about the driving task, the driver cannot take eyes off the road scene for more than a moment without risking a crash. Other theories or models of driver behavior or performance are included in Appendices A through C, illustrating the use of theory as a scientific basis for the safety relevance of each class of measures.

Archival research that links high visual demand to crash incidence has recently been attempted by Wierwille and Tijerina (1994) and Wierwille and Tijetina (in press), through work conducted as part of this project. Wierwille and Tijerina (1994) describe the results of a study in which detailed police narratives from an accident database maintained by the Highway Safety Research Institute of North Carolina were reviewed for the presence of keywords potentially indicative of in-vehicle distraction. This approach had earlier been used by Perel(1976, 1988) to examine control incompatibilities in driver/vehicle systems. Based on a review of almost 18,000 records, results showed that numerous accidents are associated with visual allocation into the vehicle. Figure 1-3 presents some of the results of that study; it shows the number of crash cases from the North Carolina data base attributed to driver attentional diversion or workload, further subdivided into interior (in-vehicle) sources of distraction and dash/console/steering column distraction sources.

Subsequently, Wierwille (see Appendix E) developed a quantitative relationship between in-vehicle visual demand (weighted by in-vehicle device use) and crash incidence for those crashes identified in the earlier research. In order to accomplish this, estimates of visual demand and frequency of device use were needed as predictor variables in a regression model that had crash incidence or number of crashes from the previous analysis as the criterion variable. The approach taken was to use data in the human factors literature to develop estimates of the frequency of use of selected in-vehicle devices (e.g., radio, speedometer, windshield wiper, etc.) and estimates of the visual demand of those same devices.

Appendix E presents the entire set of analyses used. The human factors literature was used to identify visual demand data for similar in-vehicle devices. From these, mean glance duration and average number of glances required to service various in-vehicle devices were collated for use in the present analysis. For a given device use (e.g., radio tuning), visual demand was estimated to be the product of mean glance duration and mean number of glances (See Appendix A for definitions).

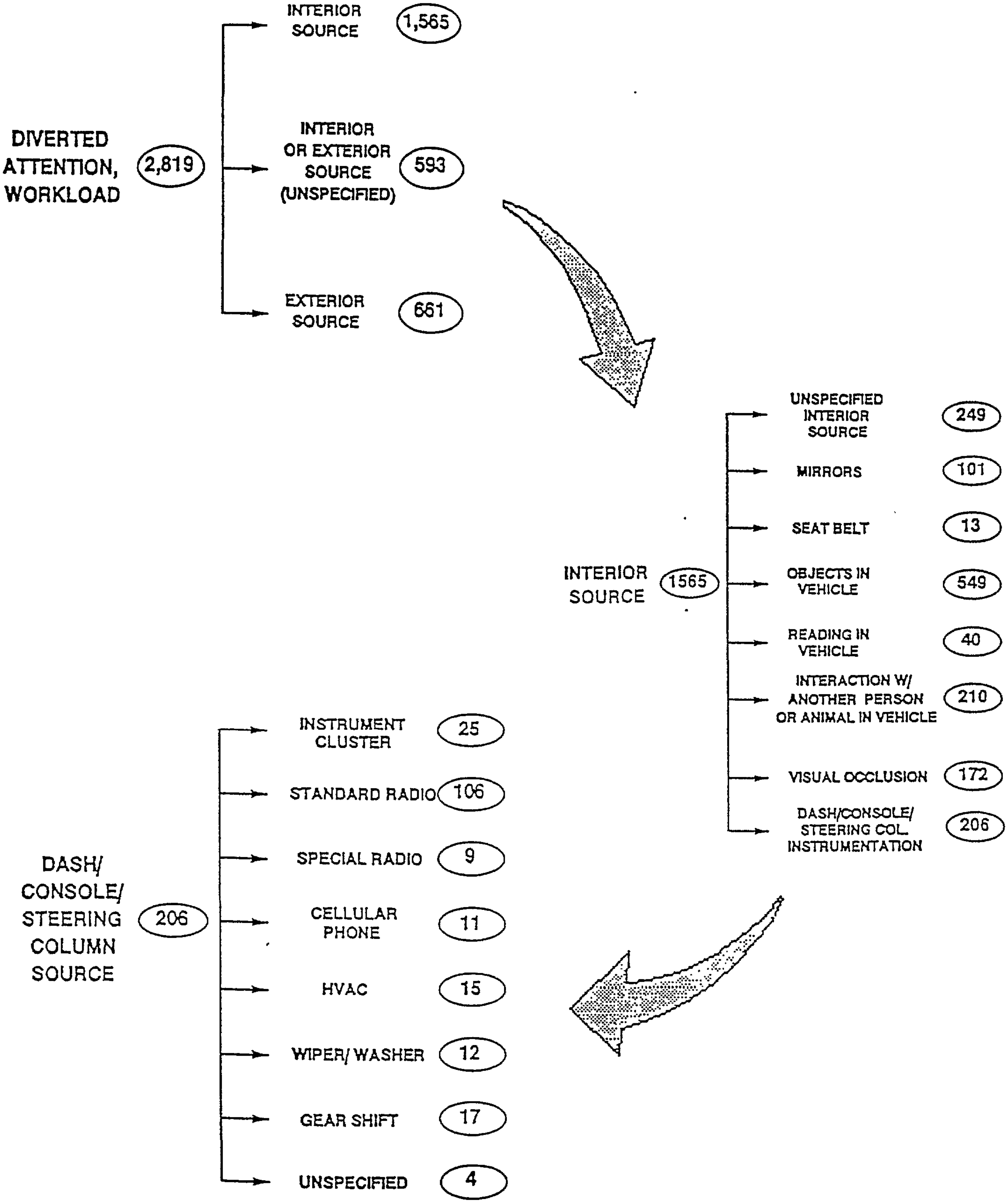


Figure 1-3. Number of Crashes Distributed by Sources of Attentional Distraction, Broken Down into Interior Source and Dash/Console/Steering Column Instrumentation Group (Source: Wierwille and Tijerina,1994).

The principal approach taken to estimate device frequency-of-use was a logical approach with engineering judgement applied when necessary to develop the necessary relative-use values. This was a difficult endeavor given the limited data available and engineering judgement was needed to arrive at a metric for frequency-of-use for several types of in-vehicle transactions.

Exposure was defined as a function of visual demand (the product of mean glance duration and mean glance frequency per in-vehicle device use) and device frequency-of-use (scaled to a common time frame, e.g., uses per week). Three types of exposure were computed for device j:

Type 1 Exposure j = $(\text{mean glance duration}_j) \times (\text{mean glance frequency}_j) \times (\text{use frequency}_j)$
Type 2 Exposure j = $(\text{mean glance duration}_j)^{3/2} \times (\text{mean glance frequency}_j) \times (\text{use frequency}_j)$
Type 3 Exposure j = $(\text{mean glance duration}_j)^2 \times (\text{mean glance frequency}_j)(\text{use frequency}_j)$

Type 2 and Type 3 exposure weight longer glance durations more heavily under the assumption that longer single glance durations increase crash hazard exposure more than might be implied by a linear increase. Appendix E includes the results of regression analyses using exposure as the predictor variable and crash incidence from Wierwille and Tijerina (1994). In general, the regression fits are excellent regardless of the exposure type, with correlations ranging from 0.898 to 0.982. This study is a unique attempt to use actuarial data to relate visual allocation measures to crash incidence.

1.7.2 Lanekeeping

It is self-evident that the driver must control the vehicle and remain in the travel lane, moving from it only in a controlled fashion. Failure to properly keep in one's lane is the proximal physical event that leads to such crash types as lane change crashes (Chovan, Tijerina, Alexander, and Hendricks, 1994), opposite-direction crashes (Chovan, Everson, Hendricks, and Pierowicz, 1994) and single vehicle roadway departures (Hendricks, Allen, Tijerina, Everson, Knipling, and Wilson, 1992; Mironer and Hendricks, 1994). Thus, measures of lanekeeping such as lane exceedences are directly safety-relevant. Furthermore, increases in lane position variability or mean lane position may also be interpreted as safety-relevant to the extent that driving closer to a lane line reduces the driver's margin for recovery in the event of an emergency, all else being equal. The same principles apply to measures such as Time-To-Line Crossing (TLC) (Godthelp, 1984), Time-to-Trajectory Divergence (TTD), and other measures related to lanekeeping performance. Thus, in addition to the logical relations contained in the simple model of driving presented earlier, there are archival principles of vehicle control and archival relationships to recommend such measures for a workload assessment protocol.

1.7.3 Speed Measures

It is not uncommon for a driver under high workload to slow the vehicle. However, there is evidence that crash incidence rises with speed variability. Cirillo (1968) used crash data to show that the driver traveling closer to the average speed of the travel stream has a lower crash risk than the driver traveling at higher or lower average speeds. More recently, Garber and Gadiraju (1989) presented regression models relating crash rate to speed variance. These plots indicate that a vehicle traveling significantly slower or faster than the prevailing travel speed (regardless of posted speed), may be more likely to be involved in a crash.

In addition to archival data supporting the safety relevance of speed variation measures, mean speed measures are also important. Nilsson (1990, as cited in Evans, 1991) examined changes in crashes and casualties associated with speed limit changes and derived quantitative prediction models that correspond well to accident statistics in the U. S. when speed limits were increased from 55 mph to 65 mph in 1987. More recently, Hendricks, et al. (1992) and Mironer and Hendricks (1994) have determined that a substantial number of roadway departure crashes at curves are associated with excessive speed with respect to roadway geometry and traction. Should a distracted driver allow speed to creep up during in-vehicle device use, this increase may be a potential safety threat.

1.7.4 Time Headway

Rear-end crashes are the single most common crash type in the United States (Tijerina, 1995). The vast majority of these crashes involve driver inattention and/or following too closely. Furthermore, Evans and Wasielewski (1982) showed that time headway adopted on a section of highway was a significant discriminator of traffic violators from non-violators, that time headways were often below 1.0 s for the traffic violators, and that such short headways greatly increase the risk of rear-end crashes. Thus, there is at least some archival evidence that car following measures such as time headway have safety relevance. Principles of physics also can be used to relate close car following to crash involvement (Knipling et al., 1993).

This section is short but illustrates the types of scientific information available to relate the workload measures in this document to safety. Additional research is needed to develop such relationships further. Application of this workload assessment protocol should contribute to such developments.

1.8 ORGANIZATION OF THE PROTOCOL DOCUMENT

This document provides guidance in the organization, planning, and execution of a device evaluation from a driver-centered perspective. Figure 1-4, based in part on Williges (1992) and

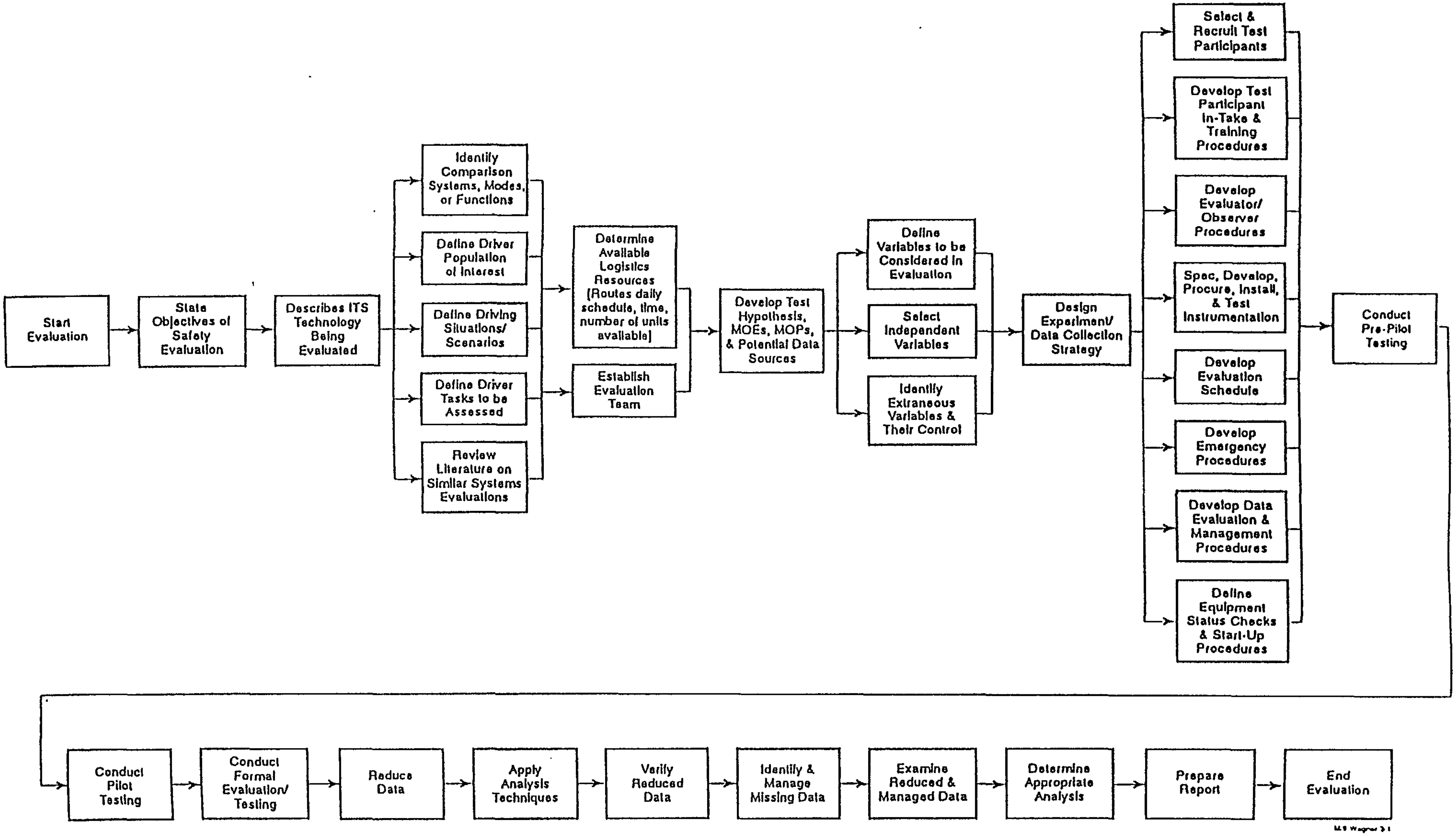


Figure 1-4. Flow Diagram of Device Assessment Process From Driver Workload Perspective.

Unisys (1987), presents a flow diagram of the overall workload assessment process. Chapter 2.0 of this document covers the design and planning stages. Chapter 3 .O addresses the conduct, analysis, and interpretation stages, Chapter 4.0 contains discussion of outstanding issues that face workload assessment. Chapter 5.0 provides an epilogue. Finally, there are number of appendices that provide additional information and guidance on the execution of a device evaluation.

As mentioned earlier, this document should be of use to several types of users:

- Managers who need to organize a workload evaluation;
- Technical staff who must implement the workload evaluation;
- Researchers who wish to adopt a consistent workload measurement system and add to that system by contributing additional findings to those presented here.

Government representatives who wish to provide contractors with a guidance document to conduct high-technology device or ITS safety evaluations.

This protocol document is intended to address many different types of devices and a broad range of ITS products. For this reason, there can be no single assessment that applies to all possible cases. The details of a safety-relevant workload assessment for a route guidance system will be different than that for a voice communication system, which in turn will be different for that carried out on a crash avoidance system (CAS). Thus, the guidance is genera1 in tone. However, when possible, recommendations are made on which of several alternatives might be of genera1 usefulness for device or system evaluations.

REFERENCES

Anastasi, A. (1988). *Psychological testing* (Sixth Edition). New York: McMillan.

Barton, S. (1994). Chaos, self-organization, and psychology. *American Psychologist, 49,* 5-14.

Chapanis, A. (1970). Plenary discussion: Relevance of physiological and psychological criteria to man-machine systems: The present state of the art. *Ergonomics, 13(2), 337-346.*

Chovan, J.D., Everson, J.J., Hendricks, D.L., & Pierowicz, J. (1994, June). *Analysis of opposite-direction crashes* (Contract No. DTRS-57-89-D-00086). Columbus, OH: Battelle.

Chovan, J.D., Tijerina, L., Alexander, G. & Hendricks, D.L. (1994, March). *Examination of lane* change *crashes and potential* IVHS *countermeasures* (Report No. DOT HS 808 071 and DOT-VNTSC-NHTSA-93-2). Cambridge, MA: John A. Volpe National Transportation Systems Center.

Cirillo, J.A. (1968). Interstate system accident research study II, interim report II. *Public Roads,* 35, 71-75.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (Second Edition). Hillsdale, NJ: Lawrence Erlbaum Associates.

Dingus, T.A. (1995). Moving from measures of performance (MOPS) to measures of effectiveness (MOEs) in the safety evaluation of ITS products or demonstrations. In D. Nelson (Ed.), *Proceedings of ITS Safety Evaluation Workshop.* Washington, DC: ITS America.

Evans, L. (1991). *Traffic safety and the driver.* New York: Van Nostrand Reinhold.

Evans, L., Wasielewski, P. (1982). Do accident involved drivers exhibit riskier everyday driving behavior? *Accident Analysis and Prevention, 14, 57-64.*

Garber, N.J. & Gadiraju, R. (1989). Factors affecting speed variance and its influence on accidents. *Transportation Research Record, 1213,* 64-71.

Godthelp, H. (1984). *Studies on human vehicle control.* Soesterberg, The Netherlands: The Institute for Perception.

Hendricks, D., Allen, J., Tijerina, L., Everson, J., Knipling, R., Wilson, C. (July, 1992). WSC IVHS *program* ***Topical Report #2: Single vehicle roadway departures*** (Volumes I and II) (Contract No. DTRS-57-89-D-00086). Columbus, OH: Battelle

Knipling, R.R., Mironer, M., Hendricks, D.L., Tijerina, L., Everson, J., Allen, J.C., & Wilson, C. (1993, May). ***Assessment*** *of* ***(IVHS countermeasures for collision avoidance: Rear-end crashes*** (Report No. DOT HS 807 995). Washington, DC: U.S. Department of Transportation National Highway Traffic Safety Administration.

Hess, R.A. (1987). Feedback control models. In G. Salvendy (Ed.), *Handbook of human factors* (pp. 1212-1241). New York: Wiley.

Meister, D. (1986). *Human factors testing and evaluation.* Amsterdam: Elsevier.

Mironer, M., & Hendricks, D.L. (1994, August). ***Examination of single vehicle roadway departure crashes and potential*** IVHS ***countermeasures*** (Report No. DOT HS 808 144 and DOT-VNTSC-NHTSA-94-3). Cambridge, MA: U.S. Department of Transportation John A. Volpe National Transportation Systems Center.

Nilsson, G. (1990). The effect of speed limits on traffic accidents in Sweden. *VT.‘. Report No. 68* (pp. l-10). Linkoping, Sweden: National Road and Traffic Research Institute.

Older, J.S., Spicer, B.R. (1976). Traffic conflicts - a development in accident research. *Human Factors, I8(4), 335-350.*

Perel, M. (1976). *Analyzing the role of driver/vehicle incompatibilities in accident causation usingpolice reports* (Report No. DOT HS 801 858). Washington, DC: National Highway Traffic Safety Administration.

Perel, M. (1988). *An analysis of pedal-related accidents using police narratives.* Washington, DC: Office of Crash Avoidance Research, National Highway Traffic Safety Administration.

Sussman, E.D., Bishop, H., Madnick, B., & Walter, R. (1985). Driver inattention and highway *safety. Transportation Research Record, 40-48.*

Tijerina, L. (1995). *Key human factors research needs in IVHS crash avoidance* (Paper No. 950250). Paper presented at the Transportation Research Board 74th Annual Meeting, January 22-28, 1995, Washington, DC. *(Also Transportation Research Record,* in press).

Tijerina, L., Freedman, M., & Farber, E. (in press). *A synopsis of the ITS AMERICA/NHTSA Workshop on ITS Safety Evaluations.* Paper to be presented at the Second World Congress on Intelligent Transportation Systems, Yokohama, Japan, November 9-l 1, 1995.

Treat, J.R., Tumbas, McDonald, ST., Shinar, D., Hume, R.D., Mayer, R.E., Stansifer, R.L., Castellan, N.J. (1977). *Trilevel study of the causes of traffic accidents: Executive summary.* (Report No. DOT HS 034 3 535). Washington, DC: U.S. Department of Transportation.

Unisys (1987, July). *User-assisted test and evaluation methodology assistance program (TEMAP): Version* I (Contract No. F33615-85-C-0533). Wright Patterson AFB, OH: Crew Systems Ergonomics Information Analysis Center (CSERIAC).

Wierwille, W., W., Tijerina, L., Kiger, S. M., Rockwell, T. H., Lauber, E., & Bittner, A. (September, 1992). *NHTSA heavy vehicle driver workload assessment Task 4 Interim Report= Review of workload and related research and a theoretical basis for safety relevance of workload measures* (Contract No. DTHN22-9 1 -C-07003). Columbus, OH: Battelle.

Wierwille, W. W., & Tijerina, L. (1994). An analysis of driving accident narratives as a means of determining problems caused by in-vehicle visual allocation and visual workload. In A. G. Gale et al. (Eds.), *Vision in Vehicles V.* Amsterdam: North-Holland.

Wierwille, W. W., & Tijerina, (in press). *Modeling the relationship between driver In-vehicle visual demands and accident occurrence.* Paper to be presented at the Sixth International Conference on Vision in Vehicles, Derby, England, September 12 - 15, 1995.

Williams, M.J. (198 1). Validity of the traffic conflicts technique. *Accident Analysis & Prevention, 13,* 133-145.

Williges, R. (1992). Experimental Design. R.W. Pew, P. Green (Eds.), *Human Factors Engineering Short Course* Notes (33rd Ed.). Ann Arbor, MI: The University of Michigan, Chrysler Center for Continuing Engineering Education.

2.0 WORKLOAD ASSESSMENT APPROACH DEVELOPMENT

2.1 DETERMINE THE PURPOSE OF THE WORKLOAD EVALUATION

The motivation behind a driver workload assessment of an in-cab device or system is this:

> Systems are being developed and marketed that provide many potentially useful functions to the driver while on the road. The concern is that any given in-cab system may introduce subsidiary tasks that may compete with the driver's primary task of safely controlling the vehicle.

From this motivation, the purpose of the driver workload evaluation is stated succinctly:

> The objective driver workload assessment for an in-vehicle system or device is to empirically assess the potential of that device to distract the driver from the driving task to the extent that safety may be compromised.

Workload assessment must be conducted within the context of in-vehicle system characteristics, driver characteristics, and driving condition characteristics, using measurable variables that are putatively related to highway safety, under the constraints of what resources are available to complete the assessment. Current knowledge of what causes highway crashes is currently poorly understood. Thus, the relationship between measured variables and safety is currently poorly understood. However, the protocol presented here is a step toward cataloguing the state-of-the-art in driver workload measurement in terms of variables that are logically related to driving safety.

2.1.1 <u>State the Objectives of the Workload Evaluation</u>

The purpose of driver workload assessment is to assess the instrusiveness of in-cab device or system use on the driving task. From this purpose and a theory of how driver workload might be manifested, three broad evaluation objectives may be pursued and, from them, hypotheses may be generated and addressed empirically. These three objectives are given below:

- Objective 1: Answer the question "Do driver behaviors with an in-cab device differ significantly from one or more comparison conditions?" Comparison conditions may be other device modes (e.g., map mode or auditory mode for a route guidance system), functions (data display vs. error correction), manual or paper analogues (e.g., paper map compared to an electronic route guidance system), or commonly accepted device uses (e.g., use of instrument panel devices or open road driving). Driver behaviors can be characterized, minimally, as visual allocation, manual activity (i.e., hand activity) required of the system or device being evaluated, and inputs to steering and pedal controls.

- Objective 2: Answer the question "Does driver-vehicle performance with in-cab device use significantly differ from a comparison condition?" Examples of driver-vehicle performance include lateral control (e.g., control of lane position and heading), longitudinal control (e.g., speed maintenance, braking), minimally. Object and event detection and wayfinding are other categories of driver-vehicle performance that may be included in an evaluation.

- Objective 3 : Answer the question "What are driver attitudes about the in-cab device?" That is, how do drivers subjectively react to a device in terms of the subjective workload experienced, the functionality provided, and the perceived safety of the device under varied driving conditions?

High technology in-cab device transactions can introduce subsidiary tasks that may compete with the primary task of safely controlling the vehicle at all times. Alternatively, the device may ease the driver's workload by, say, providing safety-critical information in a timely manner. Driver-vehicle performance in terms of lateral control, longitudinal control, and object and event detection, have *prima facie safety* relevance. For this reason, a comparison of driver-vehicle performance while the driver interacts with the device to other driving circumstances when the device is not used is an important aspect of the safety evaluation.

Safety relevant driver behaviors include visual allocation, manual activity, and directed attention. The driver's eyes cannot be taken off the road scene for more than a moment before highway safety is affected, yet almost any in-cab device with a visual display will demand some visual allocation. Similarly, manual resources that might be used to control the vehicle must be shared with the in-cab device controls (as well as other instrument panel devices in the vehicle). It is also possible that biomechanical interference effects arise while the driver attempts to manipulate a control. One example of this might be inadvertent force applied to the steering wheel while reaching over to operate a device's controls; this could lead to a lane departure. Finally, the driver may devote attention to the in-cab device directly, in which case visual allocation provides insight into driver attention. Alternatively, the driver may devote attention to thinking about information provided by a device after having picked it up from the displays. These effects are the focus of the second question. All such effects may be assessed within the context of driver behavior while using common in-vehicle devices such as radios, paper maps, and instrument panel devices.

The third question focuses on driver acceptance and attitudes toward the in-cab device. Driver acceptance of a system has important safety implications. Acceptance determines the frequency with which system functions will be used (and hence, the facility which the driver develops with that function). It also determines what features or functions will be used: A function that is seldom used may either have no impact on highway safety or negatively impact highway safety because it is always novel to the driver. Acceptance also determines how the driver will interact with the technology, ranging from slow and attention consuming search processes to well learned, almost automatic routines.

A fourth broad question that arguably might be included in a device workload assessment might be to consider the usability of the device, i.e., how legible its displays are, how easy it is to make manual inputs, the directness of error recovery, etc. This is not included explicitly in the objectives of this safety-relevant field evaluation but should be conducted in preparation for a field evaluation of workload. A great deal of useful information about usability problems may be uncovered (and subsequently corrected) in a usability review by means of checklists, structured walk&roughs, or perhaps iterative testing with a low-cost desktop driving simulator (e.g., a driving video game). A system that facilitates fast and accurate driver interaction is considered ideal for workload reduction. Usability problems may also direct the nature of the workload assessment, and thus focus the assessment to particular functions, modes, driving conditions, and perhaps even types of drivers. The usability of the device will be reflected in the answers to the three questions listed above and the answers will provide a better indication of the safety impacts of the device's usability (or lack thereof). In practice, the evaluator must do this evaluation first. It is not an efficient use of resources to take a "poor" device into an on-the-road evaluation if its flaws are already apparent.

2.2 DEFINE THE SYSTEM TO BE EVALUATED AND REFINE THE ASSESSMENT PROBLEM

In order to define the issues and questions of concern that the workload evaluation must address in detail, there is a need to describe the in-vehicle system to be evaluated, determine comparison conditions, driving scenarios, driver population, and driver tasks of interest. Each of these are discussed below.

2.2.1 <u>Describe the System or Systems to Be Assessed</u>

What is the system, device, or product being evaluated? How does it operate? What are its functions, features, and modes of operation that are to be evaluated? A key aspect of any product or device evaluation is a thorough understanding of that device's structure and function. This understanding requires access to and review of documentation such as a user's manual, the human interface design specification, an operator task inventory or task analysis, states and modes diagrams, failure modes and effects analysis (FMEA) reports, and theory of operations documents. The output of this effort should be an assessment of the visual, manual, auditory, and cognitive demand on the driver from this system. In addition, it can be invaluable for an evaluation team member to interview and learn more about the system from its designers or others knowledgeable about the system or device. If at all possible, the evaluation team members should have an opportunity to learn about and use the system directly.

2.2.2 Identify Comparison Systems, Modes or Functions to Be Assessed

The essence of evaluation is comparison against a baseline and/or alternative configurations or conditions. Therefore is it is important to ask what are the comparisons that are reasonable and important to make in the evaluation? These might include different modes of driver display, different device functions, comparison between automated and manual operation, comparison between new device demand and the demand posed by commonly used instrument panel tasks (e.g., manually tuning a radio), error modes and the demand posed by recovery procedures, and so on.

2.2.3 Define the Situational Characteristics That Are Relevant to the Workload Evaluation

What are the contexts in which the device (or comparison systems) may be used? Typically, this may be characterized by answering what are the driving tasks (e.g., backing, intersection negotiation, lane changes, open road driving) and driving conditions (e.g., lighting, traffic density, divided vs. undivided highway, reduced visibility, reduced traction) in which the system may be used.

2.2.4 Define the Relevant Driver Population

What is the user population for the system? Should the evaluation include, exclude, or sample a range of truck drivers, passenger car drivers, older drivers (55 years or more) , younger drivers (25 years or less), males versus females, inexperienced or experienced drivers, or drivers with certain abilities (or lack of certain abilities)?

2.2.5 Define the Driver's Tasks to Be Assessed in the Workload Assessment.

Based on the description of the m-cab system or device to be assessed and the driving conditions under which the system may be used, driver tasks that are to be evaluated in the workload assessment must be determined. It will be important to understand the various transactions that can be accomplished with the device and the contexts in which these transactions might arise. Inspection of the physical interface characteristics may provide an early indication of workload-inducing properties (e.g., visual display washout under high incident illumination, inadvertent control activation when gloves are worn, etc.). This background work allows the experimenter to determine what in-cab tasks should be part of the data collection session, under what scenarios those tasks might be observed, and what types of problems should be closely monitored.

2.3 REVIEW RELEVANT LITERATURE

In preparing a workload assessment, it is valuable to review relevant literature to determine what is known in the field, what techniques are in use, new methods for workload assessment, and critiques of measures and methods that have been used to date. The series of interim reports generated for the NHTSA project under which this protocol document was developed provide useful information and references.

2.4 DETERMINE AVAILABLE AND REQUIRED RESOURCES

In preparing the workload evaluation plan, it is important to consider what resources are required and what logistical constraints exist. These resources can be characterized as logistical resources and evaluation team resources. Each category of resource is discussed below.

2.4.1 Logistical Resources

Logistical resources address what is actually needed for the evaluation. These include test subjects, routing options, equipment, and time. The availability of specialized subject pools must be factored into test design and schedule. For example, the availability of professional heavy vehicle drivers for participation in a device workload evaluation may sometimes be severely limited due to factors such as the volunteer's driving schedule, selection criteria such as number of moving traffic violations received within the last three years, and special characteristics like age, gender, or experience.

A second resource that must be carefully considered is the route and data collection session schedule. To the extent that driving condition variables (e.g., lighting, road type, traffic density) will be explicitly manipulated in the study, these must also be carefully factored into the test design. For example, conducting an assessment that considers driving conditions like lighting and traffic density may prove difficult because of variations in lighting with the seasons, and changes in local (or test site) traffic patterns.

A third resource is the equipment available for the workload assessment . There may be only a single instrumented vehicle that can be used for data collection. There may be only a limited number of prototype in-cab systems or devices that are available for test purposes. There may be limitations in the data collection equipment that make certain types of measurements infeasible. There may or may not be redundant systems that can be used for data collection to improve data reliability or integrity.

A fourth resource is time. The workload assessment will have to be completed within some planning horizon. This planned schedule should ideally include an opportunity to accommodate unexpected delays due to such factors as vehicle or equipment breakdowns, union strikes, and inclement weather. Any or all of these types of unforeseeable events can significantly delay timely completion of an evaluation.

2.4.2 Establish the Workload Evaluation Team

The workload assessment team consists of the following roles, which may or may not be carried out by the same person or persons. The assessment team administrator is responsible for specification of test objectives, development of the test protocol, selection of independent and dependent variables, and the overall conduct of the test. The administrator is also ultimately responsible for report preparation and development of recommendations from the assessment. The team manager is responsible for overall coordination of the testing, review of test materials and methods, management of resources, and crisis management. A team engineer is responsible for development and specification of the hardware and software required for the assessment. This includes power supplies and conditioning, sensors and sampling rates, data acquisition and storage, time code generation, multiplexing, and so on. A technician is responsible for equipment installation, calibration, repair, and replacement. More than one technician may be needed for a given assessment. One or more experimenters or observers may be needed to initialize the systems, collect test participant biographical data, administer screening tests, secure informed consent, carry out the assessment protocol, manage the data collection equipment and prototypes as needed, serve as the tactical trouble-shooter while the test is under way, conduct test participant debriefs, administer payment and collect receipt of payment forms, as needed, and accomplish all housekeeping functions like marking the diskettes and video tapes for the date, time, and conditions of the test. A data reducer (more than one may be needed) is for receiving the data collected in written, audio, video, and magnetic media, cataloguing, and storing that data appropriately. The data reducer filters the digital data as appropriate, parsing the critical segments of the data stream for detailed analysis, and deriving measures of performance from the appropriate segments of the filtered data stream. Data reducers are responsible for video data reduction, e.g., frame-by-frame review of glance direction and duration. Data reducers may also be assigned the responsibility of collating summary demographic data or tabulating verbal responses to questions. Data reducers also extract test participant responses to written questionnaires and ensure that outputs from the data reduction phase are data, in an appropriate form, suitable for analysis. A data analyst is responsible for conducting graphical, descriptive, or inferential statistical analysis. The goal of this analysis is to answer specific questions regarding the independent variables and their effects on measured responses (i.e., dependent variables). The data analyst also works with other members of the assessment team to interpret the results of the analysis. A secretary is responsible for detailed scheduling of test participants, follow-up reminders, mailings of preliminary briefing materials (as appropriate), and support for report preparation. Note that the manager, engineer, experimenter, and analyst need experience in vehicle dynamics and driver performance.

3.0 WORKLOAD ASSESSMENT DETAILED EVALUATION PLAN

3.1 DEVELOP WORKLOAD ASSESSMENT TEST HYPOTHESES, MEASURES OF EFFECTIVENESS, MEASURES OF PERFORMANCE, AND DATA SOURCES

Burgett (1994) has recently outlined the steps used in the safety evaluation for the TravTek program in Orlando, Florida. The safety evaluation was arranged to move from stated objectives to analysis of relevant data in an orderly manner. Table 3-l presents the approach taken in the TravTek program. This same tabular approach provides a convenient way to summarize a driver workload assessment for a particular in-cab system or device.

The Burgett template shows a means to methodically move from stated objectives of a safety assessment to analysis of relevant data for sub-element within the template. Table 3-l provides an indication of the objective to be met by the evaluation, the hypothesis (or hypotheses) that are generated by the objective, Measures of Effectiveness (MOEs) that are theoretical constructs that are presumed to affect driving safety (Dingus, 1995). Measures of Performance (MOPs) are operationally-defined measured response variables that are presumed to have an impact on the theoretical construct of interest.

Table 3-2, Table 3-3, and Table 3-4, address the first, second, and third generic objectives of workload assessment. They indicate what hypotheses might be appropriate, what MOEs and MOPs should be considered, means of collecting such data, and general guidance on the nature of suitable data analysis for each. Each of these tables includes notes about the expected interpretations to be placed on each of the measures of performance. Note that interpretation of measures of performance is largely investigative. That is, exploratory data analysis will be required to determine if and in what ways the measure provides indications of safety-relevant in-cab device workload.

3.2 DEFINE THE VARIABLES TO BE CONSIDERED IN THE WORKLOAD EVALUATION

Indeuendent variables or factors are variables that may impact the workload assessment and can be manipulated by the evaluation team or fixed for the study (e.g., vehicle cab layout). Dependent variables are response measures that reflect the influences of independent variables if there is a statistically reliable relationship between the two. Extraneous variables (also called nuisance variables) are variables that can influence the outcomes of an evaluation but are not a part of *the* study *per se* and so may make interpretation of the results difficult or impossible. Selection of appropriate independent variables and dependent variables is based on the background research that leads to a description of the in-vehicle system, driving scenarios, driver population, driver tasks, etc., along with a review of relevant literature to learn what has been used before and what has been discovered. Control of extraneous variables also depends on much the same body of knowledge.

Table 3- 1. TravTek Safety Evaluation Study Definition.

OBJECTIVE	HYPOTHESIS	MEASURES OF EFFECTIVENESS	MEASURES OF PERFORMANCE	DATA SOURCES	METHODS OF ANALYSIS
Objectives are stated in terms of what to measure or what to evaluate.	These include a statement of the primary hypothesis.	Measures of Effectiveness (MOEs) are conceptual measures that convey "goodness" or ability to meet a set of criteria	Measures of Performance (MOPS) are data elements required to satisfy the MOEs (i.e., the variables needed to compute the MOEs).	This column refers to the various sources of data (e.g., sensors, video, observer logs) required to the MOPS.	This column broadly defines the types of analytical procedures that will be used.

Table 3-2. Does safety-relevant driver-vehicle performance with in-cab device use significantly differ from a comparison condition?

OBJECTIVE	**HYPOTHESIS**	**MEASURES OF EFFECTIVENESS**	**MEASURES OF PERFORMANCE**	**DATA SOURCES**	**METHODS OF ANALYSIS**
Assess the intrusion of in-cab device use on the driving task in comparison with selected alternatives.	Driving performance will vary dependent on in-cab device use and selected comparison alternatives.	Driver-vehicle performance.	Mean speed Speed variance Mean lane position Lane variance # of unplanned lane exceedences Duration of unplanned lane exceedences Abrupt lateral accelerations Abrupt longitudinal accelerations Following time headway Minimum following distance Peak closing velocity Yaw standard deviation Yaw deviation mean Minimum miss distance (for near-misses) Peak closing velocity (for near-misses)	5th wheel Lane tracker Lane tracker Lane tracker, Road scene video Lateral accelerometer Longitudinal accelerometer Laser Headway Detector Laser Headway Detector Laser Headway Detector Yaw/Yaw rate accelerometer Yaw/Yaw rate accelerometer Video Video	Inferential Statistics - t-test - ANOVA - Regression - MANOVA

Notes: In general, for any measure of performance, scaling can be made such as "more is worse." That is, greater magnitudes imply more degraded driver-vehicle performance. For near-misses, smaller minimum miss distances are worse and greater peak closing velocities are worse. Near-miss measures will, of course, be happenstance.

Table 3-3. Do driver behaviors with an in-cab device differ significantly from one or more comparison conditions?

OBJECTIVE	HYPOTHESIS	MEASURES OF EFFECTIVENESS	MEASURES OF PERFORMANCE	**DATA SOURCES**	METHODS OF **ANALYSIS**
Assess the intrusion of in-cab device use on the driving task in comparison with selected alternatives.	Driver in-vehicle behaviors will vary depending on in-cab system use and selected alternatives.	In-vehicle Driver Behavior	Glance duration Glance frequency Glance distribution Steering standard deviation Peak steering deflection Steering velocity mean Steering velocity variance Steering holds Steering hold duration Steering reversals Zero-crossings Steering response time Braking applications Mean break application time Accelerator pedal reversals Accelerator variance Accelerator holds Accelerator releases Brake RT (to near-miss) Break application foot pressure Total hands-on-wheel time Hand-off-wheel occurrences	Video Video Video String pot String pot String pot String pot String pot String pot String pot String pot String pot String pot Brake light switch Brake light switch Accelerator pedal switch Accelerator pedal switch Accelerator pedal switch Accelerator pedal switch Break light switch/road scene video Pedal pressure transducer Video Video	Inferential Statistics - t-test - ANOVA - MANOVA - Regression - Chi-square

Table 3-3 (Continued)

Notes: Longer or more frequent glances away from the road scene are considered indicative of in-vehicle intrusion. Steering measures vary on a case-by-case basis but are intended to capture intermittent open-loop lateral control by the driver while engaged in-vehicle transactions. Brake applications, high brake pressure, long RTs (to near-misses as judged by video) are also indicative of intermittent open-loop driving. Accelerator pedal reversal patterns for driver-vehicle performance assessment are investigative at this point. However, they may reflect driver workload management strategies during in-vehicle device use. Hands-on-wheel time is expected to be less with in-vehicle device use than during normal driving. Hand-off-wheel occurrences may be highly correlated with in-vehicle visual glances, indicating the presence of visually guided movements to in-vehicle device controls.

Table 3-4. What are driver attitudes about the in-cab device?

OBJECTIVE	HYPOTHESIS	MEASURES OF EFFECTIVENESS	MEASURES OF PERFORMANCE	DATA SOURCES	METHODS OF ANALYSIS
Assess the intrusion of in-cab device use on the driving task in comparison with selected alternatives.	Acceptance and attitudes will vary by in-cab device and selected alternatives for comparison.	Subjective Assessments	Subjective Workload Scales Driver acceptance ratings Driver debrief comments In-cab function frequency-of- use In-cab device function error rate In-cab device function completion time	SWAT, TLX, MCH Likert-scale ratings Experimenter notes Observer log, video Observer log, video Observer log, video	Inferential Statistics - parametrics - non-parametrics

Notes: If workload measures show greater workload with in-cab device use, this might be correlated with reduced driver acceptance and negative attitudes. Driver attitudes are expected to differ with different functions; negative attitudes may indicate a safety-relevant problem that could, in principle, be addressed in redesign. Differences in frequency of use may be correlated with greater error rates or longer transaction completion times, all of which may pose a distraction to the driver and decrease driver acceptance.

3.2.1 Select Independent Variables

Selection of the set of independent variables to include in a workload assessment is difficult. There are many factors that may influence the workload a driver experiences while engaged with an in-vehicle device or system. Ideally, only the subset of all possible independent variables is included that is likely to substantially affect driver workload measures in meaningful ways. If this subset is not determined, the number of independent variables can become so large that the evaluation becomes too cumbersome or impossible to execute. In general, expert judgement of what might count is invaluable in selecting the independent variables to be included in a workload assessment. The expert can use knowledge of previous research, operational conditions, and the research literature to guide the selection of what is important to manipulate and what is not important to manipulate. The expert should also guard against including combinations of independent variables that do not occur under normal (or perhaps even abnormal) conditions. Ultimately, the generalizability of the workload assessment will depend, at least in part, on the judicious selection of independent variables and levels thereof.

Table 3-5 provides a listing of independent variables that are potentially relevant for driver workload assessment. These include driver variables, in-cab device variables, and driving condition variables such as traffic density, roadway type or characteristics, lighting, and environmental factors. As indicated in the table, some of these may be manipulated by route selection and scheduling, while others can only be manipulated in a simulator or by capitalizing on chance occurrences of natural phenomena (e.g., schedule an impromptu run if it rains). In general, the selection of independent variables should be guided by the anticipated range of device characteristics, driver characteristics, and driving conditions that might be common to a particular device. As the number of independent factors goes up, the complexity of the evaluation increases exponentially. Thus, the minimum independent variable set should include device characteristics and a select few driving condition variables (e.g., route, traffic density), as anticipated for that device.

3.2.2 Select Dependent Measures (Measures of Performance)

The selection of dependent measures or Measures of Performance (MOPs) depends on a theory of workload and driving safety. Such a theory was presented earlier in this protocol. Appendix A through Appendix D present a system of candidate dependent measures proposed for workload assessment. These are actually measures of performance (MOPs) that bear a relationship to measures of effectiveness. These appendices provide guidance on the motivation, instrumentation needs, and operational definitions of these MOPs.

Table 3-5. Independent variables that may be manipulated in a Driver Workload Evaluation.

Independent Variable Category	**Independent Variables**	**Method of Manipulation**	**Comment**
Driver Characteristics	Age	Selection	Literature review, preliminary studies, or targeted audiences will indicate what driver variables might be appropriate to include in a given device evaluation. Specific levels for a given evaluation are context-specific.
	Gender	Selection	
	Driving Experience	Selection and Training	
	Device Experience	Selection and Training	
	Skill or Abilities	Test & Selection	
	Permanent Handicaps	Selection	
	Altered States (fatigue, intoxication, inattention)	Various techniques (.e.,g., prolonged driving, administration of alcohol, visual occlusion, distractors)	
	Motivation	Instructions, payoffs, feedback	

Table 3-5. Independent varables that may be manipulated in a Driver Workload Evaluation. (Cont.)

Independent Variable Category	**Independent Variables**	**Method of Manipulation**	**Comment**
In-cab Device Characteristics	Display, Visual Display, Auditory Device Modes Device States Controls	Location, visual angle, resolution, contrast, brightness, polarity, content, etc. Auditory display location, pitch, volume, content, etc. Device Modes are those device configurations that may be selected by the driver while Device States are conditions that may or may not be selected by the driver (e.g., failures) Controls may be varied by control type, resistance, throw, fine-tuning requirements, etc.	In a formative evaluation, factors such as these could be manipulated. In a summative evaluation, these would likely be fixed parameters within a given design. However, it is still possible to compare the workload of alternative devices, each of which represents its own complex of control, display, and logical features.

Table 3-5. Independent varables that may be manipulated in a Driver Workload Evaluation. (Cont.)

Independent Variable Category	**Independent Variables**	**Method of Manipulation**	**Comment**
Driving Conditions	Environment Lighting Precipitation Effects Wind	Lighting (day/night) Precipitation effects (visibility, sensor degradation, loss of traction) Wind (gusts)	Lighting affects sight distance. Precipitation provides obscurance, affects coeff. of friction. Wind gusts perturb lanekeeping.
	Roadway Roadway type	Roadway alignment (horizontal) Roadway alignment (vertical) Lane width Roadway skid resistance Shoulder width Shoulder skid resistance Posted Speed Limits Road obstructions Roadway type (divided/undivided) Intersection geometry	Roadway variables determine the effects of the "track" the driver is on. Specific features are scenario-dependent. May be manipulated in a simulator, by route selection on-the-road.
	Traffic	Number of other vehicles Principal Other Vehicle (POV) relative position, direction of travel, velocity, acceleration. Pedestrian/animal location, movement POV driver behavior	Traffic variables affect potential conflict situations, depending on context.

3.2.3 Identify and Determine Controls for Extraneous Variables

It should also be noted that there will be many factors that could potentially affect the outcomes of a workload evaluation that are not explicitly independent factors. These are called extraneous variables or nuisance variables. Examples of extraneous factors that would not normally be a focus of workload evaluation include varying levels of driver fatigue, boredom, and motivation during the testing. Other possible extraneous factors, for a given evaluation, might include driver age or gender, if not explicitly a part of the evaluation plan. Sometimes these factors can be controlled explicitly by one or more of the following means:

- Random assignment of subjects to test conditions;
- Counterbalancing of test conditions to effectively eliminate systematic effects of extraneous or nuisance variables like boredom, fatigue, increased familiarization with the testing procedures, and the like;
- Maintaining the extraneous factor at a constant levels (e.g., selection of only young drivers to control for age effects or only female drives to control for gender effects);
- Blocking, i.e., explicitly defining a blocking variable such as driver age and assigning volunteer drivers to levels in the block for later analysis. Note that blocking makes the blocking factor a part of the analysis; and
- Use of each subject as his or her own control. This is often maximally effective in controlling for subject variability but is only suitable for situations were there is no possibility that experience in one experimental situation carries over to the next in asymmetric ways, and when it is feasible to have the subject experience multiple conditions.

Note that research earlier in the project indicated that time stress due to running late was the largest source of workload, as reported by the drivers themselves (Kiger, Rockwell, Niswonger, Tijerina, Myers, and Nygren, 1992). Integration of time stress into an evaluation may not be feasible due to safety constraints or some other reason. If so, time stress should be eliminated to the extent possible.

3.3 DESIGN THE EXPERIMENT/EVALUATION DATA COLLECTION STRATEGY

Given the selected independent and dependent variables and controls chosen for the extraneous variables, a key planning activity is the design of the workload evaluation experiment or data collection strategy. This is the step in which an orderly method of data collection is determined. The experimental plan specifies the particular treatment conditions or combinations of independent variables that will be included in the evaluation, the assignment of subjects to those treatment conditions, the order of testing subjects, and the statistical analyses that will be most appropriate. Experimental design is discussed in many textbooks devoted to the subject (Keppel, 1991; Winer, 1971; Box, Hunter, and Hunter, 1978; Kirk, 1982). See Appendix F for more information about experimental design and data collection strategies.

3.4 DEVELOP DATA COLLECTION HARDWARE/SOFTWARE INSTRUMENTATION NEEDS

Plans for procuring and equipping an instrumented test vehicle that satisfies the requirements of the Driver Performance Tests are described in Appendix A through D. Each Appendix describes instrumentation needs particular to each class of measurement. It should be noted that instrumentation is in a constant state of evolution. New technology (e.g, the DASCAR program sponsored by NHTSA) is currently under development that can make it easier, less costly, and more reliable to equip a test participant's own vehicles for data collection and allow for data collection over longer periods of time.

Turanski and Tijerina (1992) describe the process by which a standard heavy vehicle was chosen. A conventional cab was chosen over a cab-over because it appeared at that time that the conventional cab was most common. However, the selection of heavy vehicle depends on the application at hand and the target population for the analysis. In addition, the nature of the trailer used for the workload assessment (e.g., single versus double versus triple; 48 ft versus 52 ft length; payload weight) should be tailored to the research purposes and applications at hand.

3.5 DEFINE PROCEDURES TO BE FOLLOWED

This step involves developing all procedures to be followed for pre-pilot test, pilot test, and formal test. Procedures must be developed for several categories of activity. These are discussed briefly below.

3.5.1 Equipment Status and Startup Procedures

It is very important that the evaluator or ride-along observer who will be conduct the data collection session fully understand how the equipment operates. Start-up procedures must be indicated in a checklist and the checklist must be validated by having the ride-along observer attempt to use it. The instrumentation engineer should be a witness to this validation and make necessary revisions to the startup procedures as needed. A troubleshooting list should also be prepared by the instrumentation engineer so that the evaluator or ride-along observer may be sensitive to failure modes and so minimize data loss or damage to equipment.

3.5.2 Test Participant Intake and Training Procedures

The evaluator must be provided with appropriate procedures for intake and training. The procedures for intake will usually take the form of demographic questions to characterize the driver and screen the individual for compliance with selection criteria. If the evaluation involves driver abilities (e.g., spatial ability), part of the driver intake will involve administration and grading of tests so the test to be used must be selected and provided to the evaluator conducting the data collection sessions. Finally, training procedures must be developed. These may range from very informal procedures, to demonstration of the in-vehicle technology, to formal training with printed material, perhaps video media, and tests to determine if the driver has achieved the minimal level of proficiency on the system.

3.5.3 On-site Evaluator/Observer Procedures

Procedures must be developed for the evaluator or ride-along observer to follow during the data collection session. These procedures may cover conditions of safety (e.g., do not request an in-vehicle device transaction if the headway is less than 150 ft), the evaluation of current conditions (e.g., operational definitions the evaluator is to use in characterizing driving conditions, e.g., traffic density), pacing of events in the data collection session, wording to be used in interacting with the driver, and so forth.

3.5.4 Data Evaluation and Management Procedures

There must be procedures for the checking and management of data (e.g., when to change video cassettes, high-density data cartridges, etc). These procedures should provide a means to determine if sensors are properly calibrated and to detect drift or failure. The management of data must include procedures to log or catalogue data for later analysis and minimize data loss.

3.5.5 Emergency Procedures

The ride-along observer must be informed about procedures to follow in the event of an emergency. Emergencies may vary from vehicle breakdown and instrument failure, to a mishap or crash. Contingency plans for “no show” subjects or sudden illness by the evaluator must be developed so that disruption to scheduled activities may be minimized.

3.6 SELECT AND RECRUIT TEST PARTICIPANTS

3.6.1 Test Participants

To the extent possible, test participant sample size, composition, and selection procedure should be based on both statistical and experimental considerations. Sample size might be determined by power analysis (Cohen, 1988), though this is often impractical and provides sample size estimates that are difficult to implement. In practice, sample size takes into account expected constraints on driver availability, equipment, time, and other resources.

All test participants participating in the tests should be paid volunteers. In addition, all recruits should be screened for the following criteria prior to participation:

- Valid driver's license for the type of vehicle used in the study [e.g., Class A Commercial Driver's license (CDL) for large trucks]

- At least three years of experience in driving vehicles similar to the test vehicle

- No more than three moving violations in the past three years as indicated by state records

- Insurance coverage (provided by research organization or subject must submit proof of liability insurance)

- Vision of at least 20/40 (corrective lenses acceptable) as measured with a Baush and Lomb Orthorater (Note: This test is administered during many driver's license examinations. Thus, a valid driver's license is indicative of adequate vision to drive.)

- Hearing within normal range (by age) as measured with a portable audiometer. Alternatively, one can substitute a current medical certificate from the driver's employer in lieu of a hearing test or a vision test.

- No drug or alcohol abuse as measured with self-report and experimenter observation. In practice, most subject samples are not randomly selected. In such cases, researchers should strive to recruit subjects so that the sample has characteristics that match the characteristics of the target population.

- No physical or psychological conditions that might preclude participation

Workload has been shown to be affected by many different driver factors including driver gender, age, driving experience, familiarity with the test route, familiarity with the vehicle, familiarity with the device, fatigue levels, and personality traits, among other factors. If the evaluation is to address such factors as driver indenendent variables, there will of course be a need to recruit appropriate persons. It may also be necessary to test candidates as a selection technique to ensure that drivers with particular attributes (e.g., high versus low spatial reasoning abilities) are included in the evaluation.

As a general rule, it is recommended that the test participant sample consist of people similar to those to whom the assessment results will be applied. Ideally, subjects will be a random sample from the population of interest. In practice, volunteer subjects (self-selected) are often used.

3.6.2 Test Participant Recruitment

Test participants may be recruited in a variety of ways. These might include newspaper ads, announcements at truck stops, direct coordination with dispatchers at local trucking firms, and so on. As a rule, the test participant will be contacted by phone to request participation in the testing. All test participants will be paid for their participation. This remuneration should be set at a level to provide inducement to participate in the study. If a candidate refuses to participate, another candidate should be sought, ideally from a pool of candidates by random selection.

3.6.3 Test Participants Release Form

Test participants must be briefed on the nature of the performance test objectives and methods to be used in the study in which they will participate. They will be provided with a subject consent form (see Appendix G) that provides information necessary for informed consent and release from liability. Test participants will have an opportunity to obtain a response to any questions regarding the procedures or informed consent form. Note that the informed consent form may be designed to adhere to the National Highway Traffic Safety Administration (NHTSA) Order 700-*1* on *Protection of the Rights and Welfare of Human Subjects in NHTSA-Sponsored Experiments* (NHTSA, 1981).

3.6.4 Test Participant Orientation and Training

Test participants must be informed about the procedures that will be followed in the device evaluation. In addition, the planning must address the degree of training to be provided on the to-be-evaluated in-vehicle device. In general, there should be at least a minimal amount of orientation to the device so that the driver is at least acquainted with device form and function. At the other end of the training continuum, the driver might be trained with a structured set of training materials until the driver exhibits a certain minimal level of proficiencies (e.g., complete device transactions without error, complete device transactions within 4 s or less, and so on). In between, training might involve presentation of a demonstration of the device with an opportunity for the driver to ask questions. Since training may in fact be an independent variable in the workload assessment, no definitive statements can be provided. However, it is important to realize that a worst-case analysis will involve a driver who has had little or no familiarization with the in-vehicle device.

3.7 DEVELOP WORKLOAD EVALUATION SCHEDULE

This section presents some points for consideration in developing a schedule for the evaluation.

3.7.1 Develop Initial Schedule

It is important to develop an initial schedule based on sound project management principles. All too often, insufficient time is allocated for completion of hardware and software development and checkout. There may be early lead times needed to order instrumentation or arrange for limited resources (e.g., an instrumented vehicle, a simulator) to be made available for the workload assessment. It is beneficial to develop a PERT chart that shows a network of activities that must be completed for the workload evaluation to be completed. This chart should show the sequential contingencies among the various tasks so that the program manager is able to determine the critical path that constrains completion date. This is also a useful means to provide estimated task completion times to determine when the project can actually be completed and to assess the impact of delays or increases in the actual vs. the estimated task completion times. Software such as Microsoft Project Manager ™ provides useful tools to accomplish the initial schedule.

3.7.2 Establish a Contingency Schedule

Murphy’s Law states that what can go wrong will go wrong. Murphy’s Law seems to apply double for human experimentation. Delayed starts, instrumentation failure, inclement weather, a union strike, ... all of these factors have plagued the authors of this document and can beset any evaluator. Some provision must be made to accommodate such problems should (or when) they arise. The contingency schedule applies also to the micro-schedule of a data collection session. For example, if the driver, for whatever reason, does not complete a task called for in the evaluation plan, what should be done? In essence, up-front “what if’ thinking can prove valuable when adversity strikes. It is important to develop an initial schedule based on sound project management principles.

REFERENCES

Box, G. E. P., Hunter, W..G., & Hunter, J. S. (1978). *Statisticsfor experimenters: An introduction to design, data analysis, and model building. New* York: John Wiley.

Burgett, A. L. (1994, May). Methodologies for evaluating the impact on safety of Intelligent Vehicle Highway Systems. *Proceedings of the 14th International Technical Conference on Enhanced Safety of Vehicles*, Munich, Germany, May 24-26, 1994.

Cohen, J. (1988). *Statisticalpower analysis for the behavioral sciences* (2nd Edition). Hillsdale, NJ: Erlbaum.

Dingus, T. A. (1995). Moving from measures of performance to measures of effectiveness in the safety evaluation of ITS products or demonstrations. In D. Nelson (Ed.), *Proceedings of the ITS Safety Evaluations Workshop.* Washington, DC: ITS AMERICA.

Keppel, G. (1991). *Design and analysis: A researcher 's handbook* (Third edition). Englewood Cliffs, NJ: Prentice-Hall.

Kiger, S., Rockwell, T. H., Niswonger, S., Tijerina, L., Myers, L. B., & Nygren, T. (September, 1992). *NHTSA heavy vehicle driver workload assessment Task 3 Interim Report: Task analysis data collection* (Contract No. DTHN22-91-C-07003). Columbus, OH: Battelle .

Kirk, R. E. (1982). *Experimental design: Procedures for the behavioral sciences* (Second edition). Belmont, CA: Brooks/Cole Publishing.

NHTSA (198 1). *Protection of the rights and welfare of human subjects in NHTSA-sponsored experiments.* Washington, DC: Author.

Turanski, A., & Tijerina, L. (January, 1992). *NHTSA heavy vehicle driver workload* ***assessment Task 2 Interim Report: Standard vehicle configuration/task spec@cations*** (Contract No. DTHN22-91-C-07003). Columbus, OH: Battelle .

Winer, B. J. (197 1). *Statistical principles in experimental design* (Second edition). New York: McGraw-Hill.

4.0 WORKLOAD ASSESSMENT TEST EXECUTION, ANALYSIS, AND REPORTING

The execution of a workload assessment should generally go through three phases:

- Pre-pilot testing: This testing involves members of the evaluation team as subjects to verify instrumentation and test procedures and allow for fine tuning of the test as needed;
- Pilot testing: This testing involves bringing in one or a small number of driver subjects to verify the instrumentation and test procedures. Since the driver subjects are unfamiliar with the details of the test, they will provide useful information on aspects of the testing approach that are problematic in terms of understanding and execution. At the end of this testing, there may be refinements or modifications of the test plan that are required;
- Formal Testing: This is the basic testing that comprises the workload assessment. After the refinements in procedures and apparatus are developed, a sample of driver subjects is used to collect data on workload associated with in-vehicle device use and other factors. The results of the formal testing are what are reported.

Each of these testing phases is discussed below.

4.1 PRE-PILOT TESTING

The purpose of pre-pilot testing is to allow the evaluation team to try out the evaluation protocol on themselves. This testing allows for calibration of the data capture system, and verification that all systems (e.g., in-vehicle system, data capture system) have been fully integrated and are functioning as expected. The initialization procedures can be tested for completeness and correctness. Subject instructions, training materials and procedures, test procedures, timing and sequencing of events during the data collection, potential problems and their resolution are all part of what may be addressed during the pre-pilot testing. Based on the data collected during pre-pilot testing, changes to the evaluation protocol may be made as needed.

4.2 PILOT TESTING

The pilot test is conducted after the changes made based on pre-pilot testing have been integrated into the overall procedures of the workload evaluation. The pilot test differs from the pre-pilot test in that subjects similar to the intended population are involved, rather than using members of the evaluation team as subjects. It is not uncommon to find out that persons new to the evaluation protocol reveal needs for further changes than those that were uncovered previously. Based on pilot test results, additional changes may be called for in the protocol and these must be incorporated prior to the actual testing or formal testing.

4.3 FORMAL TESTING

Formal testing is the actual data collection phase of the workload evaluation. Given all of the preparations that have gone before this, the formal testing involves completing the protocol as designed and amended. Formal testing should be monitored assessed to insure that all is going according to plan. Some potential problems that might arise include equipment malfunction, cancellations by subjects, evaluation team cancellations due to poor weather, changes in the test route that come unexpectedly (e.g., road construction), union strikes, and other factors beyond the evaluators' control. These factors can substantially alter the schedule and contingency plans developed earlier in the planning process will prove useful.

Ideally, the data collected would be reduced shortly thereafter to ensure data quality. It is sometimes possible to compile results as each subject's data becomes available and to cancel further data collection once an effect or trend has been established with a certain degree of statistical confidence. The results of an early-on evaluation may prompt additional data collection or supplemental tests. Even if concurrent data reduction and analysis is not feasible, periodic evaluations of all equipment and procedures must be done to allow for recalibration as needed.

4.4 PREPARE DATA FOR ANALYSIS

Once data have been collected, it is possible to conduct a statistical assessment. However, the data analysis process begins with data preparation, continues with data analysis, and includes interpretation of the results.

4.4.1 Reduce data

Once data have been collected and managed (i.e., logged or catalogued properly), data reduction procedures may begin. See Appendices A through D for additional information on data reduction specific to each class of workload measures. At a minimum, data reduction involves turning all measured channels into engineering units (e.g., lane position may have been recorded as a voltage but is converted to inches), filtering the channels as needed, and parsing the data stream into the major treatment conditions to be evaluated. This step is critical to subsequent analysis. If there are errors made during this step, those errors will promulgate through the analysis and potentially lead to false conclusions.

The data reducer may encounter anomalies, i.e., situations that do not conform with the evaluation plan or what was expected. An example might be how to treat a particular type of error that the data reducer uncovers but has never been considered previously. Given these are novel cases, it will be important for the evaluation team to review these anomalies and develop a means of dealing with them. Unfortunately, the only guidance that can be offered is to take whatever steps are needed to preserve as much data as possible. Beyond that, the evaluator should pursue the simplest analysis that will meet the objectives of the workload assessment.

4.4.2 Verify reduced data

Once reduced data are available, it is important for a knowledgeable person to review the data and check for any obvious errors. This may be as simple as verifying that events occurring one after the other have successively later event time codes. There is a need for the reviewer to be familiar with what the data should look like so that anomalies may be noticed. At a more rudimentary level, manual data reduction (e..g., taking data off of interview sheets and entering them manually into an ASCII file for subsequent analysis) might be double-checked by a different person than the data reducer for:

- missed data,
- misclassified data,
- transposed digits,
- the accuracy of simple intermediate calculations,
- simple engineering transformations (e.g., 32 ft./s^2 = 9.8 m/s^2).

4.4.3 Identifv and Manage Missing Data

Inevitably, there are missing data. Equipment breaks, the driver does not or cannot complete one or more in-vehicle transactions, the driving conditions planned for simply never materialize. These are but a few of the reason why, despite excellent planning, some data will be missing. While it is advisable to minimize the likelihood for missing data, procedures may be applied to deal with missing data when it arises.

The reasons why missing data are important to plan for and address are bound up in the data analysis. Certain types of data analysis are particularly sensitive to missing data; one example is multivariate analysis, a set of procedures that assumes complete data sets. There are three basic approaches that may be taken to deal with missing data (Rummel, 1970). Each is discussed below.

One simple approach is to substitute the average value for the response variable in the particular treatment condition where the missing value is located This is a simple and therefore popular method but it has its drawbacks. In particular, inserting averages in place of missing data will effectively lower any correlations that might otherwise be present in the data. The more averages that are substituted for missing data, the more the overall correlations or covariances will underestimate the true values.

A second approach is to replace missing data with the results of a regression model. In regression analysis, the available data on each variable are regressed on the available data on the other variables to determine regression estimates for the missing data. A number of regression equations equal to the number of missing variables with missing data are computed to determine regression estimates for all missing data. The equations may be recomputed, including the missing data estimates this time, to generate a new set of estimates. This method is both efficient and reliable if the variables in the data matrix are highly correlated. On the other hand, if variable intercorrelations are low, the regression approach will yield poor estimates and the margin of error may be quite high.

A third approach, perhaps used only as a last resort, is to analyze only complete data sets. Thus, any record that is missing data will be removed from the analysis. If one is attempting to conduct multivariate analysis, this can be a drastic move. If, however, the missing data are concentrated in a few measures, perhaps dropping cases only for those measures is reasonable.

4.5 CONDUCT DATA ANALYSIS

Once a reduced data set is available, data analysis may begin. The recommended approach is to carry out the simplest analysis required to meet the objectives of the workload assessment. Key steps are discussed below.

4.5.1 Examine Reduced Data

The first step in examining the reduced data from a data analysis standpoint is to plot it. Graphical displays of the data provide insights into how the sample is distributed, if the data appear normally distributed (this is rare in human factors work), what outliers are present, and if there is any connection between outliers and subjects (e.g., the same person tends to be an extremely good or extremely poor performer) By looking at the data, the evaluator may select a data transformation. The purpose of the data transformation is to allow the statistical machinery to work properly and allow for sensitive tests to be carried out. In general, conventional statistical procedures assume that the data are normally distributed, that the variances among a set of treatment conditions are similar or identical, and that there are no correlations between errors and treatment conditions. Figure 4-1 provides a display of several types of transformations that may be applied and the conditions under which they are appropriate.

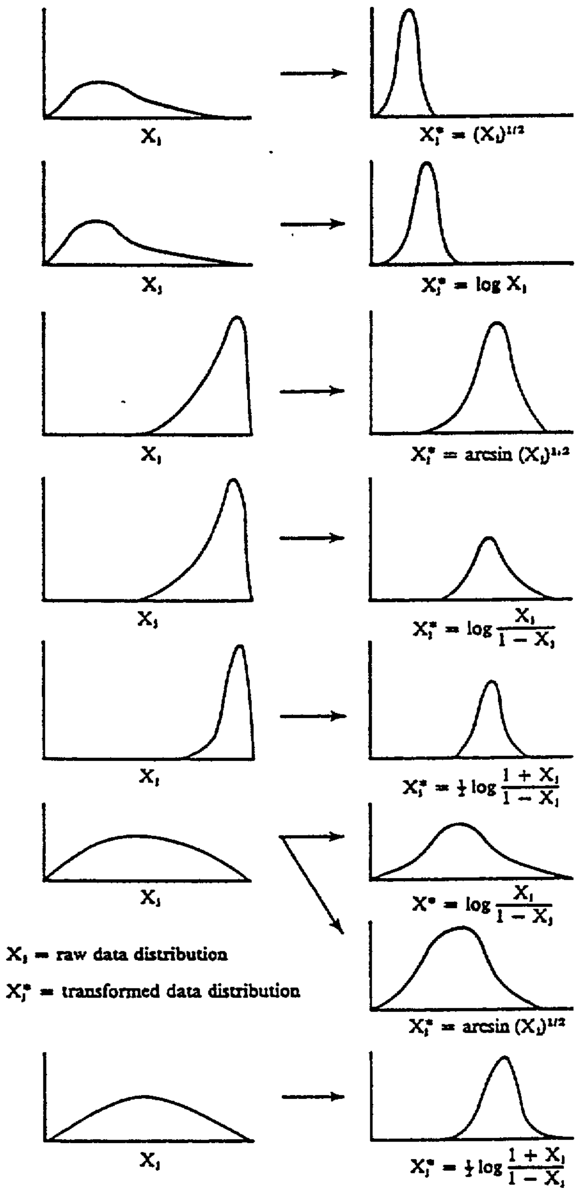


Figure 4-1. Distributional Transformations (Source: Rummel, 1970)

4.5.2 Determine Appropriate Analysis Techniques

This step in the analysis is best carried out with the consultation of an experienced data analyst or statistician. It was mentioned earlier that the simplest analysis is preferred if it can answer the questions and hypotheses that motivated the evaluation. Some basics of statistical assessment are briefly described below. This material is taken from Unisys (1987).

A statistical hypothesis is an assumption about a population based upon some provisional theory (e.g., lane exceedences with in-vehicle device use are the same as with open-road driving). A statistical test is a formal procedure for assessing whether this provisional theory or hypothesis should be rejected or not.

The procedure of the statistical. test is to give the facts of the sample data an opportunity to discredit the hypothesis (called the null hypothesis). If the sample data do, then the null hypothesis is rejected and decisions are made assuming that the provisional theory is false. On the other hand, if the sample data do not discredit the hypothesis, then decisions are made assuming that the null hypothesis or provisional theory is true. The statistical test allows for statements to be made about the likelihood that the sample results could turn out the way they did if the null hypothesis is true.

A variety of statistical procedures may be applied to workload data. In the final version of this protocol, selected procedures will be described in more detail. In general, the advice of an experienced statistician or data analyst is advised. There are almost always differences between the planned and the actual evaluation that require statistical expertise.

4.5.3 Apply Analysis Techniques

Once appropriate statistical techniques are selected, there are numerous software packages available for their execution. Some, like the Statistical Analysis Software (SAS^{TM}), the Statistical Package for the Social Sciences (SPSS^{TM}), or the BioMedical Data Program (BMDP^{TM}) are quite sophisticated and have their own programming language. On the other hand, spreadsheets like Excel^{TM} can perform many types of analysis satisfactorily. The advice of a statistical consultant will be valuable in applying as well as selecting the analysis techniques.

4.6 REPORT RESULTS

Upon completion of the analysis, the results and their interpretation must be written up and conveyed to management or an outside source (conference reviewer, journal review panel, government body, etc.) as required. This report should contain all the parts of parts of a scientific report (e.g., American Psychological Association, 1994):

- Introduction - This is both an introduction to the system under evaluation but also a review of the background that led to the evaluation carried out. This section concludes with the purpose and rationale of the workload evaluation.
- Method - The Method section describes the test participants, instrumentation, system evaluated, vehicle(s) used, and driving scenarios, as well as the procedures used.
- Results - In this section the dependent and independent variables are reviewed, the statistical methods used are introduced, and the analysis results are presented in numeric, tabular, or graphic form, as appropriate for the data.
- Discussion - This section deals with the interpretation of the results, their implications for system development, and their implications for other systems beyond that (or those) evaluated.
- References - Any literature used for the study should be cited.
- Appendices - As many appendices may be included as required to explain what was done.
- Acknowledgments- This allows the evaluation team to acknowledge individuals who contributed to the evaluation but are not co-authors of the report, as well as contributions from outside agencies (e.g., trucking firms that arranged for drivers to know about and volunteer for the study).

REFERENCES

American Psychological Association. (1994). *Publication manual **of** the American Psychological Association* (Fourth edition). Washington, DC: American Psychological Association.

Rummel, R. J. (1970). *Appliedfactor analysis.* Evanston, NJ: Northwestern University Press.

Unisys. (1987, July). *User-assisted test and evaluation methodology assistant program (TEMAP): Version* I (Contract No. F33615-85-C-0533). Wright Patterson AFB, OH: Crew System Ergonomics Information Analysis Center (CSERIAC).

5.0 EPILOGUE

In this chapter, a streamlined workload assessment protocol is introduced as an epilogue to the material that has been presented in previous chapters.

5.1 A STREAMLINED WORKLOAD ASSESSMENT PROTOCOL

The-preceding chapters provide guidance on the development and execution of a formal evaluation using on-the-road scenarios and an instrumented vehicle. The time and cost of such an assessment is prohibitive for all but the most extensive evaluations in a well-funded safety evaluation program. Given that there are many other potential applications where workload assessment might be beneficial to product development, is there a streamlined workload assessment protocol that might be employed? The answer is a tentative "yes" with caveats. A streamlined workload assessment protocol might be comprised of the steps described next.

5.1.1 Step 1. Analyze the functions of the device

Regardless of what else is done, this step is necessary to determine when and how a device might be used. This step allows the evaluator to become acquainted with device functions, how each function works, what system components (buttons, screens, knobs) are required by a function, and when and in what situations the function might be used and why the driver will make use of that function. It is device functions that place demands on drivers, so this step creates the task list that might be examined later.

5.1.2. Step 2. Apply an Ergonomics Checklist

An ergonomic checklist is a series of statements which describe the individual features that a device or procedure should have to be properly human engineered (Meister, 1985). They may be applied by inspection or by taking measurements. Ergonomic checklists are the most common form of human factors assessment conducted today (Meister, 1986). Checklists are available from the military establishment (e.g., Department of Defense, 1989). In addition, human-computer interface design guidelines can be helpful (e.g., Smith and Mosier, 1986). The Society of Automotive Engineers has not promulgated design guidelines or checklists per se but the Department of Transportation is currently funding a great deal of research in this area (e.g., COMSIS, 1993). While checklists are far from a complete and thorough human factors analysis, this approach can be of benefit in spotting poorly designed devices (or functions) from the outset. If a system shows serious violations of basic human factors principles from a checklist review, there is no technical reason to conduct an elaborate evaluation to demonstrate the obvious. On the other hand, lack of compliance in some areas may be grounds for further investigation or scrutiny about the impact it may have on device use while driving.

5.1.3 Step 3. Conduct a desktop evaluation with a video game as the primary "driving" task.

There is a need to load the test participant on a primary (driving) task in order to collect workload measures. One potential approach is to use a video game which provides some driving task load and some score. The in-cab device (or prototype) might be set up next to the video game, along with a commonly used device (e.g., radio for tuning). A video camera (with a millisecond timer) can be set up on a tripod to capture a test participant's eye movements. For a given set of tasks, then, the following measures can be captured: mean glance duration, number of glances, glance durations back to the "road scene" of the video game, and video game score. This is an inexpensive method of capturing human performance data and might be useful for early evaluations. On the other hand, video games may overload the test participant more than actual driving would. This could lead to results that are poorer than might result in the real world. Furthermore, a video game in general will not match vehicle dynamics of a real vehicle so as to provide comparable psychomotor load, nor will there be equivalent visual cues. Motion cues will be absent. Thus, this is another simplification to the kind of evaluation outlined in earlier chapters that may provide useful insights, but cannot necessarily be taken as a sufficient test of device workload.

An alternative to a video game would be a part-task siiulator (e.g., Bittner, Rowley, Lee, and Kantowitz, 1994). Note that the use of driving simulation is a potentially expensive undertaking. A part-task simulator can run tens of thousands of dollars for the basic hardware and software alone. Development of special test scenarios may be a substantial programming task and so simulator testing is not necessarily less expensive than on-the-road testing. Even with this additional expense over video game technology, fidelity may be insufficient and the validity of simulator results may be questioned. A fixed-based simulator provides no motion cues and so will not generate psychomotor loads similar to those found in real driving. It is difficult to match vehicle dynamics in simulation so that the psychomotor load remains similar to that of driving. Drivers may not place the same emphasis on the simulated driving task as they would in real driving, thereby skewing the results. For these reasons, simulation has been presented as an adjunct rather than a replacement for on-the-road testing.

Smiley (1995) has recently noted that the issue of simulator validity should be considered in light of the alternatives posed by on-the-road testing. Simulators allow for safe testing and simulators of varying degrees of sophistication may be used, depending on the goals of the evaluation. Smiley calls into question those who believe that the only valid measures are those obtained in the field. The reality of most field tests or on-the-road evaluations is that such experiments are also simulations of driving. To reduce variability, it is pointed out, test participants are given specific instructions, potential conflicts with other vehicles are strictly limited, and the presence of an observer or experimenter undoubtedly alters behavior to at least some degree. Perhaps the only conclusion that can be drawn at this time is that simulation may provide a lower-cost, safer alternative to on-the-road testing for device evaluations. The validation of this view will have to come in the form of simulator studies which yield the same results as on-the-road tests.

5.1.4 Step 4. Carry out a check-ride or simplified on-the-road test

If the device survives the checklist review and a preliminary simulator test, it is still beneficial to have an opportunity to observe device use on the highway. The driver might be able to complete a checklist such as that provided in Table 5-l.

This checklist is best reviewed after driving on the road with the device. A "yes" to any question should be considered grounds for further workload assessment or device redesign. If an instrumentation package is available, then the following might be pursued in decreasing order of usefulness:

- Videotape the driver's visual allocation (device average single glance duration, number of glances to complete device transaction, road average single glance duration; mirror sampling proportion);
- If possible, use a lane tracker or second video camera (and light source) to capture lane position (lane exceedences, lane standard deviation);
- Mean speed and speed variance;
- If possible, instrument the steering wheel and pedals;
- Include additional subjective assessments like those in Appendix D.

The sensors may be interfaced to a PC with filtering or signal conditioning provided before the data are stored to diskette or hard drive. The instrumentation may be turned on some time before a transaction (requested by the observer) and turned off sometime thereafter to preserve computer memory. This small set of measures may be sufficient to address the workload issue.

The streamlined testing may be conducted with perhaps eight to ten test participants (if statistical analyses are to be carried out, more test participants are advised). Test participants would be chosen so as to be representative of the prospective user population. A standard route of interstate roadway, in daylight, dry weather, with moderate to light traffic density may be used for the test scenario. The scenario would exercise the device and include a conventional in-cab task (manually tuning **the** radio) for comparison purposes. If each transaction is no more disruptive than the radio tuning task, then the system may be considered acceptable. If not, further

Table 5- 1. Device Workload Checklist to be completed by driver after on-the-road use of in-cab device. Any YES response merits further investigation.

	QUESTIONS		**COMMENTS**	
1.	Can the device be used when the vehicle is moving?	NO	YES	
2.	Does the driver have to look at the device to use it?	NO	YES	
3.	Does the device have controls, e.g.,buttons, keypad, knobs, touch-screen, slide levers, switches, etc. (e.g.. a radio has controls, a speedometer does not)?	**NO**	YES	
4.	If there is a visual display, does it display sets of numbers, text, map information, or other complex data?	**NO**	**YES**	
5.	In your opinion., is the device hard to read under normal conditions?	**NO**	**YES**	
6.	If the device has controls, do you have to visually attend to those controls (e.g., like a inserting a cassette into a car stereo)?	**NO**	YES	
7.	In your opinion, are the controls hard to use under normal conditions?	NO	YES	
8.	Does the device take longer than about 1.5 seconds to use?	NO	YES	
9.	Can the device prompt you to use it (e.g., cellular phone ring and you answer)?	NO	YES	
10.	Is it hard to start, stop, then pick up again what you were doing with the device, e.g., reading a display or entering data?	**NO**	YES	
11.	Is the use of the device mandatory?	NO	YES	
12.	Do you have any concerns about the safety of this device for heavy vehicle use?	NO	YES	

investigation would be warranted. Note that this streamlined procedure is much reduced from a research endeavor, guidance for which is provided in other portions of this document.

It would be ideal to provide nomographs that indicate cut-offs, i.e., go/no-go indications for device workload based on the data collected in either the streamlined evaluation, or the detailed formal evaluation schemes. Elsewhere, the authors have argued that go/no-go criteria are not easily defined or defended (Wierwille et al., 1992). For illustration purposes only, Figure 5-l illustrates the type of nomograph that might be constructed based on visual allocation measures and results of ongoing research.

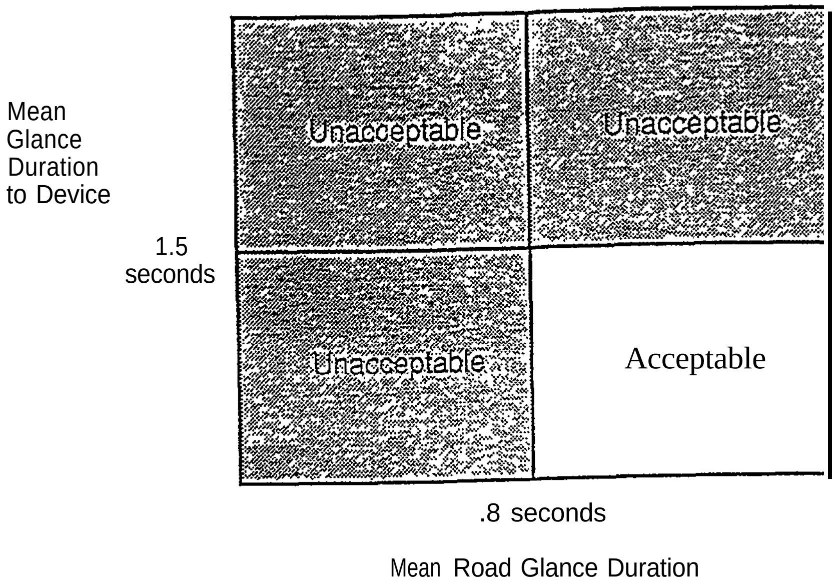


Figure 5-l. Hypothetical Workload Nomograph Based on the Visual Allocation Measures of Mean Glance Duration to the Device and Mean Glance Duration to the Road Scene. Shaded Areas Constitute Unacceptable Workload.

REFERENCES

Bitter, A. C. Jr., Rowley, M., Lee, J. D., & Kantowitz, B. H. (1994, July). *NHTSA heavy vehicle driver workload assessment Task 5 interim report: Workload measurement protocol development – simulator pilot study* (Contract No. DTN’H22-91-C-07003). Columbus, OH: Battelle.

COMSIS. (1993, October). Preliminary human factors guidelines for crash avoidance warning devices (Contract No. DTNH22-91-C-07004). Sliver Spring, MD: Comsis Corporation.

Department of Defense. (1989, March). *Military standard: Human engineering design criteria for military systems, equipment, and facilities* (MIL-STD-1472D). Washington, DC: Government Printing Office.

Meister, D. (1985). *Behavioral analysis and measurement methods. New* York: Wiley.

Meister, D. (1986). A survey of test and evaluation methods. *Proceedings of the Human Factors Society 30th Annual Meeting,* 1239 - 1243.

Smiley, A. (1995). Overview and methodologies of the ITS safety evaluation process. In D. Nelson (Ed.), *Proceedings of an ITS safety evaluations workshop.* Washington, DC: ITS AMERICA.

Smith, S. L., & Mosier, J. (1986). *Guidelines for designing user interface software* (Report No. ESD-TR-86-278). Hanscomb, AFB: Electronic Systems Division, Air Force Systems Command.

Wierwille, W., Tijerina, L., Kiger, S. M., Rockwell, T. H., Lauber, E., & Bittner, A. (1992). *NHTSA heavy vehicle driver workload assessment Task 4 Interim Report: Review* ***of*** *workload and related research and a theoretical basis* ***for safety*** *relevance* ***of*** *workload measures* (Contract No. DTHN22-91-C-07003). Columbus, OH: Battelle.

APPENDIX A

VISUAL ALLOCATION MEASURES IN DRIVER WORKLOAD ASSESSMENT

APPENDIX A. VISUAL ALLOCATION MEASURES IN DRIVER WORKLOAD ASSESSMENT

MOTIVATION

An estimated 90% of the information required for driving is acquired through the driver's sense of vision (Rockwell, 1972). This point is intuitively reasonable and has also been demonstrated in on-the-road driving and simulator studies that indicate the impact of momentary losses in visual input on drivers (Senders, Kristofferson, Levison, Dietrich, and Ward, 1967) and the effects of reduced sight distance for the road ahead (Allen and McRuer, 1977). In particular, there is but one foveal resource and it must be moved about to gather detailed visual information (Wierwille, 1993). This foveal resource is termed visual attention and its deployment by the driver is termed visual allocation. For these reasons, driver visual allocation has become an important aspect of driver human factors studies.

Early work examined the distribution of fixations across the visual field, fixation durations, fixation frequencies and percentages, and fixation sequences that are indicative of the driver's allocation of visual attention. For example, Mourant and Rockwell (1970) reported that with route familiarity, driver search and scan patterns became more compact and shifted down and to the left. Furthermore, the authors reported evidence from the eye movement records that peripheral vision is used primarily to monitor lane position and the presence of other vehicles and road signs, thus serving to direct foveal vision (and attention) as required. More recently, eye movements have been used to investigate factors that influence driver visual allocation, including vehicle factors (Kito, Haraguchi, Funatsu, Sato, and Kondo, 1989), search tasks (Louma, 1988; Hughes and Cole, 1988), visual scene complexity (Boersema, Zwaga, and Adams, 1989), and roadway parameters (Wierwille, Hulse, Fischer, and Dingus, 1988). For instance, eye movement data have revealed that drivers of large vehicles visually sample more at intersections than drivers of small vehicles (Kito, et al., 1989). Boersema et al. (1989) reported that search time and number of fixations increase systematically with the number of advertisements in search for a target word in a train station routing sign. Louma (1988) and Hughes and Cole (1988) reported that the nature of the driver's visual task affect scan patterns and direction of visual attention. Wierwille et al. (1988) found that driving-related glance times were positively correlated with increasing roadway demand characteristics (e.g., sight distance, road curvature, etc.), and in-cab navigation device glance times were negatively correlated increasing roadway demand characteristics. This suggests that drivers in the study adjusted their visual allocation appropriately to accommodate variations in driving task demand.

In addition to examination of visual allocation to elements of the road scene, work has been conducted to use eye movements as response variables to assess in-vehicle control and display workload demands. For example, Mourant, Herman, and Moussa-Hamouda (1980) reported on the use of direct looks to in-vehicle controls of different configurations and locations as a measure of driver workload. This paper explicitly posited that:

> "The positioning of controls so as to minimize direct looks will permit the driver to spend more time monitoring the forward scene for potentially dangerous events." (p, 417).

Mourant et al. (1980) found that the frequency of driver direct looks increased with increased hand travel distance to reach a control and also that look durations increased with increasingly complex control configurations.

Rockwell (1988) reported on the use of glance frequencies and glance durations as measures of driver in-vehicle visual performance. His data indicate that glance durations tend to be consistent and independent of the mean number of glances required to complete an in-cab task (e.g., radio tuning). Average glance duration is somewhat sensitive to task demand, though truncated because most drivers are unwilling to take their eyes off the road for more than perhaps 2 s. Bhise, Forbes, and Farber (1986) also reported that in-cab task demand has some effect on average glance duration but a much larger effect on glance frequency. Wierwille (1993) developed a driver visual sampling model that describes this behavior.

Most recently, extensive research has been carried out using visual allocation measures to assess attentional demand of in-cab controls and displays (see Wierwille, 1993 for a review). For example, Dingus, Antin, Hulse, and Wierwille (1989) recorded passenger car driver mean glance duration and mean number of glances for a wide variety of in-cab instruments and an operational in-vehicle route navigation system. Mean glance durations varied from approximately 0.62 s to 1.66 s while the mean number of glances to complete an in-cab transaction varied from 1.26 to 6.91 glances (see Table A-l). Tijerina, Kantowitz, Kiger, and Rockwell (1994) reported on the visual allocation of heavy vehicle drivers fixating on mirrors and various instrument panel devices during an on-the-road pilot study. As can be seen in Table A-2 the mean or average glance durations varied from approximately 1.06 s to 2.11 s while mean number of glances varied from 1.25 to 7.81 glances. Comparing across like in-vehicle device use, it appears that the heavy vehicle driver mean glance durations tended to be longer and more glances were required than was the case with the Dingus et al. (1989) passenger car drivers. Such differences underscore the need to collect baseline visual allocation data for both passenger car and heavy vehicle applications. Wierwille (1993) presents a task classification that predicts the variation in visual demand reflected in both sets of data.

While visual allocation measures are useful for driver workload assessment, they are not perfect. In particular, visual allocation measures are limited by the following points:

- The majority of interstate highway driving requires less than 50% of the driver's visual capacity (Rockwell, 1972). The driver therefore samples a large amount of extraneous information.

Table A-l. Passenger Car Driver Mean Glance Duration and Mean Number of Glances Associated with Various In-Vehicle Tasks (Source: Dingus, Antin, Hulse, and Wiexwille, 1989).

	hear Single Glance Length		Number of Glances	
Task	Maan	Standard Deviation	Mean	Standard Deviation
Speed	0.62	0.48	1.26	0.40
Following Traffic	0.75	0.36	1.31	0.57
Time	0.83	0.38	1.26	0.46
Vent	0.62	0.40	1.83	1.03
Destination Direction	1.20	0.73	1.31	0.62
Remaining Fuel	1.04	0.50	1.52	0.71
Tone Controls	0.92	0.41	1.73	0.82
Info. Lights	0.83	0.35	2.12	1.16
Destination Distance	1.06	0.56	1.73	0.93
Fan	1.10	0.48	1.78	1.00
Balance	0.86	0.35	2.59	1.18
Sentinal	1.01	0.47	2.51	1.81
Defrost	1.14	0.61	2.51	1.49
Fuel Economy	1.14	0.58	2.48	0.94
Correct Direction	1.45	0.67	2.04	1.25
Fuel Range	1.19	1.02	2.54	0.60
Cassette Tape	0.80	0.29	2.06	1.29
Temperature	1.10	0.52	3.18	1.66
Heading	1.30	0.56	2.76	1.81
Zoom Level	1.40	0.65	2.91	1.65
Cruise Control	0.82	0.36	5.88	2.81
Power Mirror	0.86	0.34	6.64	2.56
Tune Radio	1.10	0.47	5.91	2.39
Cross Street	1.66	0.82	5.21	3.20
Roadway Distance	1.53	0.65	5.78	2.85
Roadway Name	1.63	0.80	6.52	3.15

Note: Glance length given in seconds.

Table A-2. Truck Driver Visual Allocation Data to mirror and instrument panel locations. (Source: Tijerina, Kantowitz, Kiger, and Rockwell, 1994).

Command	**No, of Trials**	**Total No, of Glances**	**Average Glance Duration** (Secs.)	**Variance of** Glance **Duration** (Secs. Sq.)	**10th %tile** Glance **Duration** (Secs.)	**90th %tile** Glance **Duration** (Seer.)	**Mean No. of Glances**	**Min. No. of Glances**	**Ma. No. of** Glances	**Average Time OK** Road* (Secs)
Left Mirror-Dctcct (3)	I7	24	1.38	0.39	0.67	2.23	1.41	1	3	1.95
Right Mirror-Detect (8)	17	26	1.22	0.27	0,57	1.73	1.59	1	4	1.94
Left Mirror-Discrimination (I 5)	12	16	1.52	0.4 I	0.30	2.17	1.50	1	3	2.28
Right Mirror-Discrimination (19)	14	26	1.45	0.38	0.73	2.43	1.86	1	3	2.69
Read Exact Speed (1)	21	27	I.60	0.28	1.00	2.40	1.29	I	2	2.06
Read Speed & Comperc to Posted Limit (1 1)	16	20	1.42	0.26	0.77	2.08	1.25	I	2	I.77
Read Air Pressure (2)	19	38	2.11	1.32	0.67	3.85	2.00	I	9	4.21
Read Engine **RPM (5)**	**18**	28	1.66	0.50	0.73	2.55	1.61	1	3	2.67
Read Fuel Gage (16)	18	32	1.88	0.50	0.75	2.77	1.78	I	4	3.34
Read Clock (9)	17	3 2	1.20	0.28	0.48	I.77	1 .88	1	7	2.25
Read Elapsed Time (20)	12	32	1.65	0.27	0.98	2.33	2.67	1	6	4.40
Radio Volume Up/Down (4)	34	55.	1 . 1	0.18	0.40	1.47	1.62	I	3	1.78
Scicct Preset Station (17)	16	51	1.46	0.50	0.63	2.50	3.19.	1	7	4.65
Tune Radio to 90.5 (18)	16	125	1.77	0.4 I	0.97	2.67	7.81	3	18	13.81
Change CB Frcqucncy (G)	33	122	1.34	0.22	0.73	2.00	3.76	2	7	5.04
Turn CB Volume Up/Down (7)	24	31	1 .06	0.14	0.50	I 53	1.29	1	3	1.37
AC Temp Up/Down (21)	5	12	1.65	0.51	0.80	2.57	2.40	1	4	3.97
Fan Speed Higher/Lower (22)	7	12	1.35	0.23	0.62	1.90	1.71	1	**3**	2.3 I

•Product ** Glance * Number of Glances**

- Fovea1 vision is considered important for many aspects of driving and crash avoidance like sign reading and object and event detection. On the other hand, peripheral vision may be primary in detecting relative motion, which is also an important aspect of hazard detection (Liebowitz and Owens, 1986; Shiff and Arnone, 1995).

- The driver's gaze usually, but not always, indicates where the driver's attention is focussed.

These caveats suggest that visual allocation measures may be useful but should be augmented by other measures of attentional demand and intrusion on the driving task caused by in-cab devices. It is also important to keep in mind that there are other eye movement measures that are not included in the definition of visual allocation measures. These include measures such as pupil diameter, blink rate, eyelid closure, slow eye movements (SEMs), and others. Such measures have their uses (e.g., in drowsy driver detection) but are not considered to reflect the direct acquisition of visual information for safe driving.

INSTRUMENTATION NEEDS

The measurement of eye movements can be accomplished by a variety of methods (see Young and Sheena, 1975 for a review of basic methods). These include electro-oculographic (EOG) methods, pupil-center-corneal-reflection techniques, and film or video of the driver's face and eyes. Eye movement or visual allocation techniques for use in vehicles should ideally: a) allow the driver to use normal visual scanning strategies; b) allow the driver a full range of free head and upper-body movements; c) operate under various lighting conditions (day and night) and the vibration environment of the vehicle; d) provide sufficient resolution on where the driver is looking; and e) be reducible by automatic means. While innovative instrumentation options are currently under development (cf. Hagiwara and Zwahlen, 1995), no visual allocation system is currently available to meet all of these needs. In particular, head-mounted systems are of concern from the standpoint of driver acceptance and restricted field of view. The performance of systems that depend on infrared light sources (e.g., pupil-center-corneal reflection techniques) is degraded in bright daylight. Sensors that track head position are currently expensive to procure. There are conditions were high resolution of fixation location is required (1 degree of visual angle or less). Examples include the need to determine where a driver is looking among closely spaced instruments or among data items in a visual display or a head-up display (HUD). Currently, no instrumentation (known to the authors) is available for use in a operational environment that provides such resolution without encumbering the driver to at least some extent.

At present the simplest and most reliable means to collect visual allocation data is by means of videotape with manual data reduction. It is tedious work, but effective and so the instrumentation needs for the video method with manual data reduction are described next. The fact that manual data reduction is most common now should in no way detract from the need to develop computer-assisted data reduction (and enhanced precision in data capture) in the future.

The collection of driver visual allocation data on-the-road requires a variety of instrumentation that can be grouped into two systems: a data capture system and a data reduction system. Key components are described below.

DATA CAPTURE SYSTEM

The following components constitute parts of the data capture system for visual allocation data gathering.

- A power source is needed for all equipment. This power source should be conditioned to minimize equipment malfunction and data loss or inaccuracies due to power fluctuations.
- A camera must be mounted so that it may be directed toward the driver' s face. This view is required to record the driver's visual glances during the data collection run. The camera should be equipped with a lens so focussed as to provide a clear image of the driver's face such that minor head movements do not cause the driver's face to be lost from the recorded image. It is important to position the camera and its mount in such a location that their presence does not affect the driver's view of the driving scene.

For night data collection, there will also be a need for an infrared light source to illuminate the driver's face.

- A recording system is needed, such as a video cassette recorder (VCR). The VCR should use high-quality video tape for good resolution. The recording system should be set to run at the fastest recording speed, if possible, for best picture resolution.
- A video monitor is needed to examine the quality of the recorded image.
- A calibration video must be made wherein the driver is asked to systematically look in pre-specified locations. Periodically, recalibation video should be taken to aid in the data reduction.
- A time-code generator is needed that superimposes time information on the view of the driver's face prior to recording on the VCR. The device should provide a high-speed elapsed time clock with resolution equal to the video frame rate (e.g., l/30th of a second for one video frame assuming NTSC standard video at 30

frames per second; 1/25 of a second for one video frame assuming PAL standard video at 25 frames per second).

- It is advisable to have additional cameras and VCRs to capture the road scene ahead and in-cab activities. If so, it is recommended that there be an additional VCR and a four-into-one video splitter. One VCR can record the driver' s face. The second VCR can be used to record a split screen view of, for example, the driver's face, the road scene ahead, the in-cab scene, and an additional camera view of the driving situation. The time code generator must superimpose the same time code on both recordings. Finally, a video switcher is required if the experimenter wishes to periodically view each video recording on the video monitor to ensure proper camera aim following in-route seat adjustments or postural changes made by the driver.

- Micronhones can be readily interfaced into the video data capture system and audio recordings may be made. This option should be considered to capture driver comments or experimenter comments.

- Auditory or visual event markers can be recorded on the video tape to facilitate cuing during data reduction.

Figure A-l presents a schematic of one data capture system. The configuration depicted includes two cameras oriented to the road scene ahead, one camera to the driver's face (the "gaze" camera) and one in-cab camera oriented to capture hand movements off of the steering wheel (the "hands-on-wheel" camera).

DATA REDUCTION SYSTEM

The data reduction system is used to take visual allocation data from the videotape for subsequent analysis. The components of such a system, suitable for manual data reduction, are presented below. Note that advances in automatic image processing may automate much of the data reduction involved, though manual data reduction is currently the most common and reliable method used.

- A professional editing video cassette recorder or editing; deck is used for playback. The VCR or editing deck should not suppress the audio recording during slow motion or search-speed playback. This is because the audio is needed to detect time codes and auditory event markers, if used. An additional desirable feature is the ability to enter in a time code for automatic search and cuing. The unit should allow for frame-by-frame advance.

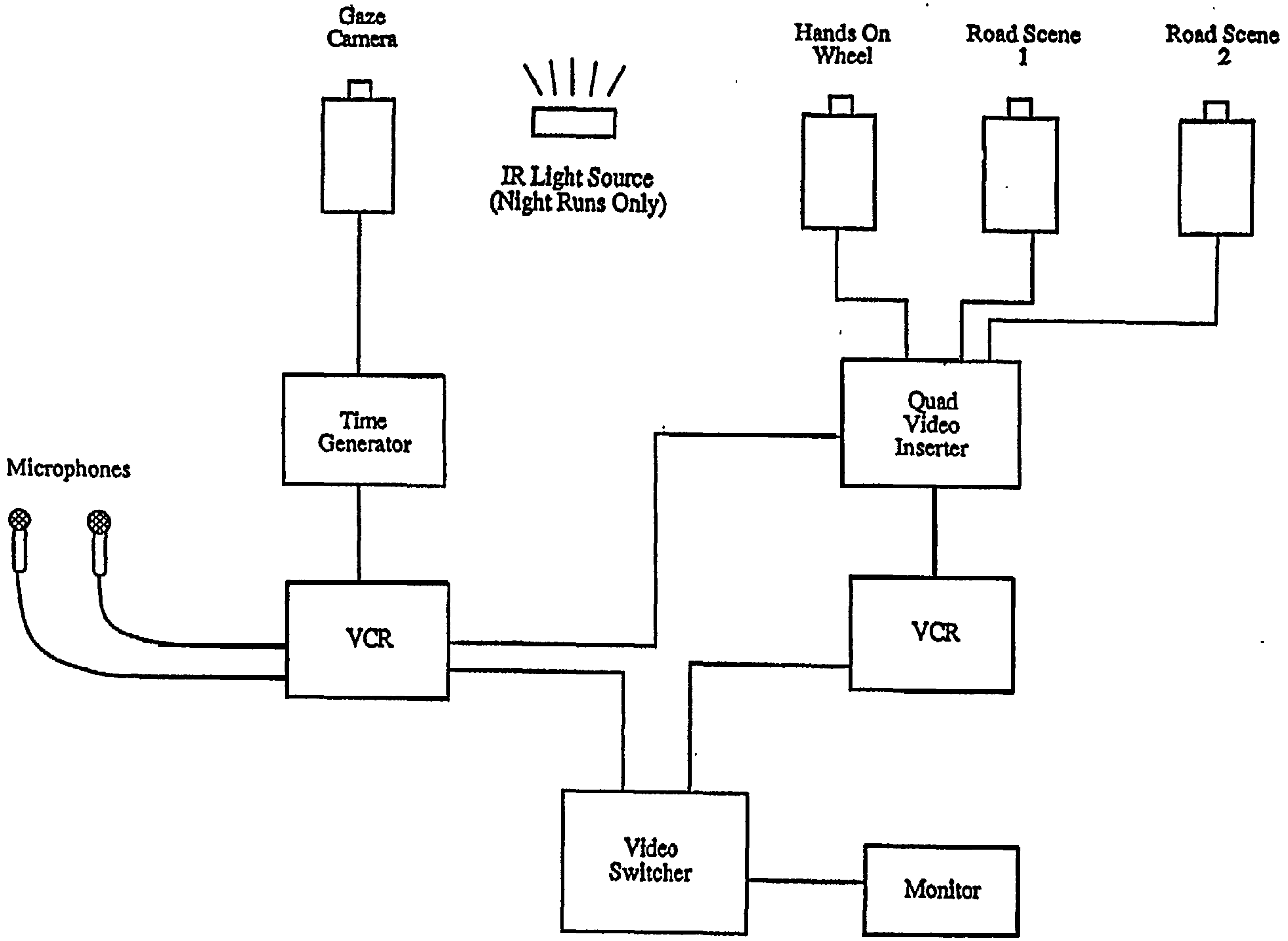


Figure A-1. Schematic of a Visual Allocation Data Capture System for Instrumented Vehicle Use.

- A <u>video playback monitor</u> is needed to allow the data reducer to review the video tape.

To expedite data reduction, it is advisable to develop a computer interface that can accept the start and stop times for each glance to a location as well as the location code entered by the data reducer. The additional equipment required is listed below:

- A <u>time code reader</u> is needed, interfaced to a PC, to capture the start and end time codes for each glance.

- A PC with custom software is needed to read time codes directly from the videotapes and store them in a database along with the location code selected by the data reducer.

The following terms are defined for the data reducer:

Sample Interval:	A time period that constitutes a sample of interest (e.g. an in-cab task) of the videotape for data reduction. Usually, this will be the time associated with an event.
Frame:	The basic unit of observation for data reduction. The data reducer examines a video display frame by frame, to determine the driver eye fixations.
Fixation Location:	Where, in a pre-defined mapping of areas, the driver is looking in a given frame. As was mentioned earlier, a calibration video must be recorded wherein the driver is asked to look at pre-specified locations so that the data reducer may allocate fixations across locations from frame to frame reliably. Furthermore, different locations must have some minimum spatial separation to be distinguishable on the videotape.
Transition:	A change in eye fixation location from one defined fixation location to another, different, fixation location.
Transition pair:	The From-To pair of fixation locations in a given transition.
Gaze Shift:	A change in the driver's eye point-of-regard, in a given frame, that is between pre-defined fixation locations.

Given these background definitions, the following procedure should be followed:

1) The data reducer advances the videotape to the start of a sample interval of interest.

2) The data reducer examines the first frame of the driver's face and determines the driver's fixation location, then enters that location code and the starting time for that fixation.

3) The data reducer advances the video tape, frame by frame, until the driver's eyes move to another location. When this occurs, the data reducer enters the new fixation location code and the time code for that frame. The data reducer also indicates that this is the first transition pair from the first location (e.g., location j) to the second location (i.e., location j).

4) If one or more frames indicate gaze shifts (i.e., the driver's eyes are in motion and between defined fixation locations), the data reducer may select one of the following options:

   - The frame(s) may be deleted. This is suitable if the analysis does not require that all of the time in the sample interval be accounted for. For example, an analysis of mean glance durations and glance frequencies would not require all of the sample interval time to be accounted for.

   - Allocate the frames containing gaze shifts to the original fixation location until the new fixation location is reached. This convention is based on plausible assumptions that a) the driver is still processing information just picked up from the original fixation location and b) does not begin picking up and appreciating information from the new fixation location until the eyes are on the location and have re-accommodated or refocussed. This option will allow for all of the sample interval time to be accounted for, subject to a small bias introduced in the glance duration data which may overstate glance duration to some extent.

   - Collect the time required for gaze shifts explicitly for analysis of transition times, times required to shift the eyes from one location to another. This will also account for the total sample interval.

5) The previous steps are repeated frame-by-frame until the sample interval has been fully reduced.

FUNDAMENTAL DATA

Given the reduced data, it is possible to determine three fundamental measures: glance durations, glance frequencies, and transition pairs. Each of these is operationally defined below.

Glance:	A series of consecutive fixations (frames) on the same location. A glance is indicated by the same fixation location across multiple consecutive video frames.
Glance Duration:	The time that a driver's eyes are stationary (disregarding small movements) on a single fixation location. This is taken as the time interval from when the driver first fixates on a location until the driver's eyes shift to a different location.
Number of Glances:	The total number of glances to a particular location in the sample interval, where each glance is separated by at least one glance to a different location.
Transition Pair:	A change in eye fixation from location j to location k where j is not equal to k.
Transition Time:	The time interval required for the eyes to move from location j to location k. This time interval is essentially linear with distance traveled during the gaze shifts. Hayes, Kurokawa, and Wierwille (1989) report that transition time also increases with age and, for their study, averaged between 100 ms (for drivers 18 to 25 years of age) and 125 ms (for drivers 49 to 72 years of age).

From these fundamental measures, the visual allocation measures of performance (MOPs) in Table A-3 can be derived. The table consists of the following elements:

- Operational Definitions of each MOP
- Workload interpretation, i.e., a prediction of how the MOP should vary with increased workload.

The analysis of the MOPs may be conducted using a variety of statistical techniques. These range from t-tests and ANOVAs on mean values to Chi-square tests for homogeneity of proportions for fixation probability data, to multivariate procedures (MANOVA, cluster analysis, regression), to exploratory graphical data analysis techniques. The references included at the end of this appendix provide examples of various analysis procedures.

REFERENCES

Allen, R. W., & McRuer, D. T. (1977). *The effect of adverse visibility on driver steering performance in an automobile simulator* (SAE Technical Paper No. 770239). New York: Society of Automotive Engineers.

Antin, J. A., Dingus, T. A., Hulse, M. C., & Wierwille, W. W. (1990). An evaluation of the effectiveness and efficiency of an automobile moving-map navigational display. *International Journal of Man-Machine Studies*, *33*, 581-594.

Bhise, V., Forbes, L., and Farber, E. (1986). *Driver behavioral data and considerations in evaluating in-vehicle controls and displays*. Paper presented at the Transportation Research Board 65th Annual Meeting, Washington, DC., January.

Boersema, T., Zwaga, H. J. G., & Adams, A. S. (1989). Conspicuity in realistic scenes: An eye-movement measure. *Applied Ergonomics*, *20*(4), 267-273.

Dingus, T. A., Antin, J. F., Hulse, M. C., & Wierwille, W. W. (1989). Attentional demand requirements of an automobile moving-map navigation system. *Transportation Research*, 23A(4), 301-315.

Hagiwara, T., & Zwahlen, H. T. (1995). *A proposed new driver eye scanning behavior recording and analysis system* (Paper No. 950948). Paper presented at the Transportation Research Board 74th Annual Meeting, Washington, DC, January.

Hayes, B. C., Kurokawa, K., & Wierwille, W. W. (1989). Age-related decrements in automobile instrument panel task performance. *Proceedings of the Human Factors Society 33rd Annual Meeting*, 159-163.

Hughes, P. K., & Cole, B. L. (1988). The effect of attentional demand on eye movement behavior when driving. In A. G. Gale et al. (Eds.), *Vision in vehicles II* (pp. 221-230). Amsterdam: Elsevier North-Holland.

Kito, T., Haraguchi, M., & Takayuki, F. (1989). Measurements of gaze movements while driving. *Perceptual and Motor Skills*, *68*, 19-25.

Kurokawa, K, & Wierwille, W. W. (1990). Validation of a driving simulation facility for instrument panel task performance. *Proceedings of the Human Factors Society 34th Annual Meeting*, 1299-1303.

Liebowitz, H. W., & Owens, D. A. (1986). We drive by night. *Psychology Today*, *20*(1), 54-58.

Louma, J. (1988). Driver's eye fixations and perceptions. In A. G. Gale et al. (Eds.), *Vision in vehicles II* (pp. 231-237). Amsterdam: Elsevier North-Holland.

Mourant, R. R., Herman, M., & Moussa-Hamouda, E. (1980). Driver looks and control location in automobiles. *Human Factors*, *22*(4), 417-425.

Mourant, R. R., & Rockwell, T. H. (1970). Mapping eye-movement patterns to the visual scene in driving: An exploratory study. *Human Factors, 12*(1), 81-87.

Rockwell, T. H. (1972). Skills, judgement, and information acquisition in driving. In. T. W. Forbes (Ed.), *Human factors in highway traffic safety research* (pp. 133 - 164). New York: Wiley Interscience.

Rockwell, T. H. (1988). Spare visual capacity in driving - revisited. In A. G. Gale et al. (Eds.), *Vision in vehicles II* (pp. 317- 324). Amsterdam: Elsevier North-Holland.

Senders, J. W., Kristofferson, A. B., Levison, W. H., Dietrich, C. W., & Ward, J. L. (1967). The attentional demand of automobile driving. *Highway Research Record, 195*, 15 - 32.

Shiff, W., & Arnone, W. (1995). Perceiving and driving: Where parallel roads meet. In J. M. Flach, P. A. Hancock, J. K. Caird, & K. J. Vicente (Eds.), *Local applications of the ecological approach to human-machine systems* (pp. 1-35). Hillsdale, NJ: Erlbaum.

Tijerina, L., Kantowitz, B., Kiger, S. M., & Rockwell, T. H. (1994). Driver workload assessment of in-cab high technology devices (Paper No. 94-S2-O-060). *Proceedings of the Fourteenth International Technical Conference on Enhanced Safety of Vehicles*, Munich, Germany, May 23 - 26, 1994.

Young, L. R., & Sheena, D. (1975). Eye-movement measurement techniques. *American Psychologist, 30*, 315-330.

Wierwille, W. W. (1981). Statistical techniques for instrument panel arrangement. In J. Moraal and K.-F. Kraiss (Eds.), *Manned systems design: Methods, equipment, and applications* (pp. 201-218). New York: Plenum Press.

Wierwille, W. W. (1993). Visual and manual demands of in-car controls and displays. In B. Peacock and W. Karwowski (Eds.), *Automotive ergonomics* (pp. 299 - 320). London: Taylor and Francis.

Wierwille, W. W., Hulse, M. C., Fischer, T. J., & Dingus, T. A. (1988). *Strategic use of visual resources by the driver while navigating with an in-car navigation system* (SAE Technical Paper No. 885180). New York: Society of Automotive Engineers.

Table A-3. Visual Allocation Measures of Performance (MOPs) used for Driver Workload research. Note: All measures are taken within a sample interval and, unless otherwise noted, are defined with respect to a given fixation location j (e.g., in-cab device, roadway, mirrors, etc.), for a single driver.

Number of Glances$_j$ =	Total number of glances to location j, where each is separated by at least one glance to a different location..
Workload Interpretation:	The number of glances needed to complete a transaction reflects the complexity of the in-cab task as a whole, i.e., the number of task components (Kurokawa and Wierwille, 1990). Thus, the greater the workload demanded by a location (e.g., device, road scene), the greater the glance frequency.
Mean Glance Duration$_j$ =	$\dfrac{\left(\sum_{i=1}^{n} Glance\ Durations(i)\right)}{Number\ of\ Glances_j}$ The mean glance duration to location j is the sum of all glance durations to location j divided by the number of glances to location j in the sample interval.
Workload Interpretation:	The average length of a single glance reflects the difficulty of a task component (Kurokawa and Wierwille, 1990). Subject to the constraint that most drivers will not take their eyes off the road for more than perhaps 2.0 to 2.5 s, longer glances, the greater the workload demand, the longer the mean glance duration. Note that glance frequency and glance duration may trade off within a fixed sample interval. That is, very long glance durations (indicative of high workload demand) may be associated of with fewer rather than more glances. Thus, it is important to consider the two measures together, especially if the sample interval is fixed rather than allowed to reflect task completion time.

Table A-3. (Continued). Visual Allocation Measures of Performance (MOPs) used for Driver Workload research. Note: All measures are taken within a sample interval and, unless otherwise noted, are defined with respect to a given fixation location j (e.g., in-cab device, roadway, mirrors, etc.), for a single driver (Continued)

Total Glance Time$_j$ =

$$\sum_{i=1}^{n} Glance\ Duration_j(i)$$

i.e., total glance time to fixation location j is the sum of all glance durations to fixation location j in the sample interval.

Proportion Total Glance Time$_j$ = [Total Glance Time$_j$ / Sample Interval]

Workload Interpretation: The total glance time (or percentage of time) associated with a fixation location j (e.g., in-cab device) provides another measure of the visual demand posed by that location. The percentage measure may be used when there is a need to normalize total time measures based on the length of the sample interval. As workload demand increases, total time and percent time should increase.

Key measures that should be considered for driver workload assessment include Total Glance Time and Proportion of total Glance Time to the following key locations: on-road, on-mirrors, in-cab (i.e., on an in-vehicle device).

Table A-3. (Continued). Visual Allocation Measures of Performance (MOPs) used for Driver Workload research. Note: All measures are taken within a sample interval and, unless otherwise noted, are defined with respect to a given fixation location j (e.g., in-cab device, roadway, mirrors, etc.), for a single driver (Continued)

Mean Transition Time$_{jk}$ =

$$\sum_{i=1}^{n} \frac{gaze\ \ shift_{jk}(i)}{n_{ik}}$$

where

gaze shift$_{jk}$(I) is the transition time for the eyes to shift gaze from location j to location k for transition I;

n_{jk} = number of transitions from location j to location k in the sample interval.

i.e., mean transition time is the sum of the gaze shift times to move the eyes from location j to location k, divided by the number of such gaze shifts in the sample interval.

Workload Interpretation: Transition times are roughly a linear function of the distance from location j to location k. During the transition gaze shift, there is relatively little new visual information available to the driver. Thus, increased mean transition times reflect reduced time available for driver information gathering.

Table A-3. (Continued). Visual Allocation Measures of Performance (MOPs) used for Driver Workload research. Note: All measures are taken within a sample interval and, unless otherwise noted, are defined with respect to a given fixation location j (e.g., in-cab device, roadway, mirrors, etc.), for a single driver (Continued)

Fixation Probability, p_i:

$$p_j = \frac{\text{number of frames with gaze on location } j}{\text{total number of frames in sample interval}}$$

i.e., Fixation probability is the probability that location j was fixated on during a sample interval.

Workload Interpretation: The fixation probability on a given location reflects the relative attentional demand associated with that location. Across a mutually exclusive and exhaustive set of locations, fixation probabilities capture where the eyes were fixated throughout a sample interval. Given such a distribution, workload assessment might statistically compare two such distributions (under two different task types, for example). For example, if device use induced a relative decrease in the fixation probabilities associated with the driving scene (e.g., road scene, rear-view mirrors), this would be considered safety relevant and indicative of the workload demand associated with the device.

Table A-3. (Continued). Visual Allocation Measures of Performance (MOPs) used for Driver Workload research. Note: All measures are taken within a sample interval and, unless otherwise noted, are defined with respect to a given fixation location j (e.g., in-cab device, roadway, mirrors, etc.), for a single driver (Continued)

Link Value Probability, p_{Ljk} : This is a measure of the strength of association between location j and location k. According to Wierwille (1981; see also Antin, Dingus, Hulse, and Wierwille, 1990) ,the link value probability between location j and location k is:

$$p_{Ljk} = \frac{\frac{n_{jk}}{N} + \frac{n_{kj}}{N}}{N - \sum_{j=1}^{Q} \frac{n_{jj}}{N}}$$

where

n_{jk} = the number of transitions from location j to location k, j not equal to k.
n_{kj} = the number of transitions from location k to location j, k not equal to j.
n_{jj} = the number of transitions from location j to location j (i.e., successive frames where the driver' s fixation location remains the same).
N = the total number of transitions (across all locations, not just j and k) in the sample interval.
Q = the number of unique fixation locations.

It should be noted that p_{Ljk} is only defined for $j < k$. Thus, the number of link probabilities for a situation in which there are Q locations is given by [Q(Q-1)]/2.

Workload Interpretation: The link value probabilities represent the relative number of transitions between one location and another and, thus, the strength of relationship between one location and another. The greater the link value probability, the stronger is the need to time-share attention between the two locations. In workload assessment, the link value probabilities may be analyzed to assess how visual attention has been affected by an in-cab device use or driving conditions.

APPENDIX B

STEERING, PEDAL, AND MANUAL ACTIVITY IN DRIVER WORKLOAD ASSESSMENT

APPENDIX B. STEERING, PEDAL, AND MANUAL ACTIVITY IN DRIVER WORKLOAD ASSESSMENT

MOTIVATION

The driver relies heavily on visual information to safely control the vehicle. This control is accomplished by manipulation of vehicle controls, i.e., the steering wheel, the accelerator, the brake, and (less often) the transmission. It is plausible to hypothesize that in-vehicle device use can disrupt the driver's control activities. This hypothesis implies that measures of steering, accelerator, and brake activities are candidate driver workload measures. Furthermore, such measures are safety-relevant to the extent that driver control inputs, mediated through the dynamics of the vehicle and driving condition factors, affect driver-vehicle performance measures of lane keeping, speed maintenance, and car following.

Steering behavior has been used to assess primary driving task difficulty and the impact of including a secondary task (e.g., an in-cab device task), as well as the effects of fatigue. For example, Safford and Rockwell (1967) found that over a 24-hr period steering reversal rates increased relative to the first hours of driving. Wiener, Curry, and Faustina (1984) similarly reported that lack of sleep led to a statistically significant increase in steering reversals in a simulated driving task. The interpretation attached to such effects is that fatigue leads to poorer or more erratic steering performance. Drory (1985) conducted a simulator study of fatigue in truck drivers and found that, compared to just driving, steering wheel reversal rates decreased with the addition of secondary tasks; this was taken as indicating the beneficial effects of in-cab tasks to offset driving fatigue effects. Studies such as these suggest that steering performance measures may be indicative of driver drowsiness and recent drowsy driver detection research has made extensive use of steering measures as indicators of driver fatigue (Wierwille, Wreggitt, and Mitchell, 1992; Wierwille, 1994). Specific predictions are that drowsiness involves periods where there is a lack of steering activity followed by abrupt, large steering corrections. It is this 'drift and jerk' steering strategy that is said to be one characteristic of drowsy drivers. As will be seen below, inattention to the driving task is also said to be indicated by periods with little or no steering activity followed by large steering corrections.

McLean and Hoffman (1975) proposed that steering wheel reversals may serve as a sensitive measure of primary driving task difficulty. They reported that steering reversals (defined as the number of times the direction of steering wheel movement is reversed through a finite angle or "gap") increased as sight distance was decreased and lane width was decreased, manipulations that effectively increase driving task difficulty. In a test track study by MacDonald and Hoffmann (as cited in MacDonald and Hoffmann, 1980), it was later found that steering wheel reversal rate increased with narrowing of lanes but decreased when drivers also had to perform a secondary task. However, a subsequent on-the-road study by MacDonald and Hoffmann (as cited in MacDonald and Hoffmann, 1980), was conducted on suburban roads and revealed that steering reversals increased with the presence of a secondary task and were lowest on the roadway segments with the most events (where events included other traffic). MacDonald and Hoffmann (1980) later elaborated on the relationship between steering wheel reversals, driving task demand, and other demands. Seemingly contradictory results might be explained by the following assumptions:

- When driving plus in-cab task demands are within the driver's capacity to cope, the driver copes with an additional task by increasing effort and this effort is reflected in higher steering reversal rates;
- When driving plus in-cab task demands match or exceed the driver's capacity to cope, the driver manages attention in such a way that less attention is available for the steering task and this is reflected in a decrease of steering reversal rates.

A test of in-vehicle attentional demand has been made using these assumptions by Dingus, Antin, Hulse, and Wierwille (1989). They found that the steering velocity measure that showed the greatest sensitivity to in-vehicle task differences was the variance of time the steering velocity measure was zero. The logic was given as follows. Normal alert driving is characterized by use of small, relatively uniform steering corrections to maintain proper lane position. As attentional demand increases past a point where attention is drawn away from the driving task, these small corrections cease and the steering wheel is held constant for some period of time (MacDonald and Hoffmann, 1980). Thus, normal small corrections would be followed by holds in steering, causing variance of the times that the steering velocity is zero during a task to increase. Indeed, the most demanding in-vehicle tasks (as measured by task completion time, visual demand, and other indicators) were those with the greatest variance in the time the steering velocity was zero. Similar results were found when the percent of task time that the steering velocity was zero was used as a response measure (Dingus, Antin, Hulse, and Wierwille, 1986). Antin, Dingus, Hulse, and Wierwille (1990) reported (without presenting numerical results) that a measure, percent steering-zero, was sensitive to alternative route guidance aids and that an electronic moving-map route guidance system induced more steering holds than a paper map. Thus, steering measures have proven to be useful in driver workload assessment as well as driver fatigue studies.

More recently, Verwey (1991) reported on the use of a variable called Steering Wheel Action Rate (SAR) to assess workload imposed by a secondary task in an on-the-road study with an instrumented vehicle. SAR was the number of steering actions per second. It was found to increase with the introduction of either auditory or visual secondary tasks in inexperienced drivers. It only increased with visual secondary tasks performed by experienced drivers. According to the earlier assumptions of MacDonald and Hoffman (1980), such secondary tasks must have been within the capacities of the drivers in the context of the primary driving task.

Burnett and Joyner (1993) evaluated the distraction potential of a map-based route guidance system, route instructions, and written notes/maps. The authors report (without presentation of numerical results) that steering wheel variability was greater for the route-guidance system than the other two (baseline) conditions, but only for a certain portion of the experimental route. No differences were found upon approach to a critical exit, though this might be accounted for by factors such as the acquisition of information on the need to exit prior to reaching the exit itself.

Most recently, Dingus and his colleagues have used steering measures in the assessment of the TravTek advanced traveler information system, the largest field study to date on an Intelligent Transportation System (ITS). Dingus, Hulse, Fleischman, McGehee, and Manakkal (in press) examined the effects of age and navigation technique on various aspects of driving by means of assessing the number of large steering reversals (i.e, reversals that exceed 6 degrees). Recall that as workload or attentional demand increases with an in-vehicle task, the frequency of steering corrections tends to decrease. Since the small centering corrections (e.g., 2 degrees to 6 degrees of steering angle) decrease the vehicle tends to drift farther from the lane center and a larger steering input is required to correct the vehicle position error. Thus, large steering inputs (e.g., greater than 6 degrees) increase. Results indicated that the drivers 65 years and older had significantly more large steering reversals per unit time than either drivers 16 to 18 years of age or drivers 35 to 45 years of age. This might be taken to indicate that older drivers were less able to manage additional information processing demands relative to the younger drivers despite more cautious driving behavior, as indicated by other driving measures (e.g., lower travel speed overall).

Dingus, Mollenauer, Hulse, McGehee, and Fleischman (in press) assessed the effects of experience and navigation configuration on driver performance. Results here indicated a decrease in the number of large steering reversals per unit time between local user's first drive and second drive across all navigation configurations (e.g., turn-by-turn iconic visual display with voice call out, turn-by-turn iconic visual display only, electronic map with voice call out, electronic map without voice, written directions, or paper map). This was taken as an indication that with experience, drivers were able to keep their eyes on the roadway a greater proportion of the time and so had to make fewer large corrections in steering.

As these selected studies illustrate, steering measures can be sensitive measures of driver workload. However, steering activity is distinct from driving performance as measured by measures of vehicle heading angle or lane position (MacDonald and Hoffmann, 1980). Heading angle at a given time is roughly the integral of steering position over time and lane position (on a straight roadway, at least) is roughly the time integral of heading angle over time. Thus, vehicle position in the lane is two time integrations away from a steering input. However, steering input is a worthwhile data channel to capture because it may be more sensitive than lanekeeping measures.

On the other hand, steering activity is influenced by many sources simultaneously and therefore may be difficult to interpret. Typical influences in addition to workload are road conditions (e.g., bumpy versus smooth), driver style (e.g., relaxed versus tense), and lack of constraints (there are many steering strategies that allow the vehicle to assume acceptable trajectories). Consequently, researchers or evaluators should be cautious in drawing conclusions about the relationship between workload and steering activity. Correlations may exist but their values may be small.

In addition to steering inputs, accelerator inputs may be assessed. Like steering inputs, driver workload assessment by analysis of throttle inputs is oriented toward indications of intermittent open-loop driving. When a driver's attention is drawn away from the driving task, there is a tendency to maintain the foot in the same position (Dingus, Hulse, Fleischman, McGehee, and Manakkal, in press). Alternatively, the driver may release the accelerator pedal altogether as a preliminary attempt to slow the vehicle down (Dingus, personal communication, April, 1995). When the driver realizes that he or she is going (generally) too slow, the accelerator is depressed to a greater degree than is usual for a normal or continuous adjustment. Thus, accelerator pedal holds, mean hold duration, and variance or standard deviation of accelerator pedal position, as well as number of accelerator pedal releases and total (or percentage) pedal release time appear to be promising indicators of workload . Dingus, Antin, Hulse, and Wierwille (1989) found that accelerator measures were not sensitive to variations in in-vehicle tasks executed while driving in an instrumented vehicle. Antin, Dingus, Hulse, and Wierwille (1990) also reported that there were no differences in accelerator measures (or brake usage) across moving-map electronic displays, paper map, or memorized route conditions. Verwey (1991) found that frequency of depressing the accelerator (as well as brake and clutch pedals) was not affected by executing secondary tasks. Such reports are common in the small body of literature that reports on attempts to use accelerator pedal activation. On the other hand, Dingus (1995, personal communication) has recently completed a study of collision warning systems and reports that warning onset reliably prompts accelerator releases, one indication of driver attention to the warning. Despite the apparent reasonableness of pedal measures, it should be noted that less is known about throttle inputs than about steering inputs for driver workload assessment.

The attentional connection to brake actuations is most evident in brake reaction time. If the driver is attending to an in-vehicle device, this may increase the reaction time to activate the brakes. The number of brake activations might be expected to increase under conditions of high attentional demand if the driver realizes this fact and adopts the strategy of riding the brakes more than usual to support a quick response to an unexpected situations (Dingus, Antin, Hulse, and Wierwille, 1986). By similar logic, the average dwell time per brake application might increase as well. Monty (1984) reported that the number of brake activations and dwell time per brake application were sensitive to the attentional demand of various in-vehicle tasks while driving. On the other hand, Dingus, Antin, Hulse, and Wierwille (1989) did not find brake pedal measures sensitive to differences in road types, or specific tasks asked of drivers while on the road. Verwey (1991) also failed to find brake pedal use sensitive to various secondary task conditions. On the other hand, reaction time measures, especially in simulator studies, have been used to attempt to discriminate among the workload of various experimental conditions. Noy (1990) used brake reaction time in a simulator study and found no statistically reliable relationship between brake reaction time and gaze direction (looking inside or outside during the onset of the deceleration). Noy speculates that this lack of an effect might have been due to the small number of braking events in the simulator. Since on-the-road studies cannot safely include staged events that prompt braking, simulator studies may be the most appropriate means to collect such data in a systematic fashion.

A caveat similar to that presented for steering measures is in order for pedal activity. Pedal actuations are a function of many aspects of vehicle control and should therefore be interpreted with caution in regard to workload influences.

INSTRUMENTATION NEEDS

The instrumentation needed for capture of driver control inputs and manual activity are discussed below.

SENSOR SUITE

Steering Sensors. There are several options for the collection of steering data. These include steering position string potentiometers wrapped around the steering shaft, steering pitman arm position stringpot or DCDT (DC differential transformer) or linear potentiometer, and turn sensor assemblies (rotary encoders) that attach to the steering column and measure angles by various means. Given a clean steering wheel position channel, steering velocity can be determined by means of numerical methods, e.g., 5-point numerical differentiation. It is usually better to obtain steering velocity from a velocity (rate) sensor on the steering column. This device is simply a tach generator whose output is proportional to rotational velocity. It should also be noted that pitman arm sensors pick up a substantial amount of road disturbance and so must be filtered judiciously to eliminate such noise.

Accelerator Sensors. A linear potentiometer attached to the accelerator can be used to measure the percent of pedal throw (0% for pedal release to 100% for throttle in full open position). Given a clean accelerator pedal channel, accelerator pedal rate can be determined by numerical differentiation of the position signal. There are also devices available that are capable of providing output proportional to linear velocity of movement.

Brake Pedal Sensors. The simplest sensor is a simple ON/OFF switch (e.g., the signal to the brake lights would work). In addition, a pressure transducer would be useful to measure the percent (or actual pounds per square inch) of force applied in a given braking maneuver.

Manual Activity Sensors. The assessment of the manual resources needed (and available) for in-vehicle device use is a potentially important part of driver workload assessment. Manual activity can be assessed through video cameras, pressure transducers, or capacitive sensitive switches. If a video tape approach is used, the same instrumentation is required as that listed in the Appendix of Visual Allocation Measures in Driver Workload Assessment (Appendix A).

Data Sampling, Range, and Resolution. A sample rate of 30 samples per second is ample for driver workload assessment. The measurement range should be ± 500 degrees with 1.0 degree accuracy for steering position. The measurement range should be ± 6 inches with accuracy to within 0.1 inch if pitman arm DCDT sensors are used. For Accelerator position with either DCDT (DC differential transformer) or linear potentiometer, sensor range should cover a pedal throw of 0-6 inches with resolution of 0.05 inches. Brake activation by means of brake switch or brake light power should handle a voltage input of 0-15 V DC and provide on/off status. On the other hand, measurement brake pressure by means of a pressure transducer should have a range of 0-1000 psi and a resolution of 5.0 psi. If manual activity is captured by means of video tape of the vehicle interior, this should be run at the highest resolution practicable, with a frame rate compatible with NTSC (30 frames per second) or PAL standards (25 frames per second).

DATA CAPTURE AND CONTROL

The efficient capture of sensor data for driver behavior measures (as well as driver-vehicle performance measures) is best managed by means of a computer on-board the instrumented vehicle. Two possible options for data capture and control of sensor data are Pulse Code Modulated (PCM) data recorders or Data Acquisition computers with analog-to-digital (A/D) converters. The PCM option provides the highest bandwidth, highest data storage density, and easiest means for data transport. The data acquisition computer has the advantage of converting all data to digital at the time of data collection. With this option, a set of anti-aliasing filters must be incorporated to ensure that digitized data have high fidelity. Additionally, an external storage device is essential, especially if a data collection run is to last for any appreciable length of time. Examples of mass data storage devices include high density disk drives and magnetic tape cartridges. It must be noted also that there is more limited bandwidth associated with an affordable direct-to-digital system. Table B-1 presents the advantages and disadvantages of the two options (Battelle, 1994). As with other parts of the data capture system, power must be available, conditioned to avoid data loss or error. A time-code generator is needed as the basic common reference point for all data channels.

DATA REDUCTION AND FILTERING

The analysis of steering and pedal inputs depends first on appropriate filtering and data reduction. In order to examine the data, software that allows for the simultaneous examination of multiple data streams is helpful, including an examination of a video tape. One type of system that can facilitate this type of data analysis is the Intelligent Transportation System Test Performance Assessment and Evaluation System (ITS TEST PAES) software prepared under government contract by Calspan (for general information see Gawron, 1994).

Table B-1. Comparison of Data Capture and Control Options.

Consideration	Option A - Record data on PCM data recorder		Option B - Record data directly to a data acquisition computer	
	Advantages	Disadvantages	Advantages	Disadvantages
Packaging	Integrated, modular, ruggedized system	Larger than most computers	Small size	Not a rugged as tape, disk drives subject to vibration
Power consumption	Moderate power consumption (up to 225 W), runs directly on 12 VDC		Low power, 50-100 W typical	May need power converter
User Interface	Well designed integrated interface with user help		User may build operating interface	Custom interface to be designed or implemented
Software	Debugged firmware		Flexibility limited only by programming	Software must be purchased and integrated with A/D hardware
Data storage	Data storage limited only by willingness to change tapes, up to 4 hr per tape			Data limited by hard disk size and archive transfer time
Data Bandwidth	Minimum of 88hz sample rate, built in anti-aliasing		Flexible data rate	Sample rate limited by data storage, separate anti-aliasing filters needed
Dynamic range	96 db dynamic range, (16 bit), 90 db S/N typical, inputs protected to 50 V			60-70 db typical, (10-12 bits) 50 db S/N typical inputs protected to 20 V
Cost	Low integration and operating cost	High hardware cost	Low hardware cost	High integration and operational cost

FUNDAMENTAL DATA

The following sensed data are required for driver workload measures based on steering inputs.

Steering Position: This is the steering wheel angle as a function of time. Assuming a neutral position (centered steering wheel) set to 0 degrees, then steering wheel positions to the left are in negative degrees while steering wheel positions to the right are in positive degrees. Throughout, the units of measure are (signed) degrees or radians.

High-pass Steering Position: High-pass steering position data is obtained by applying a high-pass digital filter with 0.075 Hz or approximately 0.471 radians/sec corner frequency and roll-off below the corner frequency of 20 dB/decade. The purpose of the high pass filter is to remove slow steering trends due to road curvature.

These sensed data provide the basis for the following fundamental measures:

Steering Hold: A steering hold is defined to occur when the steering wheel velocity falls within the zero dead band range for a duration of 0.4 sec or longer. See Figure B-1 for a graphical depiction of steering holds.

Steering Reversal: A steering reversal is defined to begin when the steering velocity leaves a zero dead band and ends when the steering velocity enters the zero dead band. The starting point of a reversal will be the 1st sample point that falls outside of the zero range, and the end point will be the 1st sample point that falls back into the zero range. See Figure B-2 for a depiction of Steering Reversals. Note that Figure B-2b depicts the magnitude of a steering reversal as well.

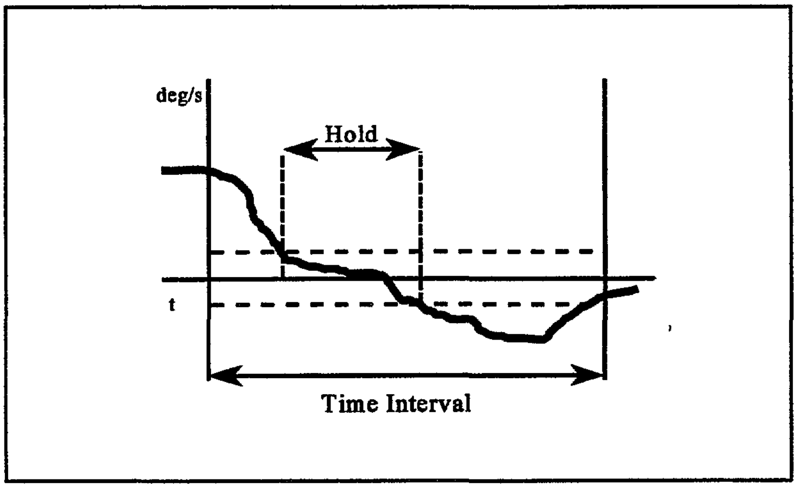


Figure B-1. Occurrence of a Steering Wheel Hold

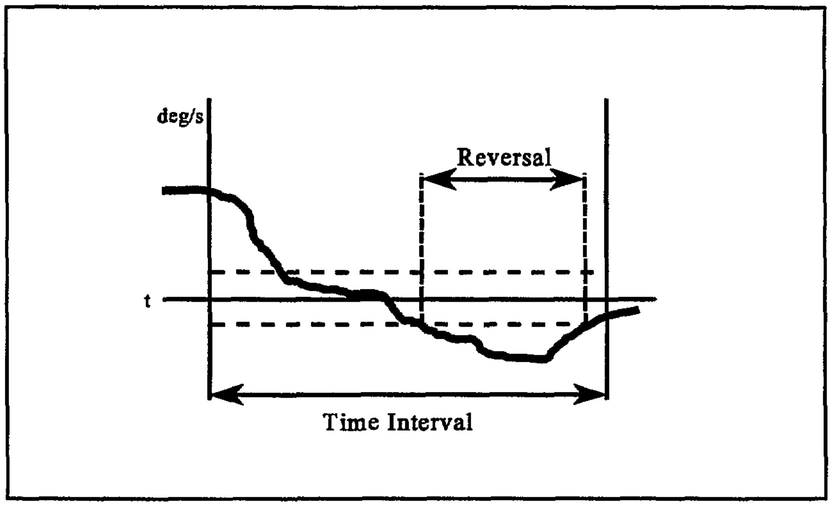


Figure B-2.a Occurrence of a Steering Wheel Reversal

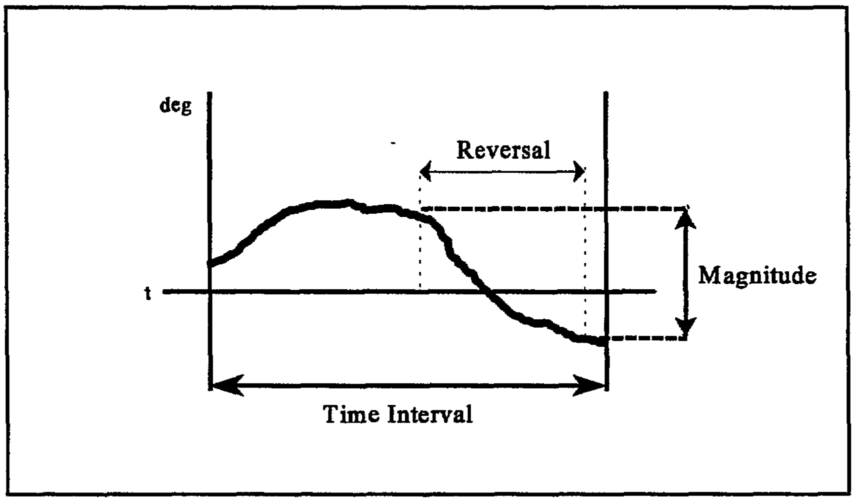


Figure B-2.b. Steering Wheel Reversal's Magnitude

Figure B-2. Graphical Depictions of Steering Reversals. See text for explanation.

Steering Zero Crossings

Steering zero crossings are defined as the number of times that the steering position passes from a magnitude of ζ or greater in one direction, through zero, and then to a magnitude of ζ or greater in the other direction. More specifically, the procedure for deriving steering zero crossings is

- Run a high pass digital filter (double pass, 0.075Hz, single pole) on the filtered steering wheel position data in order to remove low frequency noise (dc component).
- Determine the HP steering position data zero dead band based on test run data. (Dead band limits are $\pm\zeta$).
- A zero crossing occurs when steering position enters the zero band and exits the band on the other side, resulting in a change of sign in the steering position. See Figure B-3 for a depiction of Steering Zero Crossings.

From these fundamental measures, the steering measures of performance (MOPs) in Table B-2 may be derived. Similar definitions and analyses apply to the accelerator measures and brake measures (see Table B-3). The tables each consist of the following elements:

- Operational definitions of each MOP;
- A workload interpretation, i.e., a prediction of how the MOP should vary with increased workload.

The analysis of MOPs may be conducted using traditional inferential statistical methods. Specific statistical methods that are applicable include t-tests, analysis of variance (ANOVA), and various multivariate procedures (e.g., regression methods, multivariate analysis of variance, cluster analysis). In addition, graphical depictions of univariate and multivariate data are applicable. The references at the end of this appendix provide examples of various analytical procedures.

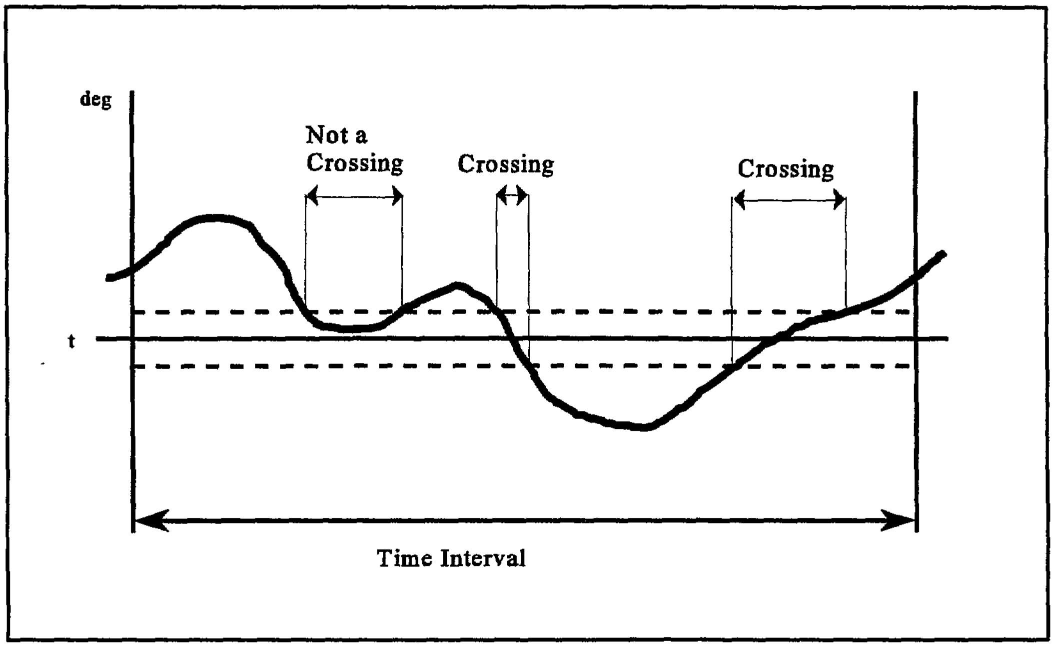


Figure B-3. Graphical Depiction of Number of Zero Crossings. See text for further explanation.

REFERENCES

Antin, J. A., Dingus, T. A., Hulse, M. C., & Wierwille, W. W. (1990). An evaluation of the effectiveness and efficiency of an automobile moving-map navigational display. *International Journal of Man-Machine Studies*, *33*, 581-594.

Battelle. (1994, June). *Functional requirements for an instrumented vehicle to support the ADVANCE safety evaluation* (Contract No. DTRS-57-89-D-00086). Columbus, OH: Battelle Memorial Institute.

Burnett. G. E., & Joyner, S. M. (1993). An investigation on the man machine interfaces to existing route guidance systems. *IEEE Vehicle Navigation and Information Systems Conference*, Ottawa, VNIS '93, 395-400.

Dingus, T., Antin, J. F., Hulse, M. C., & Wierwille, W. W. (1986). *Human factors test and evaluation of an automobile moving-map navigation system. Part I: Attentional demand requirements* (IEOR Report No. 86-03). Blacksburg, VA: Vehicle Analysis and Simulation Laboratory, Virginia Polytechnic Institute and State University.

Dingus, T. A., Antin, J. F., Hulse, M. C., & Wierwille, W. W. (1989). Attentional demand requirements of an automobile moving-map navigation system. *Transportation Research*, 23A(4), 301-315.

Dingus, T. A., Hulse, M. C., Fleischman, R. N., McGehee, D. V, & Manakkal, N. (in press). The effects of age and navigation technique on driving with an Advanced Traveler Information System. *Human Factors.*

Dingus, T. A., Mollenauer, M. A., Hulse, M. C., McGehee, D. V., & Fleischman, R. N. (in press). The effect of experience and navigation configuration on driver performance during Advance Traveler Information System use. *Human Factors.*

Drory, A. (1985). Effects of rest and secondary task on simulated truck-driving task performance. *Human Factors*, *27*(2), 201-207.

Gawron, V. (1994, April). *Test planning, analysis, and evaluation system (Test PAES) configuration notes*. Buffalo, NY: Calspan.

MacDonald, W. A., & Hoffmann, E. R. (1980). Review of relationships between steering wheel reversal rate and driving task demand. *Human Factors*, *22*(6), 733-739.

McLean, J. R., & Hoffmann, E. R. (1975). Steering reversals as a measure of driver performance steering task difficulty. *Human Factors*, *17*(3), 248-256.

Monty, R. W. (1984). *Eye movements and driver performance with electronic automotive displays*. Unpublished Master's Thesis, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Noy, I. (1990, February). *Attention and performance while driving with auxiliary in-vehicle displays* (Report No. TP 10727). Ottawa: Transport Canada.

Safford, R. S., & Rockwell, T. H. (1967). Performance decrement in twenty-four hour driving. *Highway Research Record, 163*, 68-79.

Verwey, W. B. (1991) *Towards guidelines for in-car information management: Driver workload in specific driving situations* (Report No. IZF 1991 C-13). Soesterberg, The Netherlands: TNO Institute for Perception.

Wiener, E. L., Curry, R. E., & Faustina, M. L. (1984). Vigilance and task load: In search of the inverted U. *Human Factors, 26*, 215-222.

Wierwille, W. W. (1994). *Overview of research on driver drowsiness definition and driver drowsiness detection* (Paper No. 94 S3 0 07). Paper presented at the XIVth International Technical Conference on the Enhanced Safety of Vehicles, Munich, Germany, May 23-26, 1994.

Wierwille, W. W., Wreggit, S. S., & Mitchell, M. W. (1992, April). *Research on vehicle-based driver status/performance monitoring* (Contract No. DTNH-22-91-Y-07266, ISE Report No. 92-01). Blacksburg, VA: Virginia Polytechnic Institute and State University.

Table B-2. Steering Measures of Performance (MOPs) used for Driver Workload Research. Note: All measures are taken within a sample interval of time (i.e., across n samples) and, unless otherwise indicated, are defined with respect to a single driver.

Steering PositionVariance (STPVAR) and Standard Deviation (STPSDEV):

These are two measures of the variation in steering wheel position over a sample interval of time.

$$STPVAR = \frac{\sum_{i=1}^{n} (\delta(i) - \bar{\delta})^2}{n}$$

$$STPSDEV = \sqrt{STPVAR}$$

where STPVAR is steering variance in degrees^2
STPSDEV is steering standard deviation in degrees
δ(i) is steering wheel position at sample i
δ-bar is mean steering wheel position for the sample interval
n is the number of samples in the sample interval

Workload Interpretation:

When the driver attends to the lanekeeping task, the driver makes continuous, smaller steering corrections, typically in the range of 2 to 6 degrees for passenger cars. With increased attention to in-cab tasks (or other distractions), the frequency of steering corrections per unit time tends to decrease. Since small steering corrections decrease, the vehicle tends to drift further from the lane center and this requires a larger corrective steering input (generally greater than 6 degrees for a passenger car) subsequently. If small steering corrections decrease and large corrections increase, steering wheel position variance (or standard deviation) should increase with increased workload demand.

Note: Variance measures may be more sensitive in a statistical sense because of the wider range of values results with the variance calculation. The advantage of the standard deviation measure is that it is in common engineering units, not squared units.

Table B-2. Steering Measures of Performance (MOPs) used for Driver Workload Research. Note: All measures are taken within a sample interval of time (i.e., across n samples) and, unless otherwise indicated, are defined with respect to a single driver. (Continued)

Number of Steering Holds = Total number of holds within a sample time interval.	
Workload interpretation:	If in-vehicle device demand is high, the driver will have to direct his or her attention to the device, numerous times. During such periods, the driver may hold the wheel relatively still then make a corrective input after taking a glance to the roadway. Thus, the number of steering holds may increase as task demand increases. Note that number of steering holds and steering hold duration may trade off within a fixed sample interval. That is, very long hold durations may be indicative of high workload demand yet may be associated with fewer rather than more steering holds. Thus, it is important to consider the two measures together, especially if the sample interval is fixed rather than allowed to reflect task completion time.

Table B-2. Steering Measures of Performance (MOPs) used for Driver Workload Research. Note: All measures are taken within a sample interval of time (i.e., across n samples) and, unless otherwise indicated, are defined with respect to a single driver. (Continued)

Mean duration of steering holds:

$$\textit{Mean Steering Hold} = \frac{\sum_{i=1}^{J} \textit{Steering Hold Duration}_j(i)}{J}$$

The mean steering hold duration is the sum of individual steering hold durations (steering hold duration$_j$) divided by the number of steering holds in the sample interval, J.

Workload interpretation: Given that steering holds may represent open loop control periods when the driver is attending to a task other than the driving task, longer mean steering holds are associated with higher workload demands.

. Longer holds on average imply greater attentional demand than shorter holds. It is acknowledged that number of steering holds and mean duration of steering holds measure somewhat different processes. Per unit of time, more steering holds imply shorter durations per hold. One interpretation of this is that the in-vehicle task requires multiple glances (and holds) for task completion. On the other hand, long hold durations, like long visual glance durations, imply greater demand as well and also merit assessment.

Note that mean duration of steering holds can only be evaluated if there is at least one hold within the sample interval.

Table B-2. Steering Measures of Performance (MOPs) used for Driver Workload Research. Note: All measures are taken within a sample interval of time (i.e., across n samples) and, unless otherwise indicated, are defined with respect to a single driver. (Continued)

RMS Steering Velocity Variance (STVRMS) and
Steering Velocity Variance (STVELVAR):

$$STVRMS = \sqrt{\frac{\sum_{i=1}^{n} \dot{\delta}(i)^2}{n}}$$

$$STVELVAR = \frac{\sum_{i=i}^{n} (\dot{\delta}(i) - \dot{\mu})^2}{n}$$

Workload Interpretation: RMS Steering Velocity and Steering Velocity Variance may increase or decrease with increasing in-vehicle device workload (Dingus, 1987). If the driver holds the steering wheel when attentional demand is high instead of performing the normal small corrections, the both measures would decrease with increasing workload. If on the other hand, the vehicle begins to drift off the road, the driver might "jerk" the wheel during high workload situations to correct the vehicle lane keeping error, in which case these measures should increase with increasing workload.

Note that RMS measures are identical to the square root of variance measures only if the mean of the measured variable is zero. It can be demonstrated that variance measures are not affected by a constant offset, a useful property if such a noise source is present in the data stream.

Table B-2. Steering Measures of Performance (MOPs) used for Driver Workload Research. Note: All measures are taken within a sample interval of time (i.e., across n samples) and, unless otherwise indicated, are defined with respect to a single driver. (Continued)

Steering Velocity Percent Zero Average (STPZAV):

$$STPZAV = 100 * \frac{\sum_{i=1}^{J} steering\ hold\ duration(i)}{Sample\ interval\ Duration}$$

Steering Hold Variance (STPHVAR):

$$STPHVAR = \sum_{i=1}^{J} \frac{(steering\ hold\ duration(i) - mean\ steering\ hold\ duration)^2}{J}$$

Workload Interpretation: STPZAV is basically a measure of the average time that the steering wheel is held during a task (Dingus et al., 1986). This measure might increase with increased attentional demand if, as anticipated, small steering inputs disappear when the driver is distracted from the driving task.

STPHVAR is essentially the variance of the time that the steering wheel is held during a task. If holds are followed by larger steering inputs (to correct relatively larger lane drift), then this response measure should increase with increased attentional demand (i.e., holds followed by jerks to correct lanekeeping error). Note that this value can only be evaluated if there is a least one hold in the sample interval.

Table B-2. Steering Measures of Performance (MOPs) used for Driver Workload Research. Note: All measures are taken within a sample interval of time (i.e., across n samples) and, unless otherwise indicated, are defined with respect to a single driver. (Continued)

Number of steering reversals per unit time:		See definition under Fundamental Measures. Generally defined as steering reversals of at least a 2 degree magnitude.
Number of small, medium, and large reversals:	Small reversals:	Steering reversals ≤ 2 degrees
	Medium reversals:	2 degrees <Steering reversals <6 degrees
	Large reversals:	Steering reversals ≥ 6 degrees
Workload Interpretation:	Steering reversals per second may decrease under conditions of high attentional demand away from the driving task.	
		In general, the number of small steering reversals might be expected to decrease and the number of medium or large reversals is expected to increase with greater attentional demand away from the driving task.
Number of zero crossings.	See definition under Fundamental Measures.	
Workload interpretation:	As a measure of steering activity, the number of zero crossings may decrease with increased attentional demand away from the driving task if the driver is holding the steering wheel while engaged in an in-vehicle task.	
Hand-off-Wheel time:	This is the time that one hand is off the steering wheel and engaged in in-vehicle device use.	
Workload Interpretation:	As manual demand of in-cab device use increases, the hand-off-wheel time may increase also.	

Table B-3. Accelerator and Brake Measures of Performance (MOPs) used for Driver Workload Research. Note: All measures are taken within a sample interval of time (i.e., across n samples) and, unless otherwise indicated, are defined with respect to a single driver.

Accelerator Variance (ACLVAR) and Accelerator Standard Deviation (ACLSDEV):

These are two measures of the variation in accelerator position over a sample interval of time.

$$ACLVAR = \frac{\sum_{i=1}^{n} (accelpos(i) - \overline{accelpos})^2}{n}$$

$$ACLSDEV = \sqrt{ACLVAR}$$

where ACLVAR is accelerator position variance in degrees^2
ACLSDEV is accelerator position standard deviation in degrees
accelpos(i) is accelerator position at sample i
accelpos-bar is mean accelerator position for the sample interval
n is the number of samples in the sample interval

Workload Interpretation: When the driver attends to the speed maintenance task, the driver makes continuous, smaller accelerator corrections. With increased attention to in-cab tasks (or other distractions), the frequency of such corrections per unit time tends to decrease. Since small accelerator corrections decrease, the vehicle tends to slow down and this requires a larger corrective accelerator input subsequently. Since small corrections decrease and large corrections increase, accelerator position variance (or standard deviation) should increase with increased workload demand.

Note: Variance measures may be more sensitive in a statistical sense because of the wider range of values results with the variance calculation. The advantage of the standard deviation measure is that it is in common engineering units, not squared units.

Table B-3. Accelerator and Brake Measures of Performance (MOPs) used for Driver Workload Research. Note: All measures are taken within a sample interval of time (i.e., across n samples) and, unless otherwise indicated, are defined with respect to a single driver. (Continued)

Accelerator Reversals:	See definition under Fundamental Measures. Accelerator reversals are the number of times the accelerator velocity changes sign.
Workload Interpretation:	As with steering reversals, the number of accelerator reversals is expected to decrease under conditions of high attentional demand away from the driving task.
Number of Accelerator Holds:	An accelerator hold is defined to occur when the steering wheel velocity falls within the zero dead band range for a duration of 0.4 sec or longer. This measure is a count of the number of such holds in the sample interval.
Workload interpretation:	If in-vehicle device demand is high, the driver will have to direct his or her attention to the device, numerous times. During such periods, the driver will presumably keep the accelerator pedal relatively still then make a corrective input after taking a glance to the roadway. Thus, the number of accelerator holds should increase as task demand increases. Note that number of accelerator holds and accelerator hold duration may trade off within a fixed sample interval. That is, very long hold durations may be indicative of high workload demand yet may be associated with fewer rather than more accelerator holds. Thus, it is important to consider the two measures together, especially if the sample interval is fixed rather than allowed to reflect task completion time.

Table B-3. Accelerator and Brake Measures of Performance (MOPs) used for Driver Workload Research. Note: All measures are taken within a sample interval of time (i.e., across n samples) and, unless otherwise indicated, are defined with respect to a single driver. (Continued)

Mean Accelerator Hold Duration:

$$Mean\ Accelerator\ Hold = \frac{\sum_{i=1}^{J} accelerator\ hold\ duration(i)}{J}$$

The mean accelerator hold duration is the sum of individual accelerator hold durations (accelerator hold duration_j) divided by the number of such holds in the sample interval, J.

Workload interpretation: Given that accelerator holds may represent open loop control periods when the driver is attending to a task other than the driving task, longer mean accelerator holds are associated with higher workload demands.

Note that number of accelerator holds and accelerator hold duration may trade off within a fixed sample interval. That is, very long hold durations may be indicative of high workload demand yet may be associated with fewer rather than more accelerator holds. Thus, it is important to consider the two measures together, especially if the sample interval is fixed rather than allowed to reflect task completion time. It should be mentioned that at least one accelerator hold is needed to compute this measure.

Table B-3. Accelerator and Brake Measures of Performance (MOPs) used for Driver Workload Research. Note: All measures are taken within a sample interval of time (i.e., across n samples) and, unless otherwise indicated, are defined with respect to a single driver. (Continued)

Total Accelerator Hold Time =	$\sum_{i=1}^{J} \mathit{Accelerator\ Hold\ Duration}(i)$ i.e., total accelerator hold time is the sum of all accelerator hold durations in the sample interval.
Proportion Accelerator Hold Time: = [Total Accelerator Hold Time / Sample Interval]	
Workload Interpretation:	The total accelerator hold time (or proportion of time) provide other measures of attentional demand. The percentage measure may be used when there is a need to normalize total time measures based on the length of the sample interval. As workload demand increases, total time and proportion of the time the accelerator is held should increase.
Accelerator Releases:	The number of times that the accelerator is in its null position (e.g., effectively 0% of throw).
Workload Interpretation:	Accelerator Releases may be one workload management strategy (besides holds) that allows the driver to "slow" the driving world down somewhat while engaged in an in-vehicle task or other distraction from the driving scene. If this strategy is used, the number of releases should go up with higher attentional demand.

Table B-3. Accelerator and Brake Measures of Performance (MOPs) used for Driver Workload Research. Note: All measures are taken within a sample interval of time (i.e., across n samples) and, unless otherwise indicated, are defined with respect to a single driver. (Continued)

Mean Accelerator Release Duration:	$\mathit{Mean\ Accelerator\ Release\ Duration} = \frac{\sum_{i=1}^{J} \mathit{accelerator\ release\ duration}(i)}{J}$ The mean accelerator release duration is the sum of individual accelerator release durations, accelerator release duration(i), divided by the number of releases in the sample interval, J.
Workload Interpretation:	Given that accelerator releases may represent open loop control periods when the driver is attending to a task other than the driving task, longer mean accelerator release durations are associated with higher workload demands. Note that number of accelerator releases and accelerator release duration may trade off within a fixed sample interval. That is, very long hold durations may be indicative of high workload demand yet may be associated with fewer rather than more accelerator releases. Thus, it is important to consider the two measures together, especially if the sample interval is fixed rather than allowed to reflect task completion time. It should be mentioned that at least one accelerator hold is needed to compute this measure.
Brake Application (BRNUM):	The number of times the brake pedal is depressed sufficiently to activate the vehicle brake lights.
Workload Interpretation:	It is speculated that when the in-vehicle task becomes difficult, the driver rests a foot on the brake pedal to quickly take action when the driver redirects vision to the forward view. Thus, the number of brake activations may increase with increased attentional demand away from the driving task.

Table B-3. Accelerator and Brake Measures of Performance (MOPs) used for Driver Workload Research. Note: All measures are taken within a sample interval of time (i.e., across n samples) and, unless otherwise indicated, are defined with respect to a single driver. (Continued)

Brake Reaction Time:	The time interval from the onset of an event (e.g., traffic control signal changes, lead vehicle brakes, pedestrian steps into the driving lane) to application of brakes (see previous definition).
Workload Interpretation:	Longer brake reaction times are normally associated with higher attentional demand away from the driving task.
Brake Dwell Time Average (BRTIMEAV):	This the mean time the driver spent with a foot on the brake pedal.
	$BRTIMEAV = \frac{\sum_{i=1}^{J} Dwelltime(i)}{BRNUM}$
where	BRTIMEAV is the brake dwell time average *Dwelltime(i)* is the dwell duration for brake application *i* BRNUM is the total number of brake applications in the sample interval.
Workload Interpretation:	The average brake dwell time may increase if the driver rests a foot on the brake pedal in the face of increased attentional demand away from the driving scene in order to quickly react to unexpected events when the driver returns vision to the road scene ahead.

APPENDIX C

DRIVER-VEHICLE PERFORMANCE MEASURES IN WORKLOAD ASSESSMENT

APPENDIX C. DRIVER-VEHICLE PERFORMANCE MEASURES IN WORKLOAD ASSESSMENT

MOTIVATION

Driver-vehicle performance measures are measures of various aspects of lateral and longitudinal control. Because these measures are closely related to vehicle trajectories in space and over time, such measures are closely related to safety. Visual allocation and other in-vehicle behaviors may ultimately result in changes in driver-vehicle performance measures, thereby affecting safety. Consequently, examination of driver-vehicle performance while the drivers perform in-vehicle tasks may give an indication of safety relevance. Each of several key classes of measures will be introduced briefly below.

Uncontrolled or inappropriate lane excursions are the precursor to a great many crashes each year, including lane change crashes (Chovan, Tijerina, Alexander, and Hendricks, 1994), opposite-direction crashes (Chovan, Everson, Hendricks, and Pierowicz, 1994), and single vehicle roadway departures (Hendricks, Allen, Tijerina, Everson, Knipling, and Wilson, 1992; Mironer and Hendricks, 1994). From a safety perspective, lanekeeping performance during in-vehicle device use merits scrutiny in a workload evaluation. There is evidence that various lane keeping measures have demonstrated sensitivity to workload demand, both primary driving task demand and in-vehicle distraction, as well as value as part of a set of indicators or driver fatigue or incapacitation (e.g., Wierwille, 1994).

Lane keeping measures have played a prominent role in driver workload assessment. Zwahlen, Adams and DeBald (1988) evaluated CRT touch panels in an automobile driven on a closed airport runway and recorded path deviations over a short distance while the vehicle maintained a 40 mph forward speed. Using paper mock-ups of the CRT touch displays, the path standard deviation significantly increased with in-vehicle tasks, even though drivers had the opportunity to glance to the road ahead. Green (1993) has correctly noted that in this study there was no traffic and so lanekeeping was probably given lower priority than might have been the case under realistic driving conditions. In a series of simulator studies of in-vehicle visual and auditory displays, Noy (1990) reported that lane standard deviation increased with auxiliary cognitive tasks. Perhaps because participants knew they were in a simulator, they may have altered their prioritization of performance such that in-vehicle tasks were accorded more attention than in actual driving. Green, Hoekstra, and Williams (1993) conducted on-the-road data collection in an instrumented vehicle to assess the workload effects of using route guidance and car phone devices and generally found no significant effects of in-vehicle device use on either mean lane position or lane position standard deviation. Dingus, Hulse, Fleischman, McGehee, and Manakkal (in press) reported that older drivers made a significant number of lane deviations (exceedences) and that these varied reliably with the nature of the in-vehicle route guidance display. Older drivers also tended to drive more slowly and cautiously. Dingus, Mollenhauer, Hulse, McGehee, and Fleischman (in press) reported that the results of the TravTek evaluation showed that the number of unplanned lane deviations decreased from a local driver's first to second drive with TravTek. This varied pattern of results suggests that, while lane keeping measures have safety relevance as well as intuitive appeal as workload measures, these measures are subject to the influence of many factors besides workload. Therefore, interpretation of such measures must be conducted with care and in combination with other measures taken to converge on a more accurate device assessment.

Excessive speed is involved in many crashes, particularly those involving younger drivers (Lonero, Clinton, Brock, Wilde, Laurie, and Black, 1995). In addition, it has been noted that older drivers often drive more slowly as a compensation for slower reflexes or reduced capacity to monitor driving conditions (Dingus, Mollenhauer, Hulse, McGehee, and Fleischman, in press). Speed measures for workload assessment often uncover a "slowing" of the vehicle with increased attentional demand. For example, Monty (1984) reported that speed maintenance deteriorated with in-vehicle device use. A similar finding was reported by Noy (1990) in a simulator study. Labiale (1990) carried out an on-the-road study to compare auditory and visual displays of motorist advisory information of different complexity levels. Labiale reported that approximately half of the drivers in the experiments reduced vehicle speed during long vs. short messages and more reduced speed with visual messages than with auditory messages. On the other hand, Dingus, Antin, Hulse, and Wierwille (1989) found no impact on speed measures with the use of the ETAK navigator in various modes. Verwey (1991) conducted a simulator study that involved visual and auditory secondary (in-vehicle) tasks and also reported no impact of secondary task on speed measures. These varied results illustrate that many other factors besides workload enter into speed maintenance. These include driving style, road disturbances (e.g., vertical roadway alignment), and driver workload management strategies, to name a few. Thus, changes in speed measures, like many others in workload assessment, must be interpreted with caution.

Inadequacies in car following performance are directly related to rear-end crashes (Knipling, Mironer, Hendricks, Tijerina, Everson, Allen, and Wilson, 1993), the single most common crash type in the United States. The vast majority of these crashes involve driver inattention and/or following too closely. Furthermore, Evans and Wasielewski (1982) showed that time headway adopted on a section of highway was a significant discriminator of traffic violators from non-violators, that time headways were often below 1.0 s for the traffic violators, and that such short headways greatly increase the risk of rear-end crashes. Thus, there is at least some archival evidence that car following measures such as time headway have safety relevance. The principles of physics also can be used to relate close car following to crash involvement (Mironer and Hendricks, 1994). In particular, more aggressive evasive braking or steering maneuvers are required if the following vehicle is following too closely and/or closing fast on the lead vehicle.

The literature on car following measures is somewhat limited but does indicate the value of considering such measures in a workload assessment protocol. Colbourn, Brown, and Copeman (1978) reported that drivers adopt longer headway distances, on average, when traveling at higher speeds and that the variability of following distance increases with increased travel speed as well.

Noy (1990), in a simulator study, reported that adding an auxiliary task of visually scanning a CRT as associated with degraded headway distance maintenance and that following distance mean error (from a desired following distance) was sensitive to different levels of auxiliary task difficulty.

One potential difficulty with car following measures is that because they are closely related to safety, ethical data collection will require that the evaluation safeguard against car following deterioration. For example, one requirement may be to insure that there is at least, say, 150 ft of car following distance before an in-vehicle task can be presented to the driver. This type of safeguard may eliminate any effects of the in-vehicle device distraction. Another consideration is that the driver may elect to adopt a risky car following behavior (e.g., time headway less than 1 s); engagement in in-vehicle device use may be ill-advised under such circumstances.

INSTRUMENTATION NEEDS

The instrumentation needed to capture lateral and longitudinal driver-vehicle performance measures is discussed below.

SENSOR SUITE

Lane Tracker. Optical sensors (including video cameras) are available that sense the luminance difference between the lane marker line and the surrounding pavement. In sophisticated systems, machine vision processing is applied to determine the location of the vehicle over time. An alternative to this approach is the use of a video camera followed by video post-processing. In post-processing, a data reducer could, with a digitizing tablet, encode the lane line position manually. Other types of lane tracking technologies make use of radar that senses special retroflective material used for the lane line markings. Such approaches are generally only suitable for closed courses or test tracks that can be fitted with cooperative infrastructure modifications. It should be noted that optically-based lane trackers are susceptible to numerous sources of noise and data loss. These sources include specular reflections from the pavement, shadows or dappled sunlight, precipitation build-up on the road surface, and worn lane line markings. Furthermore, it should be noted that the lane tracker will track the lane line at exit and entrance ramps, a phenomenon that leads to momentary errors. The Department of Transportation has invested in Small Business Innovative Research (SBIR) to fund the development of cheaper and more reliable lane trackers. If video or optical means are used for capturing lane position, suitable light sources will be required to illuminate the sensor/camera's field of view. Mounting hardware for the lane tracker is also required. The sample rate should be equal to the video frame rate (e.g., 30 samples per second for NTSC or 25 samples per second for PAL video). The resolution should be ±1 inch.

Accelerometers A longitudinal accelerometer is needed to capture decelerations (e.g., effective braking levels) or accelerations in the x-axis. A lateral accelerometer is needed to capture lateral accelerations (e.g., magnitude of evasive steering). Finally, yaw rate may be computed from the difference of front and rear-body lateral accelerometer signals, by means of an angular accelerometer, or by means of a rate gyroscope. The lateral and longitudinal accelerometers should be capable of recording 0 - 2 g's with resolution to ±0.001 g. The yaw rate system should be capable of recording 0-20 deg/sec and resolution of 0.01 deg/sec. If only a single set of accelerometers can be installed in the vehicle, it should be located at or near the vehicle center of mass. The sampling rate should match that of the (concurrent) video frame rate.

Speed Sensor. Speed may be measured by means of a fifth wheel tachometer or by a magnetic or inductive pickup on a wheel or the drive shaft. The speed measurement should be over a range from 0 to 80 mph and accurate to within ±0.5 mph. The sampling rate should match that of the (concurrent) video frame rate, i.e., be between 25 and 30 Hz.

Headway and Closing Rate Sensors. The headway distance or distance between the subject vehicle and a lead vehicle can be measured by means of a laser rangefinder or a radar rangefinder. Since these are line-of-sight systems, they are prone to data loss or noise caused by horizontal and vertical roadway geometry, roadside appurtenances, and the like. Furthermore, the headway distances can jump suddenly with cut-in maneuvers by other vehicles in adjacent lanes moving into the subject vehicle's travel lane. Closing rate or relative velocity can be measured by means of a Doppler radar system. The system should have a distance range of up to 350 ft and resolution of ± 1ft. The closing rate or relative velocity sensor should have a range of up to 60 mph and a resolution of ± 1 mph. A sampling rate of 25 to 30 Hz should be sufficient for driver workload assessment.

DATA CAPTURE AND CONTROL

As was explained in the Appendix on steering, accelerator, and manual activity measurement, the efficient capture of sensor data for driver-vehicle performance measures is best managed by means of a computer on-board the instrumented vehicle. Two possible options for data capture and control of sensor data are Pulse Code Modulated (PCM) data recorders or Data Acquisition computers with analog-to-digital (A/D) converters. The PCM option provides the highest bandwidth, highest data storage density, and easiest means for data transport. The data acquisition computer has the advantage of converting all data to digital form at the time of data collection. With this option, a set of anti-aliasing filters must be incorporated to ensure that digitized data have high fidelity. Additionally, an external storage device is essential, especially if a data collection run is to last for any appreciable length of time. Examples of mass data storage devices include high density disk drives and magnetic tape cartridges. It must be noted also that there is more limited bandwidth associated with an affordable direct-to-digital system.

Table B-1 in Appendix B presents the advantages and disadvantages of the two options (Battelle, 1994). As with other parts of the data capture system, power must be available, conditioned to avoid data loss or error. A time-code generator is needed as the basic common reference point for all data channels.

DATA REDUCTION AND FILTERING

The analysis of driver-vehicle performance data , like that for steering and pedal inputs, depends first on appropriate filtering and data reduction. To examine the data, software that allows for the simultaneous examination of multiple data streams is helpful, including an examination of a videotape. One type of system that can facilitate this type of data analysis is the Intelligent Transportation System Test Performance Assessment and Evaluation System (ITS TEST PAES) software prepared under government contract by Calspan (for general information see Gawron, 1994).

FUNDAMENTAL DATA

The following sensed data are required for workload measures based on driver-vehicle performance: lane position, yaw rate, speed, following distance, and closing velocity. From these fundamental data, the measures provided in Table C-1 may be derived. As with similar tables in other appendices, this table consists of the following elements:

- Operational definitions of each Measure Of Performance (MOP);
- A workload interpretation, i.e., a prediction of how the MOP may vary with increased workload.

The analysis of MOPs may be conducted using traditional inferential statistical methods. Specific statistical methods that are applicable include t-tests, analysis of variance (ANOVA), and various multivariate procedures (e.g., regression methods, multivariate analysis of variance, cluster analysis). In addition, graphical depictions of univariate and multivariate data are applicable. The references for this appendix provide examples of various analytical procedures.

REFERENCES

Battelle. (1994, June). *Functional requirements for an instrumented vehicle to support the ADVANCE safety evaluation* (Contract No. DTRS-57-89-D-00086). Columbus, OH: Battelle Memorial Institute.

Chovan, J.D., Everson, J.J., Hendricks, D.L., & Pierowicz, J. (1994, June). *Analysis of opposite-direction crashes* (Contract No. DTRS-57-89-D-00086). Columbus, OH: Battelle.

Chovan, J.D., Tijerina, L., Alexander, G. & Hendricks, D.L. (1994, March). *Examination of lane change crashes and potential IVHS countermeasures* (Report No. DOT HS 808 071 and DOT-VNTSC-NHTSA-93-2). Cambridge, MA: John A. Volpe National Transportation Systems Center.

Colbourn, C. J., Brown, I. D., & Copeman, A. K. (1978). Driver's judgements of safe distances in vehicle following. *Human Factors, 20*(1), 1-12.

Dingus, T., Antin, J. F., Hulse, M. C., & Wierwille, W. W. (1986). *Human factors test and evaluation of an automobile moving-map navigation system. Part I: Attentional demand requirements* (IEOR Report No. 86-03). Blacksburg, VA: Vehicle Analysis and Simulation Laboratory, Virginia Polytechnic Institute and State University.

Dingus, T. A., Antin, J. F., Hulse, M. C., & Wierwille, W. W. (1989). Attentional demand requirements of an automobile moving-map navigation system. *Transportation Research,* 23A(4), 301-315.

Dingus, T. A., & Hulse, M. C. (1993, January). *TravTek evaluation - Task C3 camera car study - final detailed test plan* (Contract No. DTFH61-91-C-00106). Washington, DC: Federal Highway Administration.

Dingus, T. A., Hulse, M. C., Fleischman, R. N., McGehee, D. V, & Manakkal, N. (in press). The effects of age and navigation technique on driving with an Advanced Traveler Information System. *Human Factors.*

Dingus, T. A., Mollenauer, M. A., Hulse, M. C., McGehee, D. V., & Fleischman, R. N. (in press). The effect of experience and navigation configuration on driver performance during Advance Traveler Information System use. *Human Factors.*

Evans, L., & Wasielewski, P. (1982). Do accident involved drivers exhibit riskier everyday driving behavior? *Accident Prevention and Analysis, 14,* 57-64.

Gawron, V. (1994, April). *Test planning, analysis, and evaluation system (Test PAES) configuration notes.* Buffalo, NY: Calspan.

Green, P. (1993, April). *Measures and methods used to assess the safety and usability of driver information systems* (Technical Report No. UMTRI-93-12). Ann Arbor, MI: University of Michigan Transportation Research Institute.

Green, P., Hoekstra, E., & Williams, M. (1993, November). *Further on-the-road tests of driver interfaces: Examination of a route guidance system and a car phone* (Technical Report No. UMTRI-93-35). Ann Arbor, MI: University of Michigan Transportation Research Institute.

Hendricks, D., Allen, J., Tijerina, L., Everson, J., Knipling, R., Wilson, C. (July, 1992). *VNTSC IVHS program Topical Report #2: Single vehicle roadway departures* (Volumes I and II) (Contract No. DTRS-57-89-D-00086). Columbus, OH: Battelle

Knipling, R.R., Mironer, M., Hendricks, D.L., Tijerina, L., Everson, J., Allen, J.C., & Wilson, C. (1993, May). *Assessment of (IVHS countermeasures for collision avoidance: Rear-end crashes* (Report No. DOT HS 807 995). Washington, DC: U.S. Department of Transportation National Highway Traffic Safety Administration.

Labiale, G. (1990). In-car road information: Comparisons of auditory and visual presentations. *Proceedings of the Human Factors Society 34th Annual Meeting*, 623-627.

Lonero, L., Clinton, K. Brock, J. Wilde, G., Laurie, I., & Black, D. (1995, March). *Novice driver education model curriculum outline*. Washington, DC: AAA Foundation for Traffic Safety.

Mironer, M., & Hendricks, D.L. (1994, August). *Examination of single vehicle roadway departure crashes and potential IVHS countermeasures* (Report No. DOT HS 808 144 and DOT-VNTSC-NHTSA-94-3). Cambridge, MA: U.S. Department of Transportation John A. Volpe National Transportation Systems Center.

Monty, R. W. (1984). *Eye movements and driver performance with electronic automotive displays*. Unpublished Master's Thesis, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Noy, I. (1990, February). *Attention and performance while driving with auxiliary in-vehicle displays* (Report No. TP 10727). Ottawa: Transport Canada.

Verwey, W. B. (1991) *Towards guidelines for in-car information management: Driver workload in specific driving situations* (Report No. IZF 1991 C-13). Soesterberg, The Netherlands: TNO Institute for Perception.

Wierwille, W. W. (1994). *Overview of research on driver drowsiness definition and driver drowsiness detection* (Paper No. 94 S3 0 07). Paper presented at the XIVth International Technical Conference on the Enhanced Safety of Vehicles, Munich, Germany, May 23-26, 1994.

Zwahlen, H., Adams, C. C., & DeBald, D. P. (1988). Safety aspects of CRT touch panel controls in automobiles. In A. G. Gale et al. (Eds.), *Vision in Vehicles II* (pp. 335 - 344). Amsterdam: Elsevier.

Table C-1. Driver-Vehicle Measures of Performance (MOPs) used for Driver Workload Research. Note: All measures are taken within a sample interval of time (i.e., across n samples) and, unless otherwise indicated, are defined with respect to a single driver.

Lane PositionVariance (LANEPVAR)
and Standard Deviation (LANEDEV):

These are two measures of the variation in lane position over a sample interval of time.

$$LANEPVAR = \frac{\sum_{i=1}^{n} (y(i) - \bar{y})^2}{n}$$

$$LANEDEV = \sqrt{LANEPVAR}$$

where LANEPVAR is lane position variance in inches^2
LANEDEV is lane position standard deviation in inches
y(i) is lane position at sample i
y-bar is mean lane position for the sample interval
n is the number of samples in the sample interval

Workload Interpretation: While no driver maintains the vehicle perfectly at a selected lateral position in the lane, normally the attentive driver makes continuous, smaller steering corrections that yield a certain variability in lane position. With increased attention to in-vehicle tasks (or other distractions), the frequency of steering corrections per unit time tends to decrease. Since small steering corrections decrease, the vehicle tends to drift farther from the selected lane position and this requires a larger corrective steering input subsequently. If this pattern of behavior is exhibited, lane position variance (or standard deviation) might be expected to increase with increased attentional demand.

Note: Variance measures may be more sensitive in a statistical sense because of the wider range of values that result with the variance calculation. The advantage of the standard deviation measure is that it is in common engineering units, not squared units.

Table C-1. Driver-Vehicle Measures of Performance (MOPs) used for Driver Workload Research. Note: All measures are taken within a sample interval of time (i.e., across n samples) and, unless otherwise indicated, are defined with respect to a single driver. (Continued)

Mean Lane Position (LANEPOSM):	$$LANEPOSM = \frac{\sum_{i=1}^{n} y(i)}{n}$$ The mean lane position is the sum of individual lane position samples, y(i), in the sample interval divided by the number of samples in the sample interval, n. (The zero position of y(i) corresponds to having the vehicle perfectly centered in the lane.)
Workload interpretation:	With increased attentional demand away from the lanekeeping task, the driver may drift laterally. If so, then the average lane position during such inattentive moments would perhaps be further from the lane center than under conditions of less attentional demand.

Table C-1. Driver-Vehicle Measures of Performance (MOPs) used for Driver Workload Research. Note: All measures are taken within a sample interval of time (i.e., across n samples) and, unless otherwise indicated, are defined with respect to a single driver. (Continued)

Lane RMS Deviation (RMSLANE):

$$RMSLANE = \sqrt{\frac{\sum_{i=1}^{n} y(i)^2}{n}}$$

Workload Interpretation: Lane RMS deviation would be expected to increase with increased attentional demand away from the lanekeeping task, as explained under the lane position variance and standard deviation measures. (As indicated earlier, the zero position of y(i) corresponds to having the vehicle perfectly centered in the lane.)

Note that RMS measures are identical to the square root of variance measures only if the mean of the measured variable is zero. It can be demonstrated that variance measures are not affected by a constant offset, a useful property if such a noise source is present in the data stream.

Peak Lane Deviation:
(PKLANDEV)

$$PKLANDEV = \max(|y(i)|);\ i = 1,2,\ldots,n$$

Workload Interpretation: As attentional demand increases, the driver may inadvertently allow the vehicle to veer farther from the center portion of the lane . Thus, the maximum lane deviation in the sample may be expected to increase with increased attentional demand.

Table C-1. Driver-Vehicle Measures of Performance (MOPs) used for Driver Workload Research. Note: All measures are taken within a sample interval of time (i.e., across n samples) and, unless otherwise indicated, are defined with respect to a single driver. (Continued)

Number of Lane Exceedences:	This is defined as the number of occurrences when any part of the vehicle is extended beyond either lane boundary.
Mean Lane Exceedence Duration:	This is the average time spent in a lane exceedence.
Workload Interpretation:	The interpretation of these measures is that with increased attentional demand away from the driving task, the number of lane exceedences might increase and the average duration of lane exceedences might increase. This interpretation follows from the assumption that less attention would be allocated to the lane-tracking task (Dingus et al., 1986).
Yaw Standard Deviation:.	$YAWSTDEV = \sqrt{\frac{\sum_{i=1}^{n}(\psi(i) - \bar{\psi})^2}{n}}$ Where ψ(i) is the angular difference, in degrees, between the vehicle longitudinal axis and the instantaneous roadway tangent (measured in the horizontal plane) at sample i; $\bar{\psi}$-bar is the mean of the yaw samples.
Workload interpretation:	In a moving base simulator, Hicks and Wierwille (1979) found yaw standard deviation to be a sensitive measure of primary driving task workload where the driving task workload was increased by simulated crosswind gusts. With increasing attentional demands, yaw standard deviation might be expected to increase. It is possibly a more sensitive measure than lane position measures because of the dynamics of the vehicle. However, lane position measures are more directly related to safety and are preferred.

Table C-1. Driver-Vehicle Measures of Performance (MOPs) used for Driver Workload Research. Note: All measures are taken within a sample interval of time (i.e., across n samples) and, unless otherwise indicated, are defined with respect to a single driver. (Continued)

Peak Lateral Acceleration:
(PKLATACC) This is the maximum lateral acceleration, a_y, observed in a sample interval.

$$PKLATACC = \max\left(a_y(i)\right);\ i = 1,2,\ldots,n$$

Peak Longitudinal Deceleration(PKLNGDEC): This is the maximum longitudinal deceleration, a_x, observed in a sample interval.

$$PKLNGDEC = \max\left(-a_x(i)\right);\ i = 1,2,\ldots,n$$

Workload Interpretation: Abrupt lateral maneuvers are taken to be indicative of a vehicle which is off lane-center track due to driver inattention (Dingus and Hulse, 1993). As such lateral acceleration measures should be correlated with steering measures. The workload prediction is that peak lateral accelerations, indicative of abrupt lateral maneuvers, may increase with increased attentional demand.

Abrupt braking maneuvers have also proven to be sensitive to in-vehicle device workload (Monty, 1984). If drivers look away from the road scene and glance back to discover an unexpected object or event, then the deceleration level is anticipated to be higher than under conditions of normal attention.

Table C-1. Driver-Vehicle Measures of Performance (MOPs) used for Driver Workload Research. Note: All measures are taken within a sample interval of time (i.e., across n samples) and, unless otherwise indicated, are defined with respect to a single driver. (Continued)

Mean Speed (MEANSPEED):

$$MEANSPEED = \frac{\sum_{i=1}^{n} u(i)}{n}$$

where

MEANSPEED is the average speed over the sample interval;
u(i) is the measured forward velocity at sample i; and
n is the number of samples in the sample interval.

Workload Interpretation:	Previous research has shown that under conditions of increased attentional demand, the driver often slows the vehicle. Thus, the workload prediction is that with increased attentional demand, average speed may decrease.

Table C-1. Driver-Vehicle Measures of Performance (MOPs) used for Driver Workload Research. Note: All measures are taken within a sample interval of time (i.e., across n samples) and, unless otherwise indicated, are defined with respect to a single driver. (Continued)

Speed Variance (SPEEDVAR)
or **Speed Standard Deviation**
(SPEEDEV): These are two measures of the variation in travel speed over the a sample interval of time.

$$SPEEDVAR = \frac{\sum_{i=1}^{n} (u(i) - \bar{u})^2}{n}$$

$$SPEEDEV = \sqrt{SPEEDVAR}$$

where SPEEDVAR is speed variance in $(ft/s)^2$
SPEEDEV is speed standard deviation in ft/s
u(i) is forward travel velocity at sample i
u-bar is mean travel velocity for the sample interval
n is the number of samples in the sample interval

Workload Interpretation: With increased attentional demand away from the driving task, the driver may exhibit more erratic longitudinal control. If so, then the variability of speed might be expected to increase with increased attentional demand.

Table C-1. Driver-Vehicle Measures of Performance (MOPs) used for Driver Workload Research. Note: All measures are taken within a sample interval of time (i.e., across n samples) and, unless otherwise indicated, are defined with respect to a single driver. (Continued)

Mean Following Distance (FOLDISTM):

$$FOLDISTM = \frac{\sum_{i=1}^{n} d(i)}{n}$$

where d(i) is the following distance measured at sample i;
n is the total number of following distance measures in the sample interval.

Minimum Following Distance:
(MNFOLDIS) This is the shortest following distance recorded in the sample interval.

$$MINFOLDIS = \min(d(i));\ i = 1,2,\ldots,n$$

Workload Interpretation: If the driver is not attending to the primary driving task, it is possible that car following performance will be adversely affected. In this case, the lack of attention to the car following task leads to closer (i.e., smaller) mean following distances, on average, with increased attentional demand. Minimum following distance is expected to be closer with increased attentional demand.

Peak Closing Velocity:
(PKCLOSV) This is the maximum closing rate during a sample interval, i.e.,

$$PKCLOSV = \max(u_c(i));\ i = 1,2,\ldots,n$$

Where $u_c(i)$ is the closing velocity in ft/s for sample i (positive for following vehicle gainaing on lead vehicle).

Workload Interpretation: The hypothesis is that with increased attentional demand, peak closing velocity may be expected to increase due to deterioration of car following performance.

Table C-1. Driver-Vehicle Measures of Performance (MOPs) used for Driver Workload Research. Note: All measures are taken within a sample interval of time (i.e., across n samples) and, unless otherwise indicated, are defined with respect to a single driver. (Continued)

Mean Time Headway (MHEADWAY):

$$MHEADWAY = \frac{\sum_{i=1}^{n} headway(i)}{n}$$

where headway(i) is the measured headway in sample i_j,
n is the number of samples in the sample interval.

Minimum Time Headway: (MINHDWY) This is the smallest time headway in the sample interval.

$$MINHDWY = \min(headway(i));\ i = 1,2,\ldots,n$$

Workload Interpretation: Time headway is the instantaneous following distance divided by the instantaneous subject vehicle velocity (not closing velocity) and is measured in time units. For example, if the following distance is 240 ft and the subject vehicle (following vehicle) velocity is 60 ft/s, then the time headway is 4 s. The general prediction for time headway is that may decrease with increased attentional demand under the rationale given for following distance. Minimum time headway might decrease with increased attentional demand away from the car following task.

APPENDIX D

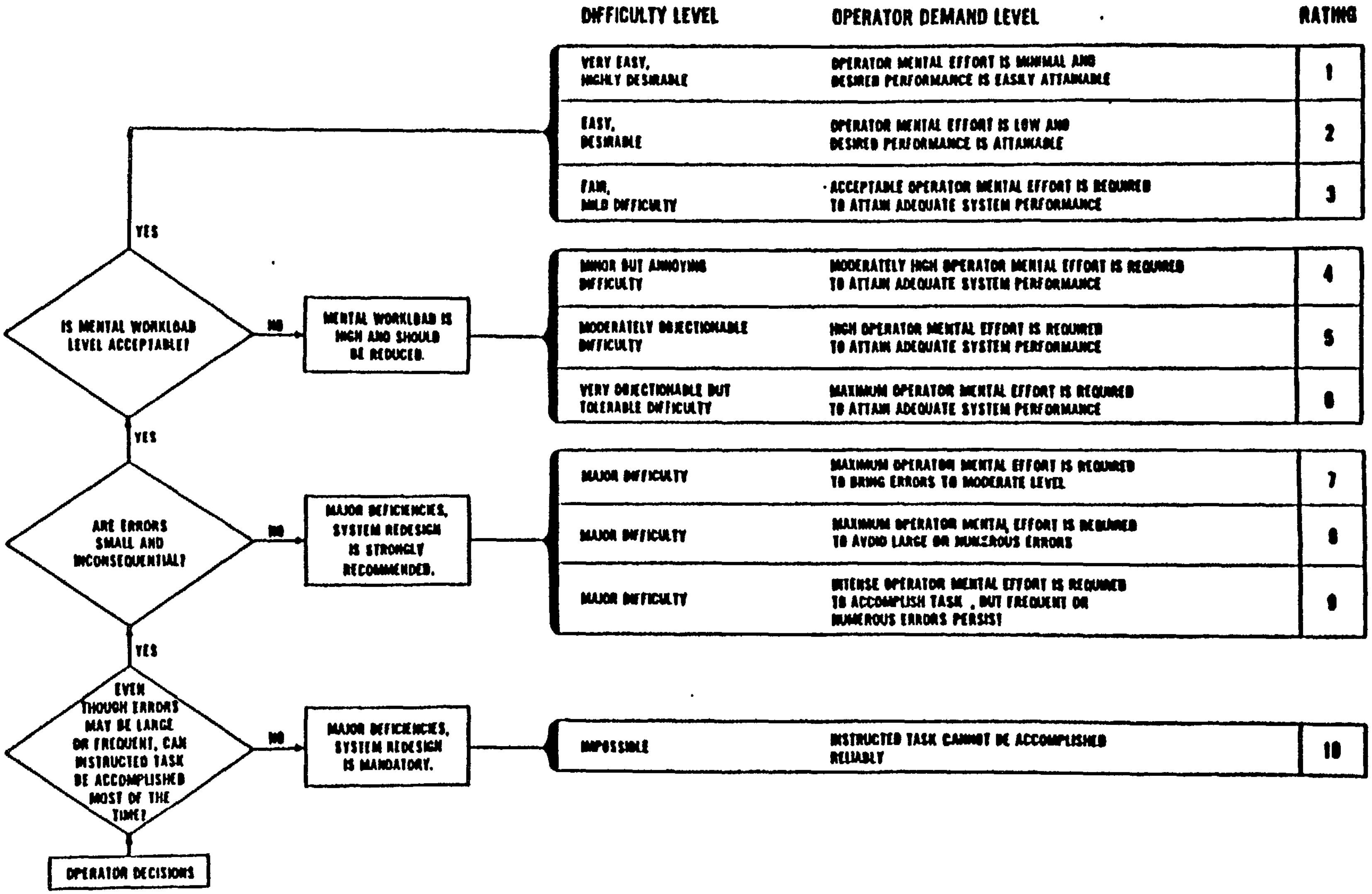
ALTERNATIVE SUBJECTIVE WORKLOAD ASSESSMENT TOOLS:

Modified Cooper Harper (MCH) Scale

Task Load Index (TLX)

Subjective Workload Assessment Technology (SWAT)

Operator Workload (OW) Scale



DIFFICULTY LEVEL	OPERATOR DEMAND LEVEL	RATING
VERY EASY, HIGHLY DESIRABLE	OPERATOR MENTAL EFFORT IS MINIMAL AND DESIRED PERFORMANCE IS EASILY ATTAINABLE	1
EASY, DESIRABLE	OPERATOR MENTAL EFFORT IS LOW AND DESIRED PERFORMANCE IS ATTAINABLE	2
FAIR, MILD DIFFICULTY	ACCEPTABLE OPERATOR MENTAL EFFORT IS REQUIRED TO ATTAIN ADEQUATE SYSTEM PERFORMANCE	3
MINOR BUT ANNOYING DIFFICULTY	MODERATELY HIGH OPERATOR MENTAL EFFORT IS REQUIRED TO ATTAIN ADEQUATE SYSTEM PERFORMANCE	4
MODERATELY OBJECTIONABLE DIFFICULTY	HIGH OPERATOR MENTAL EFFORT IS REQUIRED TO ATTAIN ADEQUATE SYSTEM PERFORMANCE	5
VERY OBJECTIONABLE BUT TOLERABLE DIFFICULTY	MAXIMUM OPERATOR MENTAL EFFORT IS REQUIRED TO ATTAIN ADEQUATE SYSTEM PERFORMANCE	6
MAJOR DIFFICULTY	MAXIMUM OPERATOR MENTAL EFFORT IS REQUIRED TO BRING ERRORS TO MODERATE LEVEL	7
MAJOR DIFFICULTY	MAXIMUM OPERATOR MENTAL EFFORT IS REQUIRED TO AVOID LARGE OR NUMEROUS ERRORS	8
MAJOR DIFFICULTY	INTENSE OPERATOR MENTAL EFFORT IS REQUIRED TO ACCOMPLISH TASK, BUT FREQUENT OR NUMEROUS ERRORS PERSIST	9
IMPOSSIBLE	INSTRUCTED TASK CANNOT BE ACCOMPLISHED RELIABLY	10

Figure D-1. The Modified Cooper Harper (MCH) Scale[1]

Task or Mission Segment: ______________________

Please rate the task or mission segment by putting a mark on each of the six scales at the point which matches your experience.

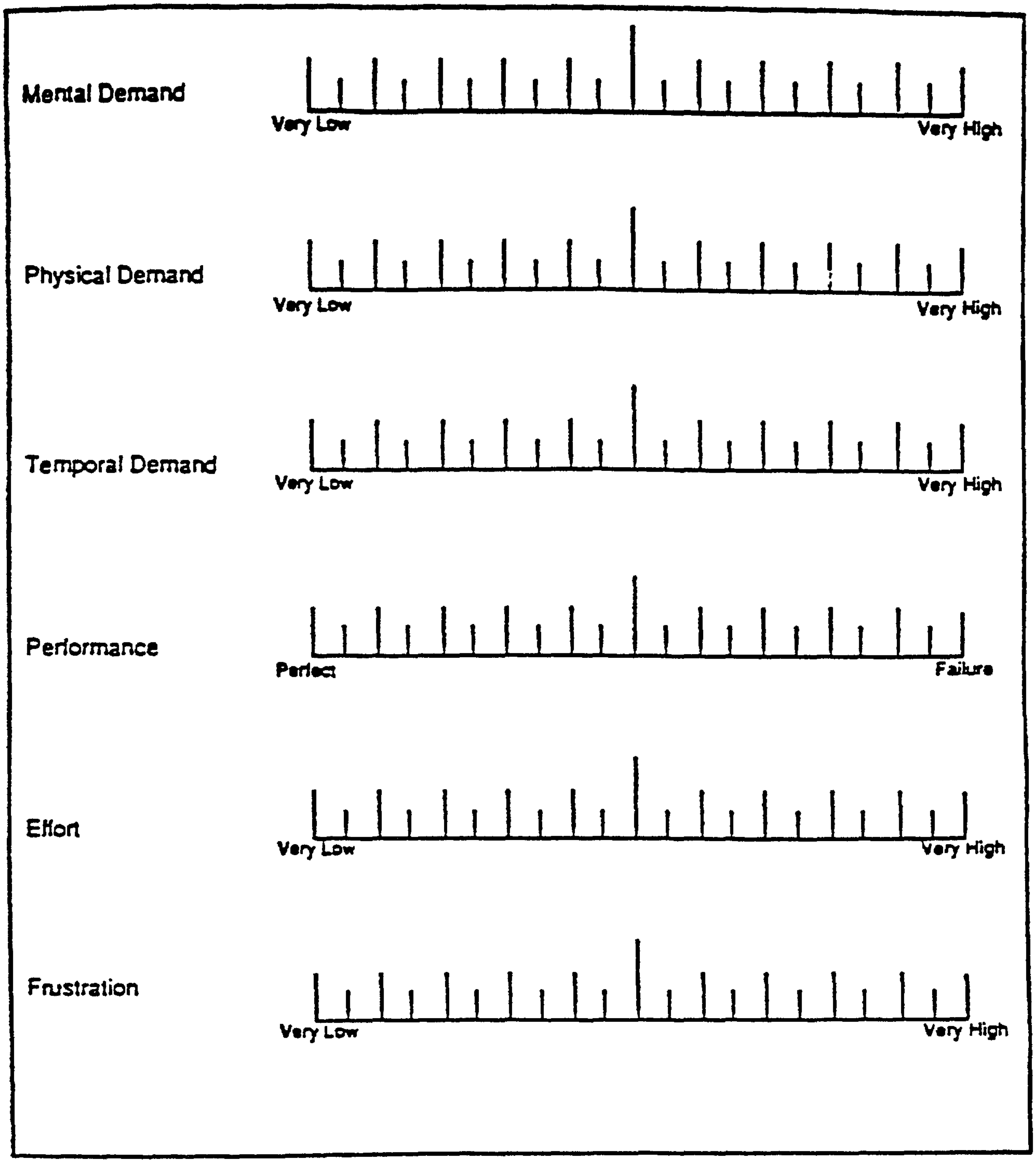


Figure D-2. The TLX Scale[2]

NASA TLX Scale:
Rating scale definitions

Title	Endpoints	Descriptions
MENTAL DEMAND	*Low/High*	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
PHYSICAL DEMAND	*Low/High*	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
TEMPORAL DEMAND	*Low/High*	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
PERFORMANCE	*Perfect/Failure*	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
EFFORT	*Low/High*	How hard did you have to work (mentally and physically) to accomplish your level of performance?
FRUSTRATION LEVEL	*Low/High*	How insecure, discouraged, irritated, stressed, and annoyed versus secure, gratified, content, relaxed, and complacent did you feel during the task?

Figure D-3. TLX Definitions[3]

SWAT Scale

Time load	Mental effort load	Stress load
1. Often have spare time. Interruptions or overlap among activities occur infrequently or not at all.	1. Very little conscious mental effort or concentration required. Activity is almost automatic, requiring little or no attention.	1. Little confusion, risk, frustration, or anxiety exists and can be easily accommodated.
2. Occasionally have spare time. Interruptions or overlap among activities occur frequently.	2. Moderate conscious mental effort or concentration required. Complexity of activity is moderately high due to uncertainty, unpredictability, or unfamiliarity. Considerable attention required.	2. Moderate stress due to confusion, frustration, or anxiety noticeably adds to workload. Significant compensation is required to maintain adequate performance.
3. Almost never have spare time. Interruptions or overlap among activities are very frequent, or occur all the time.	3. Extensive mental effort and concentration are necessary. Very complex activity requiring total attention.	3. High to very intense stress due to confusion, frustration, or anxiety. High to extreme determination and self-control required.

Figure D-4. The SWAT Scale[3]

Task or Mission Segment: ______________________

Please put a mark on the scale at the point which best corresponds to how you rate your overall workload.

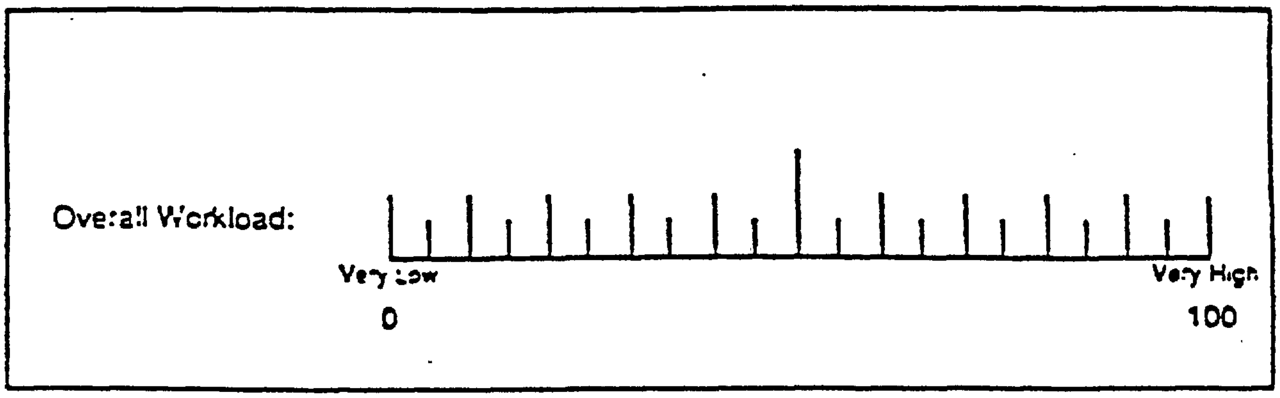


Figure D-5. The OW Scale[2]

REFERENCES

1. Wierwille, W. W., & Casali, J. G. (1983). A validated rating scale for global workload measurement applications. *Proceedings of the Human Factors Society 27th Annual Meeting,* 129-133.

2. Christ, R. E., Hill, S. G., Byers, J. C., Iavecchia, H. M., Zaklad, A. L., Bittner, A. C. (1993, March). *Application and validation of workload assessment techniques* (Technical Report No. 974). Washington, DC: United States Army Research Institute for the Behavioral and Social Sciences.

3. Wickens, C. D. (1992). *Engineering psychology and human performance* (2nd Ed.) New York: HarperCollins. (Pp. 364-411.)

APPENDIX E

AN ACTUARIAL APPROACH TO DEVELOPMENT OF SAFETY-RELEVANT CRITERIA FOR IN-VEHICLE DEVICE USE

BY

WALTER W. WIERWILLE

APPENDIX E. AN ACTUARIAL APPROACH TO DEVELOPMENT OF SAFETY-RELEVANT CRITERIA FOR IN-VEHICLE DEVICE USE

by

Walter W. Wierwille

1. INTRODUCTION

The Task 4 Interim Report (Wierwille, Tijerina, Kiger, Rockwell, Lauber, and Bittner, 1992) outlined an actuarial approach to establishing the safety relevance of in-vehicle device workloads (pp. 1-10 to 1-12). The concept, fundamentally, was to examine an accident data base and then attempt to relate it to in-vehicle task demands. The concept is an extension of Perel's (1976) approach to accident narrative search.

A block diagram of the proposed procedure was presented in the Task 4 Interim Report, which became the guiding philosophy in carrying out the approach. However, small changes were made as the analysis proceeded. Figure 1 shows the revised block diagram of the analysis procedure. The database search procedure has been previously reported in a technical paper (Wierwille and Tijerina, 1993) to which the reader is referred. The results of the four blocks on the left side of Figure 1 are contained in that technical paper. The remainder of the steps depicted in the block diagram are presented herein. These steps emphasize the exposure analysis aspects (right-hand column of Figure 1) as well as the combining of data base search and exposure analysis work (bottom two blocks of Figure 1).

In carrying out the exposure analysis, emphasis was placed on visual demands of in-vehicle devices. While visual workload is not the only form of driver workload, it is by far the most important. Because of the relatively coarse nature of accident data base analyses, it appeared that attempting to introduce other forms of workload, at least in this first analysis, would be unwarranted and unlikely to provide any additional insight.

This paper is organized in accordance with Figure 1, with sections in sequence from the top right to the bottom center of the figure. The earlier technical paper (Wierwille and Tijerina, 1993) is referred to whenever needed and should be on hand when reading this paper.

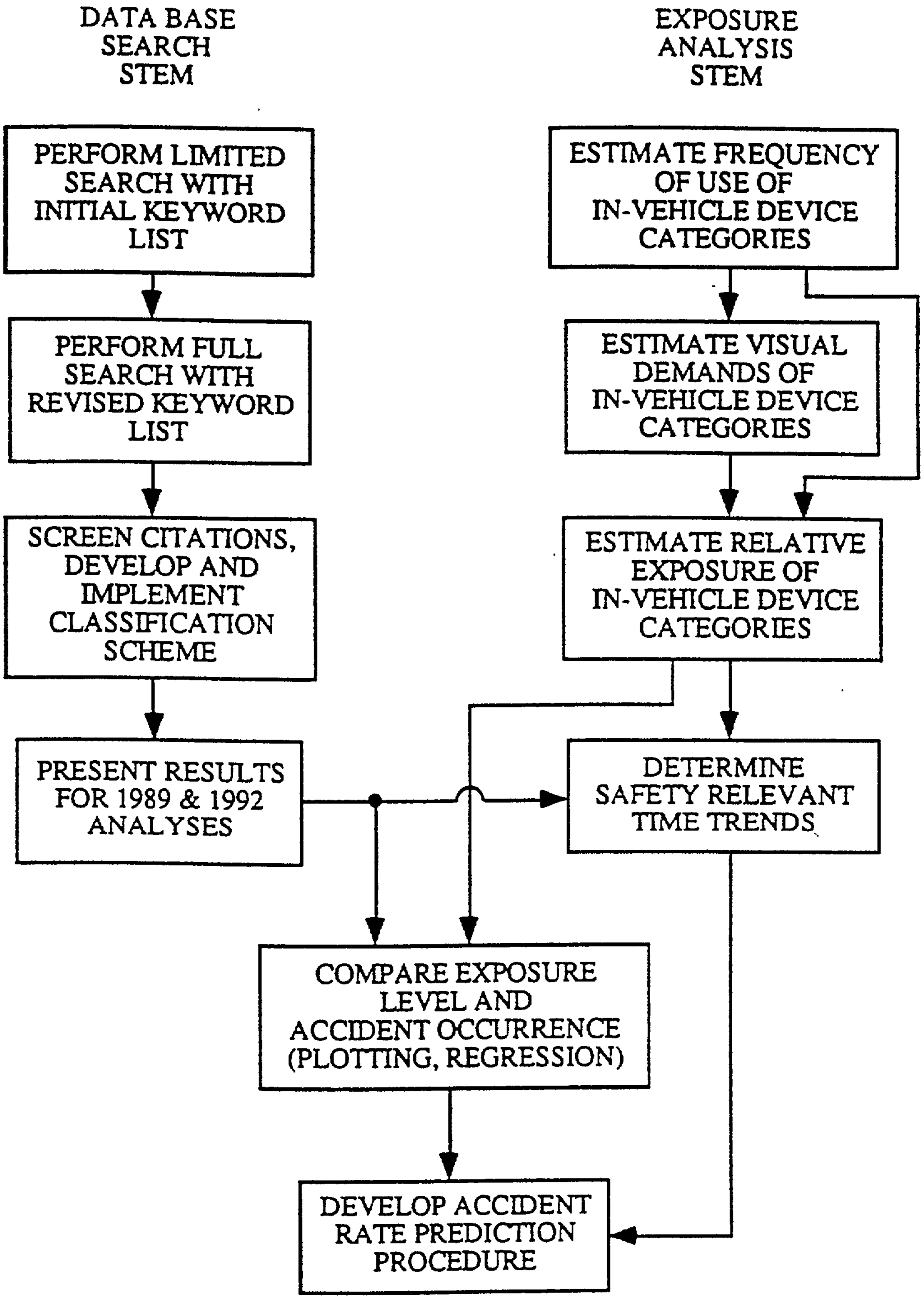


Figure 1. Revised Block Diagram of the Actuarial Approach to Establishing Safety Relevance

2. ESTIMATING FREQUENCY OF USE OF DEVICE CATEGORIES

It must be recognized that various in-vehicle devices are used with different frequencies. For example, the speedometer is used frequently, whereas map reading is performed infrequently. The accident statistics are likely to reflect a combination of frequency of use and the cost in driver resources per use. In fact, it is reasonable to assume that the product of these two factors is reflected in the accident data. Thus, both factors must be taken into account.

The analysis described in this section is limited to devices that appear on almost every production automobile. If a device exists only on a small fraction of vehicles, it must be handled separately. Its exposure is lower than that of devices appearing on most vehicles. Such low-production devices are analyzed separately in Section 6 of this paper.

Three sources of information were found on frequency of use. The first of these is a paper presented at the Transportation Research Board by Bhise, Forbes, and Farber (1986). Although the text of the paper was never published, copies of the transparencies used for the presentation were obtained from the authors. One of the transparencies contained information on usage as shown in Table 1. These data, were gathered using an instrumented vehicle in which drivers' eye and hand positions were videotaped while they drove. The right column in Table 1 has been added to the authors' table, because usage per week of driving is more easily interpreted than usage per year.

The second source of data was found by Michael Perel of NHTSA. A report by Woodson, Conover, Miller, and Selby (1969) of Man Factors, Inc., contains data on <u>relative</u> frequency of use. The data were obtained by observing and recording usage occurrences in three actual driving situations: freeway, suburban, and urban. The average of these occurrences yielded the relative frequency count information shown in Table 2. As can be seen the usage data contain information on some items that are not in the instrument panel per se. Furthermore, it is not clear from the text whether the data include visual reference only or visual reference and manual manipulation. Nevertheless, the data do provide some relative information on usage.The third source is that developed from questionnaire data obtained by Anacapa Sciences (1976). The data are summarized in Table 3. These are the only known questionnaire data available and are in a form that requires further processing before they can be ranked, or placed in frequency-of-use form. To accomplish ranking, it was necessary to weight the various columns. The following weighting equation was developed and used:

$$\begin{aligned}\text{Relative use value} = {} & 20X\ (\%\ \text{at least once a day}) \\ & + 5X\ (\%\ \text{at least once a week}) \\ & + (\%\ \text{at least once a month})\end{aligned}$$

The results of the weighting procedure appear in Table 4, with data arranged from highest to lowest values. As can be seen, the relative weighting does include results in a ranking of usage that seem to coincide with intuition.

Table 1. Relative usage of various in-vehicle devices (Bhise, Forbes, and Farber, 1986). (The right column has been added for ease of interpretation.)

INSTRUMENT PANEL DEVICE	USAGE PER YEAR	USAGE PER WEEK
SPEEDOMETER	48,000	920
TURN SIGNAL	5,300	102
RADIO CONTROLS	2,900	56
CLIMATE CONTROLS	1,900	37
WINDSHIELD WIPERS	1,500	29
FUEL GAGE	1,300	25
HEADLAMP SWITCH	500	10

Table 2. Relative usage of various in-vehicle devices (Woodson, Conover, Miller, and Selby (1969).

DEVICE	RELATIVE USAGE
Braking	65
Steering	61
Accelerator	33
Turn signal control	33
Turn signal indicator	33
Rear view mirror	31
Gear selector control	11
Gear selection indicator	11
Speedometer	10
Ignition switch	7
Engine instruments	4
Windshield wiper/washer	3

Table 3. Summary of frequency of use data from Anacapa studies (1976)

ITEM	RATED FREQUENCY OF USE					
	% AT LEAST ONCE A DAY	% AT LEAST ONCE A WEEK	% AT LEAST ONCE A MONTH	% SEASONALLY	% RARELY OR NEVER	% NOT IN MY CAR
Headlight	54.6	38.4	4.3	0.1	2.6	0
Wiper	3.2	22.5	43.5	2.5	28.2	0.1
Radio	76.2	9.5	1.8	0.1	6.1	6.3
Heater	5.3	14.7	31.1	7.0	40.4	1.4
Defroster	4.6	13.4	29.0	4.7	43.8	4.5
Cigarette Lighter	19.3	5.4	1.8	0.0	58.7	14.8
Ashtray	28.8	6.3	3.7	0.0	58.6	2.7
Hazard Flasher	3.5	3.8	9.3	0.1	67.4	15.9
Air Vent	48.4	20.6	9.4	2.4	16.2	3.0

Table 4. Results of the weighting procedure applied to the Anacapa data.

DEVICE	RELATIVE WEIGHTING
Radio	1573
Headlights	1288
Air Vent	1080
Ashtray	611
Lighter	415
Heater	221
Wiper	220
Defroster	188
Hazard Flasher	98

To take full advantage of the three sources of data, it is necessary to combine them. Doing so, however, requires some interpretation, because the same information does not appear in all three sources. For the Anacapa data, the categories of air vent, heater, and defroster were temporarily combined to form an equivalent "climate controls" category. Similarly the lighter and ashtray categories were combined to form a lighter/ashtray category. In regard to the Woodson, et al., data, certain categories were temporarily deleted, namely those associated with braking, steering, and accelerator. With these modifications, it becomes possible to perform a relative fit of the three sources, as shown in Table 5. As can be seen in the table, anchor points of radio, climate control, and wiper or wiper/washer can be made to align with one another. Then, items can be arranged around these rankings to complete the table.

It is clear from the table that a conflict exists for turn signals and speedometer. The Woodson et al., data show the turn signals to have the highest ranking, whereas the Bhise et al., data show the speedometer to have the highest ranking. Obviously, differences could be a result of scenario or driving environment differences. However, it does seem inconsistent with intuition that the speedometer would be used nine times more often than the turn signals (Bhise et al.) in any type of situation except straight road. Thus, some modification of the Bhise, et al., data for speedometer usage seems to be justified.

In addition, a conflict exists between the Anacapa and Bhise et al., data in regard to the headlamps. If beam switching is included, then the Anacapa data appear reasonable, whereas, without it the Bhise et al., data appear reasonable.

The data in Table 5 can now be used in developing a single graphical representation, as shown in Figure 2. In the graph, average usage per week is shown logarithmically from highest usage at the top to lowest usage at the bottom. The usage numerical values are based primarily on the Bhise et al., data, which are the only absolute data available. (Both the Woodson et al. and Anacapa data are relative data.) When conflicts in ranking exist, they have been resolved by examining the conflict and making prudent adjustments. In the case of headlamps, values are supplied for use with beam switching and without. Hazard flashers have been estimated on the basis that a driver might use them once every five weeks. Figure 2 represents the best available estimates on frequency of use for the items shown.

Unfortunately, the usage items shown in Figure 2 do not correspond exactly with the accident categories developed during the data base search. Consequently, further interpretation is necessary. To begin the process, categories for which data are needed must be determined. This can be accomplished by listing the categories appearing in the data-base search analysis, considering only those for which it is possible to estimate frequency of use. (An example of a category for which frequency of use data cannot be obtained is "interaction with another person or animal in vehicle.") Once the relevant categories are listed, their frequency of use can be estimated by comparison with (and interpolation of) the data in Figure 2. As in the case of previous analysis, estimation and judgment are necessary to complete the frequency-of-use

Table 5. Best fit of frequency-of-use data from three sources.

ANACAPA	Woodson, et al.	Bhise, et al.
	Turn-Signals 33	Speedometer 920
	Mirrors 31	
	Speedometer 10	Turn-Signals 102
Radio 1573		Radio 56
Climate 1479		Climate 37
Headlamp 1288		
Lighter/Ashtray 1026	Engine Inst. 4	
Wiper 220	Wiper/Washer 3	Wiper 29
		Fuel Gage 25
		Headlamp 10
Hazard 98		

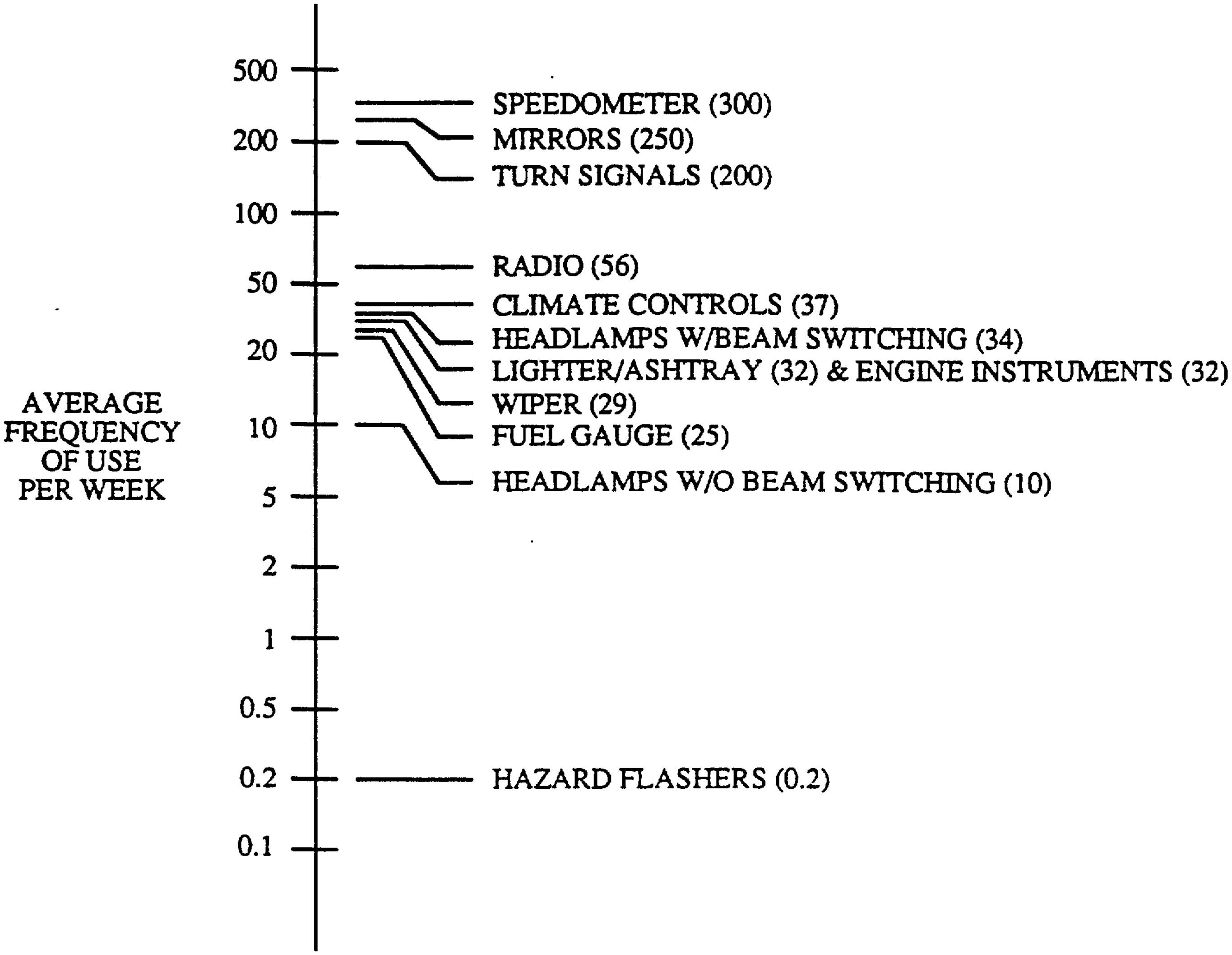


Figure 2. Composite frequency of use chart based on available data and prudent interpretation.

analysis. Table 6 contains the results, with frequency of use listed in descending order. The table has been developed assuming that the vehicle is in motion. The table is the culmination of all analysis work on frequency of use, relevant to the accident data base search.

It should be noted that the category of "turn signals" has been deleted from Table 6. Turn signal-related accidents could not be searched because there were more than 10,000 citations. Many accidents involve failure to signal, or signaling inappropriately. The narratives for these could not be separated from narratives for accidents in which the driver's attention was diverted by using turn signals or checking the turn signal indicators. Similarly, hazard flashers could not be examined because many "non-workload" accidents involve hazard flashers. The categories listed in Table 6 are ones for which accident occurrence data have been obtained.

3. ESTIMATING VISUAL DEMAND OF DEVICE CATEGORIES

There are several good sources of data on the visual demands created by in-vehicle devices. These have been previously reviewed and presented in Chapter 2.0 of the Task 4 Interim Report (Wierwille et al., 1992), as well as in Wierwille (1993). The reader is referred to these documents for a more comprehensive review. In this analysis of visual demands, three sources were used. The first (Rockwell, 1987) provides glance duration information for radios and for mirrors and is summarized in Table 7. The second source is that of Bhise et al. (1986) and is summarized in Table 8. It consists of glance duration and number of glances required for speedometer, fuel gauge, clock, radio, and multiple pushbuttons. The third source is the most comprehensive (Dingus, Antin, Hulse, and Wierwille (1989), as shown in Table 9. It contains information for a variety of conventional in-vehicle tasks and also for navigation tasks using an in-car display. The three references, taken together provide a good background on which to base visual demand estimates.

When the three sources are examined, it becomes clear that the agreement is excellent. Accordingly, all three can be used in estimating visual demands of device categories for which accident data exist. Because the Dingus et al. data are the most comprehensive, they have been relied upon most heavily. The other two sources (Bhise et al., and Rockwell) can be used to supplement the Dingus et al., data.

It must be mentioned that visual demand of in-vehicle devices involves two important parameters, as discussed in Wierwille (1993), namely, mean single glance time and mean number of glances, required to service the in-vehicle device. Both parameters are needed to obtain an estimate of visual demand.

The results of the visual demand estimation process appear in Table 10. The data are presented in the same order as the ranked frequency-of-use data (Table 6). Both mean single glance times and mean number of glances have been estimated.

Table 6. Listing of items from accident data base categorization analysis and estimated frequency of use (while the vehicle is in motion).

Category	Frequency of Use/Week
Speedometer	300
Mirrors	250
Standard Radio	56
Climate Controls	37
Smoking/Lighting	32
Wiper/Washer	29
Fuel Gauge	25
High Beam Indicator	24
Clock	15
Vent	15
Heater & Air Conditioner	15
Lock, Side Window, and Related Hardware	15
Visor	12
Gearshift	10
Defroster/Defogger	7
Seat Belt	5
Seat	3
Personal Timepiece	3
Map	1

Table 7. Summary of glance duration data for conventional tasks (Rockwell, 1987).

	Study	# Runs	x	Median	s	5%	95%
RADIO	A	35	1.27	1.20	.48	.82	2.16
	B	100	1.28	1.29	.50	.89	1.83
	C	72	1.42	1.30	.42	.80	2.50
LEFT MIRROR	A	35	1.06	.96	.40	.80	.20
	*B	100	1.22	1.15	.28	.94	1.80
	C	72	1.10	1.10	.33	.70	1.70

* Commanded mirror looks of discrimination.

Note: All data given in seconds.

Table 8. Summary of results presented by Bhise, Forbes, and Farber (1986).

Tasks Requiring a Single Glance

Task	Mean Glance Duration (Seconds)
Read Analog Speedometer	
• Normal	0.4 to 0.7
• Check	0.8
• Exact Value	1.2
Read Analog Fuel Gauge	1.3
Read Digital Clock	1.0 to 1.2

Tasks Requiring Several Glances

Task	Number of Glances	Mean Glance Duration (Seconds)
Turn on Radio, Find Station, Adjust Volume	2 to 7	1.1
Read all Labels on a 12-Button Panel	7 to 15	1.0

Table 9. Average length and number of in-car glances for a variety of conventional and navigation tasks (Dingus et al., 1989).

	In-car Single Glance Length		Number of Glances	
Task	Mean	Standard Deviation	Mean	Standard Deviation
Speed	0.62	0.48	1.26	0.40
Following Traffic	0.75	0.36	1.31	0.57
Time	0.83	0.38	1.26	0.46
Vent	0.62	0.40	1.83	1.03
Destination Direction	1.20	0.73	1.31	0.62
Remaining Fuel	1.04	0.50	1.52	0.71
Tone Controls	0.92	0.41	1.73	0.82
Info. Lights	0.83	0.35	2.12	1.16
Destination Distance	1.06	0.56	1.73	0.93
Fan	1.10	0.48	1.78	1.00
Balance	0.86	0.35	2.59	1.18
Sentinel	1.01	0.47	2.51	1.81
Defrost	1.14	0.61	2.51	1.49
Fuel Economy	1.14	0.58	2.48	0.94
Correct Direction	1.45	0.67	2.04	1.25
Fuel Range	1.19	1.02	2.54	0.60
Cassette Tape	0.80	0.29	2.06	1.29
Temperature	1.10	0.52	3.18	1.66
Heading	1.30	0.56	2.76	1.81
Zoom Level	1.40	0.65	2.91	1.65
Cruise Control	0.82	0.36	5.88	2.81
Power Mirror	0.86	0.34	6.64	2.56
Tune Radio	1.10	0.47	5.91	2.39
Cross Street	1.66	0.82	5.21	3.20
Roadway Distance	1.53	0.65	5.78	2.85
Roadway Name	1.63	0.80	6.52	3.15

Note: Glance length given in seconds.

Table 10. Estimated visual demand parameters for categories appearing in the accident data base analysis. (Order is the same as that in Table 6.)

Category	Mean Single Glance Time	Mean Number of Glances
Speedometer	0.62	1.26
Mirrors	1.00	1.00
Standard Radio	1.20	3.50
Climate Controls	1.10	1.75
Smoking/Lighting	1.50	4.00
Wiper/Washer	1.10	1.20
Fuel Gauge	1.30	1.20
High Beam Indicator	0.62	1.00
Clock	0.83	1.26
Vents	0.62	1.83
Heater & Air Conditioner	1.10	1.75
Lock, Side Window, and Related Hardware	1.40	1.60
Visor	0.80	2.00
Gearshift	1.50	1.75
Defroster/Defogger	1.10	1.20
Seat Belt	1.50	2.00
Seat	1.50	2.50
Personal Timepiece	0.83	1.26
Map	1.70	5.00

Note: Mean single glance length in seconds

It should be mentioned that some categories involve multiple activities. For example, for standard radio, the driver might simply adjust the volume, or the driver might tune the radio to a specific digital frequency. The former requires little in the way of visual demand, whereas the latter requires a great deal. Thus, for such a category, an average visual glance time and an average number of glances must be estimated.

4. DETERMINE EXPOSURE LEVEL OF IN-VEHICLE DEVICES

There are no set standards for determining exposure level caused by in-vehicle devices. However, researchers have long suspected that any time the driver's resources are allocated to invehicle devices, there is an increase in exposure (that is, an increase in the likelihood or risk of being involved in an accident). Indeed the main reason for examining the visual demands of invehicle devices has been the tacit assumption that the greater such demands are, the more likely it is that their use will result in an accident.

The two previous sections have been directed at obtaining the components that should go into an exposure level estimate, namely visual demand and number of times a device is used per unit time (say, for example, per week). It would seem reasonable that the exposure should be related to the visual demand per se multiplied by the frequency of use.

As previously indicated, visual demand can be assessed in terms of two parameters, mean single glance time and mean number of glances. These two parameters are important because drivers ordinarily perform visual sampling to complete in-vehicle tasks. Both parameters are needed to obtain an assessment of visual demand.

To a first approximation the product of mean single glance time and mean number of glances equals mean total glance time. This is the total visual glance time that is needed to service the in-vehicle device each time it is used. Correspondingly, total glance time can be multiplied by frequency of use to assess exposure. In terms of the parameters, exposure then becomes:

$$\begin{array}{c}\text{TYPE 1}\\ \text{EXPOSURE}\end{array} = \left(\begin{array}{c}\text{mean}\\ \text{single-glance}\\ \text{time}\end{array}\right) \text{X} \left(\begin{array}{c}\text{mean}\\ \text{number of}\\ \text{glances}\end{array}\right) \text{X} \left(\begin{array}{c}\text{frequency}\\ \text{of}\\ \text{use}\end{array}\right)$$

This exposure represents the total time that the driver's eyes are allocated to the in-vehicle device, say, per week (while the vehicle is in motion). This type of exposure has been called Type I to distinguish it from two additional types that are yet to be described.

Type I exposure involves the assumption that a glance into the vehicle of, say, 1.6 seconds involves a risk or exposure that is exactly twice as great as that of a glance into the vehicle of 0.8 second. In other words, exposure increases linearly with mean single-glance time. Such an assumption may not be fully warranted, however. It can be argued that exposure increases more rapidly than a simple linear function of mean single glance time, because longer eyes-off-road times substantially increase the likelihood of not detecting a hazard in the forward view, or at least of not detecting a hazard soon enough to prevent an accident. Thus, it is argued that mean single glance time should enter the exposure assessment with substantially heavier weighting for longer times. Two alternative types of exposure are thus defined, one in which mean single glance time is taken to the three-halves power and one in which it is taken to the second power (that is, it is squared) as shown in the following equations:

$$\begin{array}{c}\text{TYPE 2}\\ \text{EXPOSURE}\end{array} = \left(\begin{array}{c}\text{mean}\\ \text{single-glance}\\ \text{time}\end{array}\right)^{3/2} \times \left(\begin{array}{c}\text{mean}\\ \text{number of}\\ \text{glances}\end{array}\right) \times \left(\begin{array}{c}\text{frequency}\\ \text{of}\\ \text{use}\end{array}\right)$$

$$\begin{array}{c}\text{TYPE 3}\\ \text{EXPOSURE}\end{array} = \left(\begin{array}{c}\text{mean}\\ \text{single-glance}\\ \text{time}\end{array}\right)^{2} \times \left(\begin{array}{c}\text{mean}\\ \text{number of}\\ \text{glances}\end{array}\right) \times \left(\begin{array}{c}\text{frequency}\\ \text{of}\\ \text{use}\end{array}\right)$$

All three types of exposure are carried through the remainder of the analysis.

Using Tables 6 and 10, it is possible to calculate all three types of exposure for the accident data base categories. The results appear in Table 11.

5. RELATING EXPOSURE LEVEL TO ACCIDENT OCCURRENCE

All of the relevant information has now been developed to allow plotting of accident occurrence as a function of exposure level for various accident categories. Accident occurrence data are presented in Figures 1 through 15 of Wierwille and Tijerina (1993). These figures are for 1989, as previously described. They represent a full year of accident data base results and are probably the most accurate accident occurrence data available. Exposure data are contained in Table 11 (of this paper) for the three types of exposure.

For Type 1 and Type 2 exposures, regression analyses were conducted. These are shown in Figures 3 through 6. Figure 3 shows the plot of accident occurrence vs. exposure, the corresponding regression line, and the 95% confidence limits on the regression line for all of the Type 1 exposure data. As can be seen in the figure, all of the data points except one fall near the regression line, indicating qualitatively that there is a relationship between exposure and accident occurrence. The slope of the regression line is significantly different from zero, $p = 7.7\times10^{-8}$. The correlation coefficient associated with the data is $R = 0.898$, which is also significant, $p < 0.00001$. (It should be mentioned that two points in the graph are not plotted because they fall on top of existing points. Thus, 19 values are plotted with only 17 points appearing.)

The outlying point is that associated with the speedometer. The speedometer is used very often, but does not appear to cause as many accidents as the regression line would seem to predict.

Table 11. Exposure values for categories appearing in the accident data base analysis.

	Exposure*		
Category	Type 1	Type 2	Type 3
Speedometer	234	185	145
Mirrors	250	250	250
Standard Radio	235	257	287
Climate Controls	77	75	80
Smoking/Lighting	192	235	288
Wiper/Washer	38	40	42
Fuel Gauge	39	44	51
High Beam Indicator	15	12	9
Clock	16	14	13
Vents	17	13	11
Heater & Air Conditioner	29	30	32
Lock, Side Window, and Related Hardware	34	40	47
Visor	19	17	15
Gearshift	26	32	39
Defroster/Defogger	9	10	10
Seat Belt	15	18	22
Seat	11	14	3
Personal Timepiece	3	3	3
Map	9	11	14

* Values are rounded to the nearest integer.

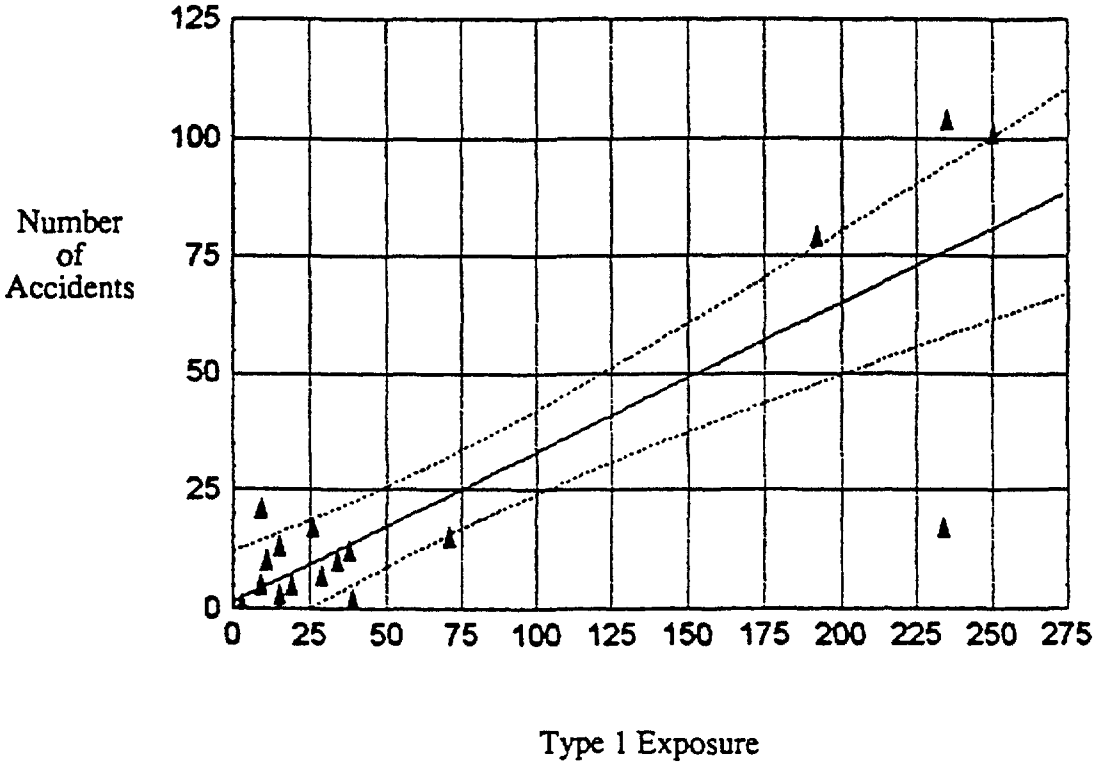


Figure 3. Plot of accident occurrence vs. Type 1 exposure, with regression line and 95% confidence limits shown.

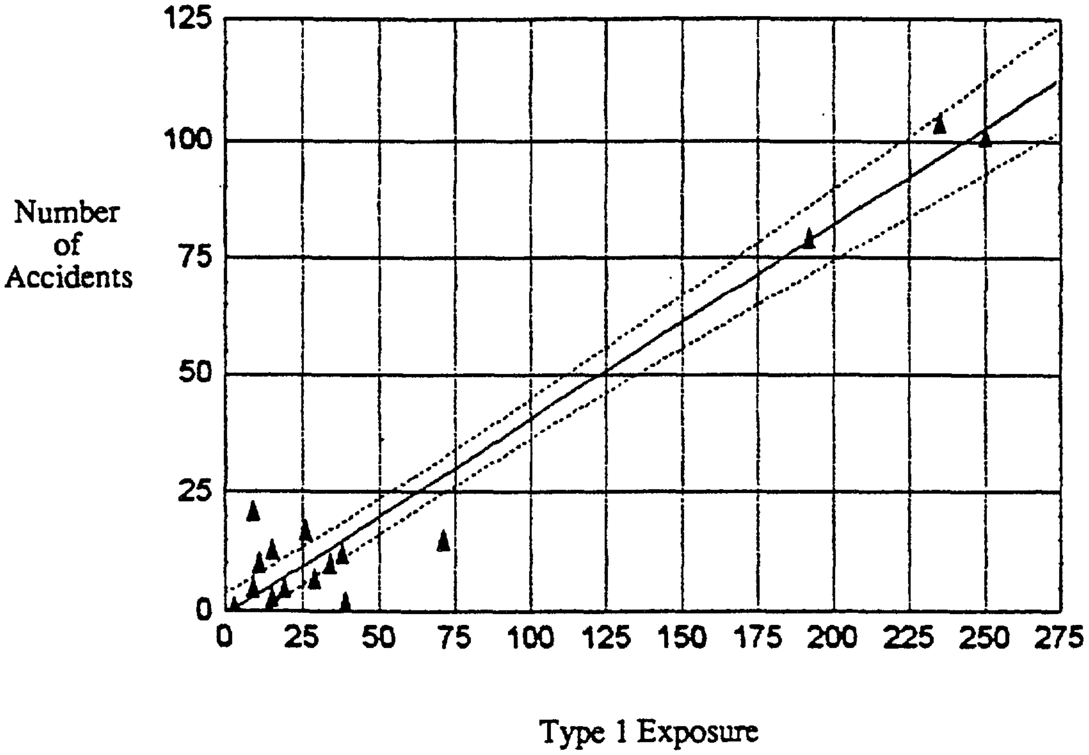


Figure 4. Plot of accident occurrence vs. Type 1 exposure, with regression line and 95% confidence limits shown. (Outlier removed from data set.)

This under-representation may be a result of the prime location of the speedometer directly below the forward line of sight, and also a result of the overlearned usage of the speedometer.

Because of the appearance of the speedometer outlier, the regression analysis for Type 1 exposure was rerun with this point removed. The result appears in Figure 4. As can be seen, there is a much better fit of the regression line to the data and the 95% regression line confidence limits are now much narrower. The slope of the regression line is of course significantly different from zero, $p = 6.35\text{x}10^{-14}$, and the corresponding correlation coefficient is now 0.982 which is also significant, $p < 0.000001$.

The corresponding results for Type 2 exposure are shown in Figures 5 and 6. Figure 5 shows results with the outlier included, and Figure 6 shows them with the outlier deleted. In Figure 5 the slope is significantly different from zero, , and the corresponding correlation coefficient is 0.941 which is significant . Similarly in Figure 6, the slope of the regression line is significantly different from zero, and the corresponding correlation coefficient is 0.982 which is significant .

For Type 3 exposure, no attempt was made at linear regression, because the plotted data clearly showed a nonlinear trend. This nonlinearity is even more pronounced when the outlier point for speedometer is included. Accordingly, an analytical function was developed by trial and error and is shown along with the plotted data in Figure 7. The curve shown has the equation

$$Y = 0.185X_3 + 0.1e^{0.0223X_3}$$

where X_3 is the Type 3 exposure value and Y is the corresponding level of accident occurrence. It should be noted in Figures 3, 5, and 7 that the outlier point moves closer to the other data as exposure goes from Type 1 to Type 2 to Type 3. This is a result of the short mean single glance time for the speedometer. When this smaller glance time is taken to the three-halves power or second power, it reduces the corresponding value of exposure, causing the point to move to the left when plotted, relative to most of the other points.

6. SAFETY RELEVANT TIME TRENDS

Additional accident trend information can be obtained by directly comparing the 1989 and 1992 accident occurrences by category. Three such comparisons are presented here: cellular telephone-related, standard radio-related, and special radio-related.

The Cellular Telecommunications Industry Association (CTIA) has estimated that in 1989 there were 3.5 million cellular telephone users. They have similarly estimated that in 1992 there were 11 million such users, a very substantial increase. Figure 11 of Wierwille and Tijerina (1993) shows that there were 11 cellular phone accidents in 1989 and Figure 18 of that paper shows that

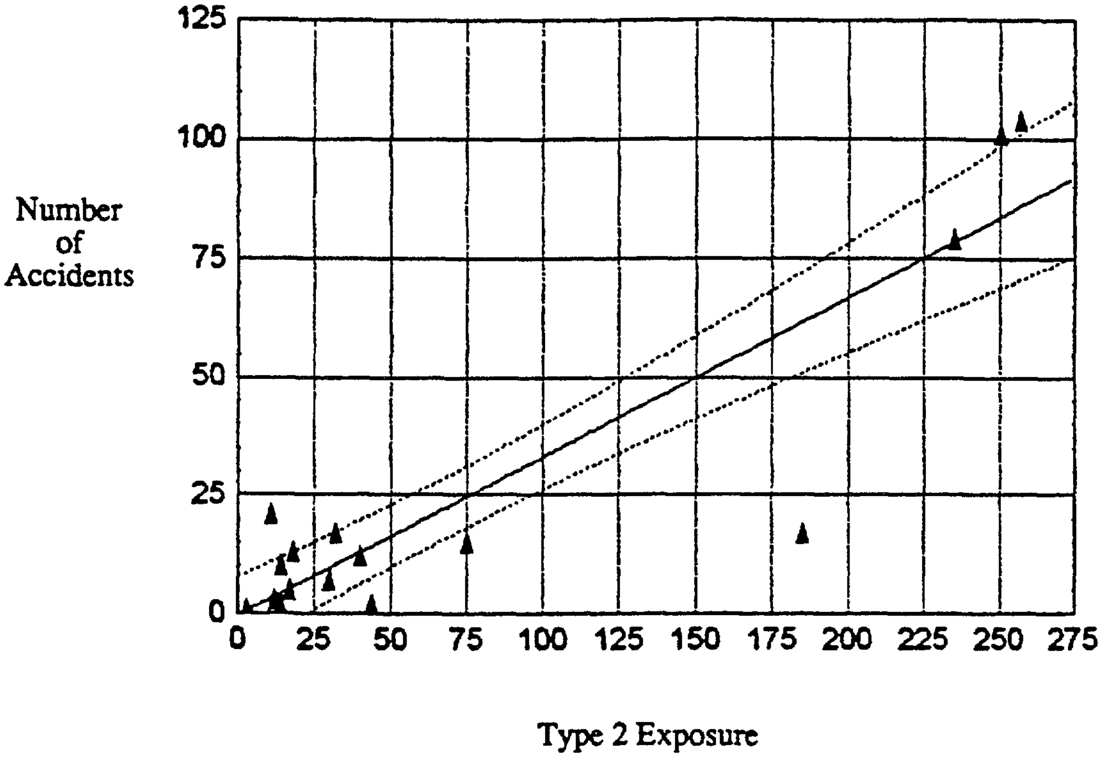


Figure 5. Plot of accident occurrence vs. Type 2 exposure, with regression line and 95% confidence limits shown.

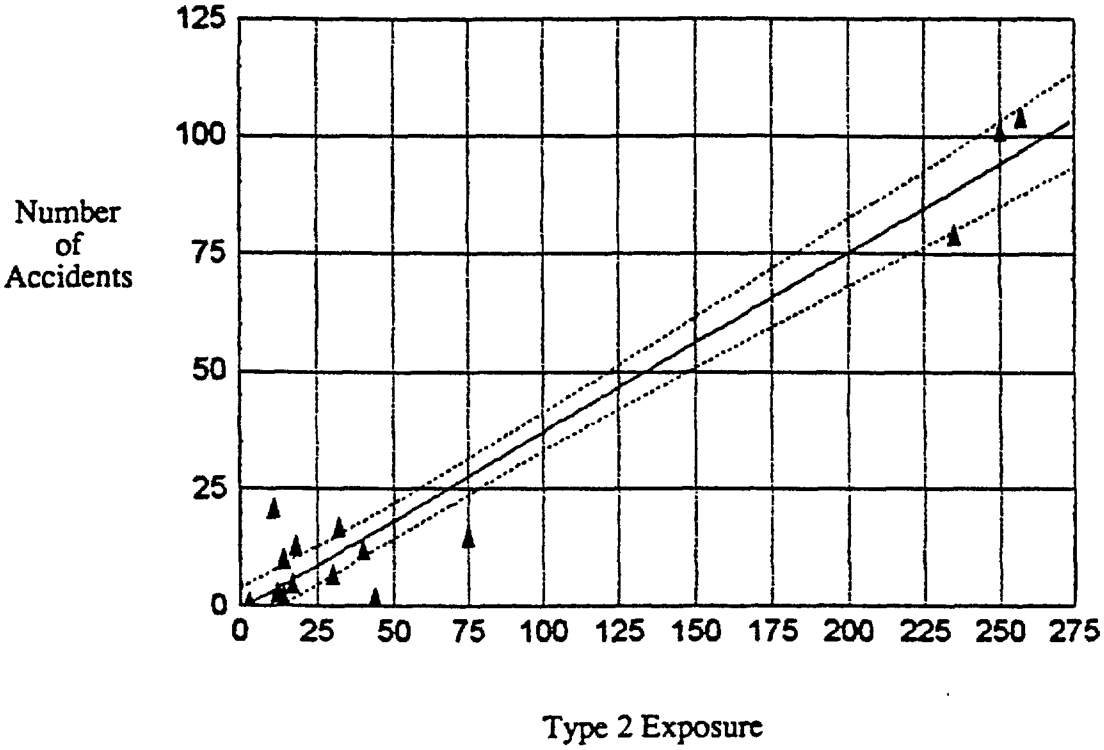


Figure 6. Plot of accident occurrence vs. Type 2 exposure, with regression line and 95% confidence limits shown. (Outlier removed from data set.)

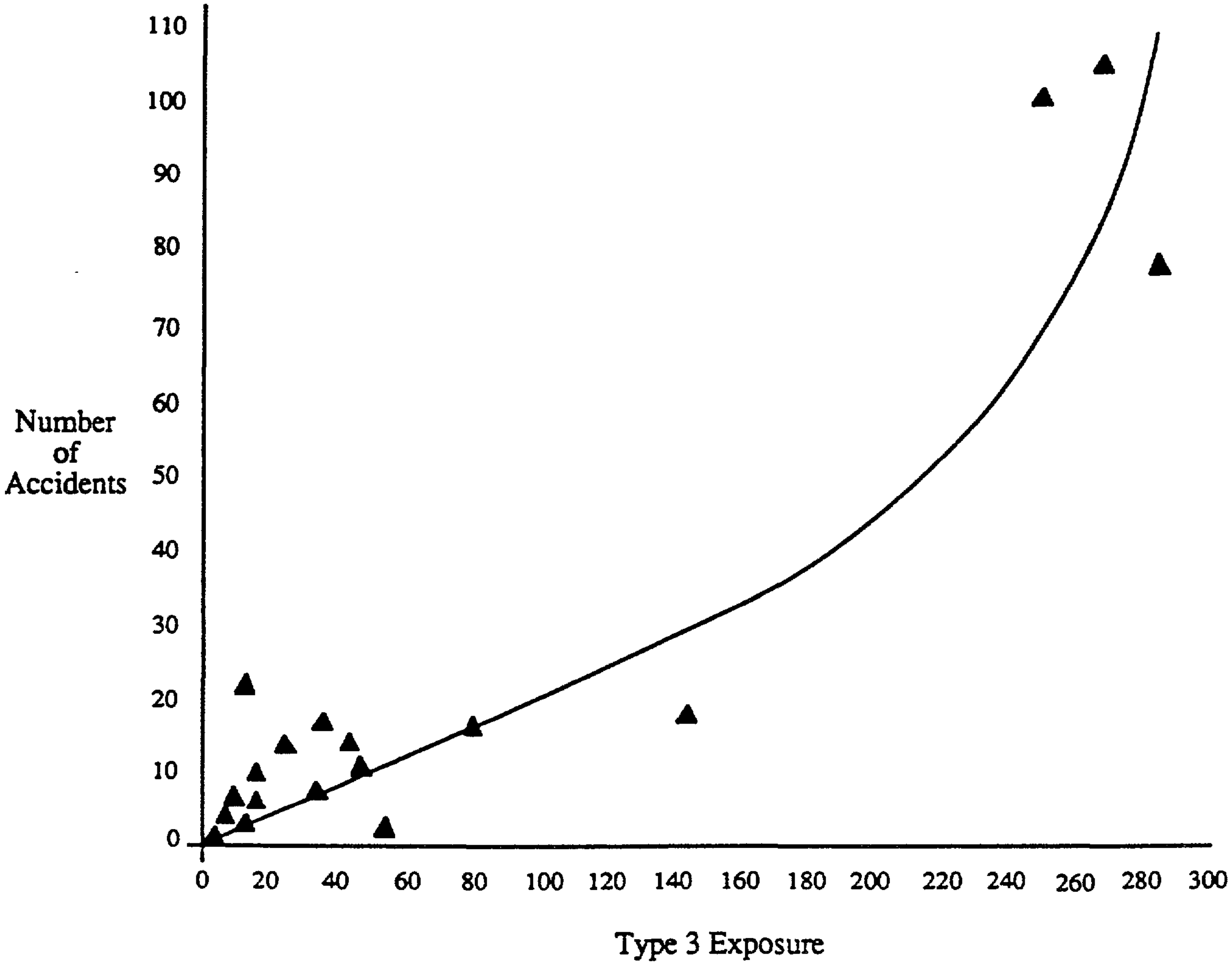


Figure 7. Plot of accident occurrence vs. Type 3 exposure, with analytical curve fitted to the data.

there were estimated to be 27 in 1992. Figure 8 herein compares those data in plotted form. It is very clear from the figure that the number of accidents associated with cellular phones is rising rapidly as the number of these devices increases in the vehicle population. For standard radio-related accidents, data can be obtained from Figures 5 and 16 of Wierwille and Tijerina (1993). The data are compared herein in Figure 9 and suggest that there may be a slight increase in accident occurrence between 1989 and 1992. This trend could be a result of the increasing complexity of recent vehicle radios, which may be inducing somewhat greater visual demands on the driver.

Finally, special radio-related accident data are presented in Figures 6 and 17 of Wierwille and Tijerina (I 993) and are compared here in plotted form in Figure 10. While the data are sparse, there appears to be a decreasing trend in special radio-related accidents. This trend may be a result of the apparent decline in the popularity of CB radios over the last several years. No figures were found for CB usage. However, it is probable that their use did decline between 1989 to 1992.

7. PREDICTION OF ACCIDENT OCCURRENCE

The previous analyses and data make it possible to predict the expected number of accidents that will occur when a new device is introduced into the vehicle population. There are several ways of doing this, and each would result in slightly different estimates. To avoid unnecessary complexity, a specific procedure has been developed and is presented here.

In examining the regression lines presented earlier, it is clear that Type 2 exposure provides the best fit to the data (Figure 6). Accordingly, the accident occurrence prediction model is based on Type 2 exposure. In Figure 6, the slope of the regression line is 0.375 accident per exposure unit. The regression line very nearly passes through zero. Thus, the number of accident narratives cited in North Carolina is 0.375 times the Type 2 exposure value for the in-vehicle device.

The steps for prediction are then as follows:

1. Determine the mean single glance time and the mean number of glances required to service the in-vehicle device.

2. Determine the number of times per week that the device is likely to be used by the average user.

3. Calculate the Type 2 exposure:

$$\begin{array}{c}\text{TYPE 2}\\\text{EXPOSURE}\end{array} = \left(\begin{array}{c}\text{mean}\\\text{single-glance}\\\text{time}\end{array}\right)^{3/2} \times \left(\begin{array}{c}\text{mean}\\\text{number of}\\\text{glances}\end{array}\right) \times \left(\begin{array}{c}\text{frequency}\\\text{of use}\\\text{per week}\end{array}\right)$$

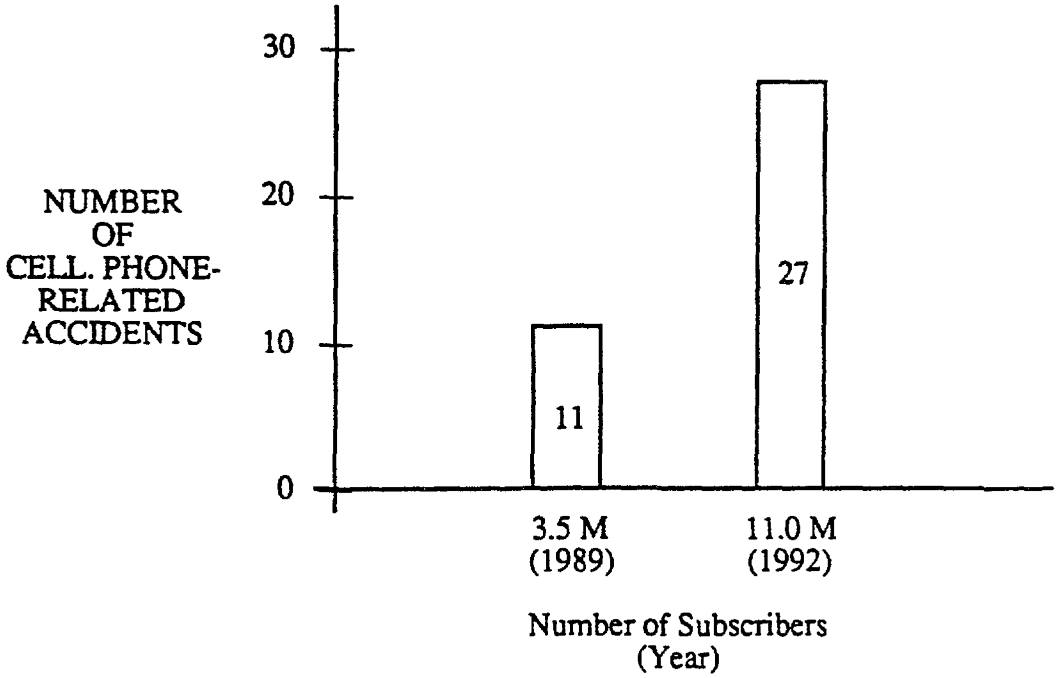


Figure 8. Number of cellular phone-related accidents as a function of number of subscribers in millions (also, years).

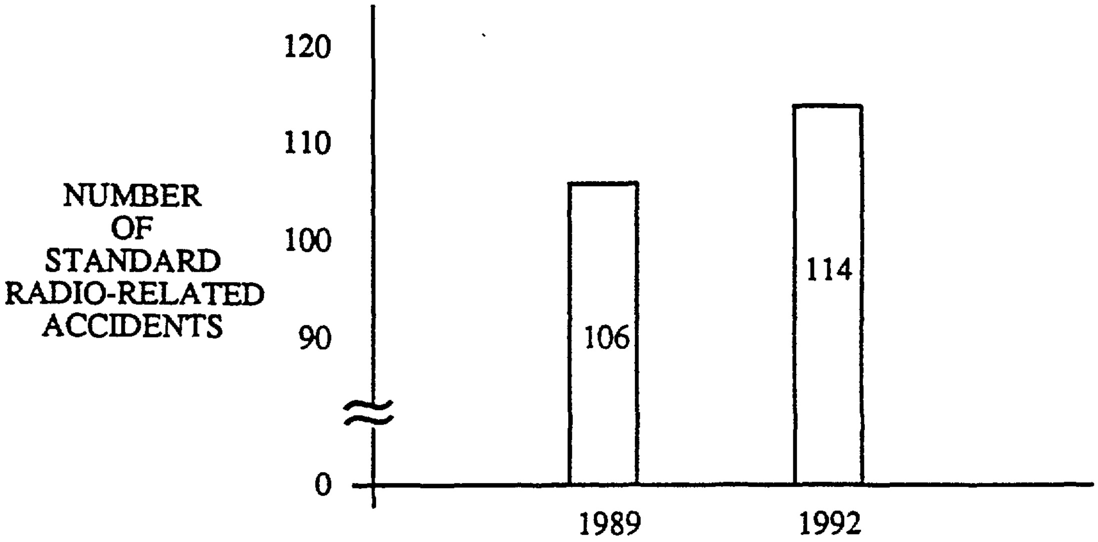


Figure 9. Number of standard radio-related accidents as a function of year.

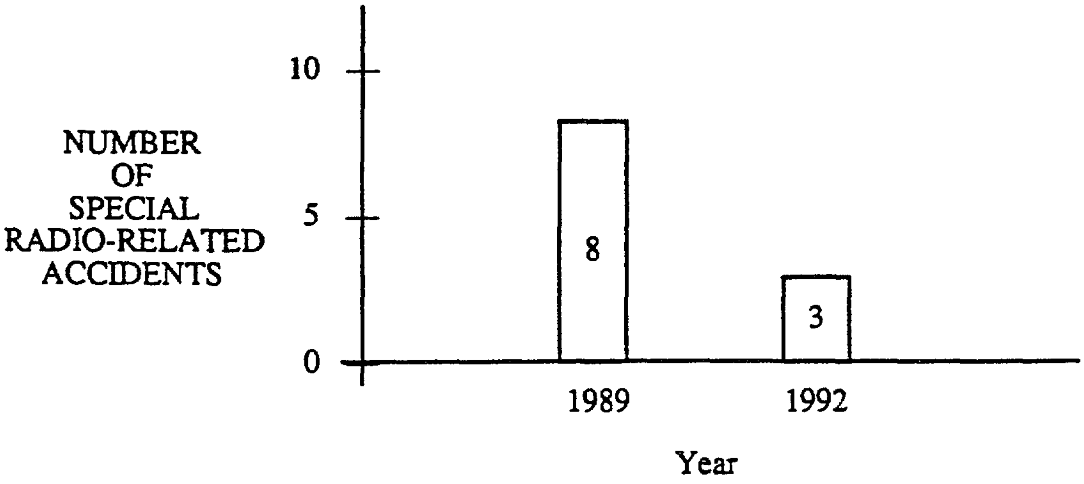


Figure 10. Number of special radio-related accidents as a function of year.

4. Calculate the number of accident narratives that would be projected to appear in North Carolina (which is assumed to be a typical state.)

$$\begin{array}{c}\text{Number of}\\ \text{occurrences,}\\ \text{N.C.}\end{array} = 0.375 \times \left(\begin{array}{c}\text{Type 2}\\ \text{Exposure}\end{array}\right)$$

5. Multiply the number of occurrences in North Carolina by 50 to estimate the occurrences nationwide.

$$\begin{array}{c}\text{Number of}\\ \text{occurrences,}\\ \text{U.S.}\end{array} = 50 \times \left(\begin{array}{c}\text{Number of}\\ \text{occurrences,}\\ \text{N.C.}\end{array}\right)$$

6. The result in 5. is based on the assumption that approximately 90% of all vehicles would have the device installed. If more or fewer have the device, a correction must be made as follows:

$$\begin{array}{c}\text{Number of}\\ \text{occurrences, U.S.}\\ \text{corrected for}\\ \text{availability}\end{array} = \left(\begin{array}{c}\text{Number of}\\ \text{occurrences,}\\ \text{U.S.}\end{array}\right) \times \left(\frac{\begin{array}{c}\text{\% of vehicles}\\ \text{equipped}\end{array}}{90}\right)$$

As an example of the accident occurrence prediction procedure, consider the following hypothetical example:

> A traffic advisory display is to be introduced widely in the U.S. The display has been determined to require a mean single glance time of 1.35 seconds and to require 3.0 glances per use. Furthermore, it is expected that the device would be used 20 times per week. By 1998, it is anticipated that 30% of all vehicles would have such devices installed.

For this example, the Type 2 exposure is

$$\begin{array}{c}\text{TYPE 2}\\ \text{EXPOSURE}\end{array} = (1.35)^{3/2}\,(3.0)\,(20) = 94$$

The corresponding N.C. occurrences (for 90% of vehicles so equipped) is

$$\begin{array}{c}\text{Number of}\\ \text{occurrences,}\\ \text{N.C.}\end{array} = 0.375\,(94) = 35$$

The corresponding U.S. occurrences (for 90% of vehicles so equipped) is

$$\begin{matrix}\text{Number of}\\ \text{occurrences,}\\ \text{U.S.}\end{matrix} = 50\,(35) = 1750$$

and
The corresponding U.S. occurrences, corrected for availability is

$$\begin{matrix}\text{Number of}\\ \text{occurrences, U.S.}\\ \text{corrected for}\\ \text{availability}\end{matrix} = 1750\left(\frac{30}{90}\right) = 583$$

The procedure described in this section does not take into account the fact that many accidents cannot be attributed directly to a specific device. Figure 1 of Wierwille and Tijerina (1993), shows that many accidents involving visual allocation cannot be classified. Therefore, it must be remembered that predicted values are necessarily conservative and that the actual number of accidents attributed to a given in-vehicle device may be much greater than the value obtained from steps 4, 5, and 6 of the estimation procedure. Nevertheless, the procedure can be particularly helpful from a relative standpoint. The previous example for a traffic advisory display could, for example, be compared with standard radio usage to determine the relative likelihood of accidents with such a device.

8. CONCLUSIONS

The analyses contained in this paper provide a compelling argument that the amount and frequency of visual attention to in-vehicle devices is directly safely relevant. The clustering of data about the regression line relating exposure, particularly Type 2 exposure, to accident occurrence provides a powerful argument that the relative number of accidents is directly related to visual resource allocation for in-vehicle tasks. Considering the "noise level" associated with actuarial data, it is surprising that the results obtained are as good as they are. It should be mentioned that the data and analyses that have been presented are as unbiased as they could be made. All "exposure" values were computed before any plotting was done, and no exposure value was ever changed to make it "fit" closer to the trend. In other words, the analyses were performed with the philosophy of "let the chips fall where they may."

It should be obvious that many interpretations and assumptions were necessary in completing the analyses. These were unavoidable because of differences in data sets and data gathering techniques. Such assumptions and interpretations need to be carefully questioned and may require some later revision. Nevertheless, small changes in assumptions and interpretations are not likely to alter the main outcomes of the analyses.

ACKNOWLEDGMENTS

This work was sponsored by the National Highway Traffic Safety Administration (NHTSA), Office of Crash Avoidance Research, under Contract No. DTNH22-91-C-07003 to Battelle, Inc., Columbus, Ohio. The author wishes to thank Louis Tijerina of Battelle, and Michael Goodman, Michael Perel, and Robert Clarke of NHTSA, Office of Crash Avoidance Research, for helpful suggestions.

REFERENCES

Anacapa Sciences (1976). Driver expectancy in locating automotive controls. Warrendale, PA: Society of Automotive Engineers (SAE), Report No. SP-407.

Bhise, V. D., Forbes, L. M., Farber, E. I. (1986). Driver behavioral data and considerations in evaluating in-vehicle controls and displays. Paper presented at the *Transportation Research Board, National Academy of Sciences, 65th Annual Meeting*, Washington, D.C. (January).

Dingus, T. A., Antin, J. F., Hulse, M. C., and Wierwille, W. W. (1989). Attentional demand requirements of an automobile moving-map navigation system. *Transportation Research A*, 23(A), 301-315.

Perel, M. (1976). Analyzing the role of driver/vehicle incompatibilities in accident causation using police reports. Report No. DOT-HS-801-858. Washington, D.C.: National Highway Traffic Safety Administration.

Rockwell, T. H. (1987). Spare visual capacity in driving revisited: New empirical results for an old idea. In A. G. Gale, et al. (Eds.) *Vision in Vehicles II — Proceedings of the Second International Conference on Vision in Vehicles*, Nottingham, UK, 14-17 September, 1987 (pp. 317-324). Amsterdam, Netherlands: Elsevier.

Wierwille, W. W., Tijerina, L., Kiger, S. M., Rockwell, T. H., Lauber, E., and Bittner, A. (1992). Task 4 Interim Report: Review of workload and related research and a theoretical basis for safety relevance of workload measures. Columbus, Ohio: Battelle, Inc. (Report on NHTSA Heavy Vehicle Driver Workload Assessment, September.)

Wierwille, W. W. (1993). Visual and manual demands of in-car controls and displays. In B. Peacock and W. Karwowski (Eds.) *Automotive Ergonomics*, London: Taylor and Francis Ltd., 299-320.

Wierwille, W. W. and Tijerina, L. (1993). An analysis of driving accident narratives as a means of determining problems caused by in-vehicle visual allocation and visual workload. Paper accepted for presentation at the *Fifth International Conference on Vision in Vehicles*, Glasgow, Scotland (September).

Woodson, W. E., Conover, D. W., Miller, G. E., and Selby, p. H. (1969). Instrument and control location, accessibility and identification. San Diego, CA: Man Factors, Inc., Report No. MF1-69-106. Prepared for National Highway Safety Bureau, U.S. Department of Transportation, Washington, D.C.

APPENDIX F

EXPERIMENTAL DESIGN STRATEGIES FOR DATA COLLECTION AND ANALYSIS

APPENDIX F. EXPERIMENTAL DESIGN STRATEGIES FOR DATA COLLECTION AND ANALYSIS

This appendix provides an introduction to data collection strategies and experimental design. These strategies are particularly useful for driver workload research where multiple factors must be included in an evaluation and orderly data collection is of great importance. Table F-1 provides definitions of terminology used in this appendix. Experimental design is discussed in many textbooks devoted to the subject (Keppel, 1991; Winer, 1971; Box, Hunter, and Hunter, 1978; Kirk, 1982).

FACTORIAL DESIGN APPROACHES

In discussing alternative data collection strategies and experimental designs, some specialized terminology will be used (see Table F-1). The most straightforward approach to experimental design is to use a factorial design. In this approach, all levels of all independent factors are crossed. For example, if there are 3 factors, each with two levels, there will be 2^3 or 8 combinations of the three factors in a full factorial experiment. The full factorial design is perhaps feasible with a relatively small number of factors (e.g., 2 to 4 factors). Full factorial experiments are appropriate when the evaluator requires that:

- Every main effect (i.e., effect of each factor averaged over all other factors) be estimated independently of every other factor, and
- all interactions be assessed.

It is important to note that for any experiment with n independent factors, there are 2^n effects that could be assessed. Thus for an experiment with 4 factors, there are 2^4 or 16 terms in a full regression model (4 main effects, 6 two-factor interactions, 4 three-factor interactions, 1 four-factor interaction, and the intercept). Generally speaking, higher-order (e.g., 3-factor and higher) interactions tend to be non-significant or account for a trivial proportion of the variability of the response measure (Box, Hunter, and Hunter, 1978). Furthermore, the interpretability of a higher-order interaction is usually extremely limited in supporting workload evaluations.

The selection of within-subjects experimental designs is generally beneficial from several standpoints. This type of design uses fewer subjects, as a rule, and reduces the variability within in each treatment level or treatment combination because subjects serve as their own controls. It is often the case that individual differences account for much larger proportions of variability in a measured response than any of equipment or roadway factors. However, repeated measures designs cannot be used if a) the number of treatment levels or treatment combinations is prohibitively larger, b) if there are potential practice effects (i.e., learning the experimental procedures, as apart from the in-vehicle technology), or c) if there are potential carry-over effects

Table 3-7. Experimental Design Terminology.

Experimental Design Term	**Definition**
Factor	An independent variable in the evaluation.
Factor Levels	The specific values of a factor included in the evaluation.
Factorial Design	An experimental design involving more than two or more factors, in which every combination of factor levels has been included. For example, if there are three factors (a, B, C) and each has two levels (a_1, a_2; b_1, b_2; c_1, c_2), there are 2^3 or 8 factorial combinations of the three factors: a_1, b_1, c_1 a_2, b_1, c_1 a_1, b_2, c_1 a_2, b_2, c_1 a_1, b_1, c_2 a_2, b_1, c_2 a_1, b_2, c_2 a_2, b_2, c_2
Treatment or Treatment Combination	a particular level (if only one factor is included) or combination of levels of factors (if two or more factors are included) in the evaluation.
Crossed Factors	Factors in which every level of one factor is combined with each level of another factor. In the example provided above, factors a, B, and C are crossed. Note that the number of treatments or treatment combinations can be calculated by multiplying the levels of each factor with the levels of the others. In the above example, each of three factors had the same number of levels and so there were $2 \times 2 \times 2 = 2^3 = 8$ treatment combinations.
Nested factor	a factor B is said to be nested within one level of factor a if each level of factor B appears with only one level of factor a.
Main effect	The impact of a factor (independent variable) on a dependent variable independent of or averaged over the impact of all other independent variables.

Experimental Design Term	**Definition**
Interaction	An interaction is present when the effects of one factor depend on the levels of another factor. Only crossed factors may interact. For this reason, nested factors cannot be assessed for interaction with factor or factors within which they are nested.
Dependent Variable	This is the measured response variable analyzed to assess the effects of the factors included in the study. Also called Measure of Performance (MOP), criterion measure, and response measure.
Confounding (Aliasing)	A confound or alias is an effect that cannot be distinguished from another effect. Thus, confounds or aliases are confusions or uncertainties about the source of some effect. When considered formally in an experimental design, confounds and aliases refer to main effects or interactions in the model that cannot be distinguished statistically from other main effects or interactions in the model. If an extraneous or nuisance variable is not controlled properly, there can be a confound between that nuisance variable and some independent variable of interest.
Replication	The observation of two or more experimental units (e.g., subjects) under identical treatment conditions to obtain an estimate of experimental error or error variation and permit a more precise estimate of treatment effects.
Experimental Error (Residual)	Variation in a dependent variable that is attributable to factors not relevant to the research hypotheses, i.e., by random factors.
Between-subjects Design (Completely Random zed Design)	An experimental design in which each subject experiences only one treatment combination in the evaluation.
Within-subjects Design (Repeated Measures Design)	An experimental design in which each subject experiences all treatment combinations in the evaluation.
Mixed Design	An experimental design in which some factors in the evaluation are between-subject factors and other factors are within-subjects factors.
Random Factor	Factor for which the treatment levels are a random sample from a larger population and inferences will be drawn about this population.

Experimental Design Term	**Definition**
Fixed Factor	Independent factor for which all treatment levels about which statistical inferences are to be drawn are included in the study.

such as learning about the device or the roadway, noting that a simulator study includes crash hazard events, and so on. Mixed designs are useful for data economy and are mandatory when, for example, subject variables are a formal part of the study. For example, gender or age are between-subjects variables that might be included in a mixed design. Perhaps all subjects would be measured repeatedly under different driving conditions, in which case driving condition factors are within-subjects variables. Care must be taken in the choice of these experimental design alternatives and the choice of which approach to adopt depends on the availability of test subjects and the characteristics of the specific evaluation itself (Williges, 1984). Appendix F provides further descriptions of alternative experimental designs.

ECONOMIC DATA COLLECTION DESIGNS

Oftentimes, there will be constraints to the use of full factorial designs in device evaluations (Williges, 198 1). Some factors cannot be crossed in the real world. It-may not be possible to collect data from the entire factorial design at one time. There may be more factors than can be reasonably included in a full factorial experimental design. For these reasons, economical data collection approaches are a necessity and several different classes of approach will be presented here.

HIERARCHICAL DESIGNS

Hierarchical designs are suitable when it is appropriate to nest one or more factors into other factors. That is, levels of one factor appear only at one level of another factor in hierarchical fashion. The hierarchical design results in a smaller number of treatment combinations when compared to a complete factorial design. Because nested variables are not crossed with the factors they are nested within, it is not possible to assess any interactions that may be present. Thus, care must be taken in planning hierarchical designs when used for purposes of data collection economy (Williges, 1984).

CONFOUNDED FACTORIAL DESIGNS

Sometimes it is not possible to collect all the workload assessment data needed in a single data collection session. Thus, the evaluator must collect the data in stages or in “blocks”. There are procedures (e.g., Kirk, 1983) that can be used to define blocks and systematically confound or alias block differences with higher-order interactions. This allows for the assessment of all main effects and lower-order interactions at the expense of the higher order interaction used in developing the treatment combinations for the blocks. Tijerina et al. (in press) used a confounded factorial design to assess workload effects of lighting (night vs. day), traffic density (low vs. high) and road type (divided vs. undivided). It was not feasible to collect workload measures on a single driver in all 2^3 or 8 treatment combinations. Instead, it was only feasible to collect workload measures on four of the 8 treatment combinations. Figure F-l shows how the design was blocked across two groups of subjects such that the main effects and two-way interactions would not be confounded with possible differences across the two groups of subjects. Only the three-way interaction is completely confounded with the subject group effects. Since three-way and higher order interactions tend to be either not statistically significant or account for only a trivial proportion of the variability in a measure of performance, this was judged to be a worthwhile tradeoff.

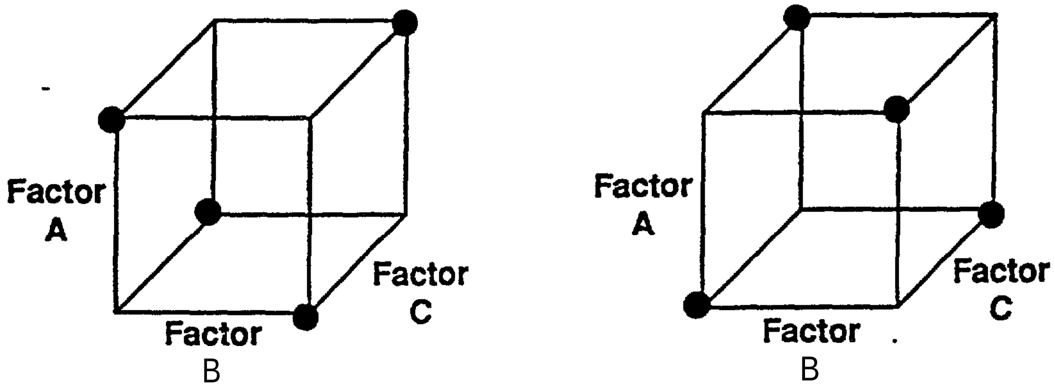
FRACTIONAL FACTORIAL DESIGNS

When it is impossible to collect data on all combinations of all factors or variables of interest from a higher-order design, a fractional factorial design may be used. As its name implies, a fractional factorial experimental design employs only a fraction of a full factorial design. That is, this experimental design approach and data collection procedure uses only a carefully chosen fraction of all possible combinations of factors to estimate the effects of those variables and interactions. For example, if there are eight (8) factors that might be included in an experiment or evaluation and each factor has only two levels, there are $2^8 = 256$ possible experimental treatments. On the other hand, a one-quarter fractional factorial design (2^{8-2}) requires only 64 experimental treatments to be run in an experiment. The data collection economy is bought with aliasing, i.e., confounding main effects and two-factor interactions with higher-order interactions. Higher-order interactions are assumed to be non-existent or trivial, So, if, for example, a two-factor interaction is statistically significant, it is attributed to that pair of factors rather than higher-order interactions aliased with it. The reasonableness of this approach comes from the fact that main effects tend to be larger than two-factor interactions, two-factor interactions tend to be larger than three-factor interactions, and so on (Box, Hunter, and Hunter, 1978). In addition to the general result that higher-order interactions contribute little or no additional explanatory power, the ability to comprehend and explain such interactions in a parsimonious way becomes impossible.

For most practical purposes, there is little advantage to evaluating the impact of factors alone and in two-factor interactions. This leads to the concept of design resolution. Design resolution refers to the precision of the estimated effects from the experiment and the types of aliases that might exist. In general:

- A design of Resolution R = III does not confound single-factor effects (called main effects) with one another but does confound main effects with two-factor interactions.

- A design of Resolution R = IV does not confound single-factor effects and two-factor interactions but does confound two-factor interactions with other two-factor interactions.



Conditions	Conditions
A_1, B_2, C_1	A_1, B_1, C_1
A_1, B_1, C_2	A_1, B_2, C_2
A_2, B_1, C_1	A_2, B_2, C_1
A_2, B_2, C_2	A_2, B_1, C_2

Figure F-l. Confounded Factorial Design Strategy for Assessing Driving Condition Factors and Their Effects on Driver Workload.

As mentioned previously, higher-order interactions (i.e., three-factor interactions and higher) are often not significant or account for so little of the variation in the response that they are trivial. For this reason, a significant main effect aliased with a higher-order interaction is attributed to the main effect alone. Similarly, a two-factor interaction that aliases with a three-factor interaction is attributed to the two-factor interaction alone. Since human performance and behavioral data sometimes yield two-factor interactions of interest, fractional factorial designs of resolution V or higher are recommended.

Originally, the fractional factorial approach to experimental design was developed to allow for sequential experimentation. That is, research or evaluations would be done in stages and ambiguities that arise in the first stage of the research could be investigated in subsequent data collection stages that dis-ambiguate the first stage results. It is often the case that product evaluations do not allow for sequential investigations, Thus, a resolution V design is recommended if only a single investigation is to be carried out. Finally, it should be noted that fractional factorial designs typically involve all factors at the same number of levels, most commonly 2 levels. Thus, the factors may be continuous variables with levels chosen for “high” and “low” or they may be dichotomous variables (e.g., male or female drivers).

The attachment to this appendix contains examples for Resolution III, IV, V, VI, and VII designs to study up to eight factors and indicates the number of unique runs required. The negative (-) sign and positive (+) signs represent the two levels of each factor. Assignment of factors and factor levels to codes is arbitrary. A software program has been developed, entitled the Automated Experimental Design (AED) Assistant, that will automatically generate fractional factorial design assignments for 2 level designs with up to 20 factors and a maximum of 256 treatment combinations, for 3-level designs with up to **12** factors and maximum of 243 treatment combinations, and for 5-level designs with up to 8 factors and a maximum of 125 treatment combinations (System Development Corporation, 1986).

CENTRAL COMPOSITE DESIGN APPROACHES

The fractional factorial design (along with confounded factorial designs and single-observation factorial designs) supports economical data collection for hypothesis testing, i.e., to test whether a driver-performance measure reliably varies with some workload factor. It is well suited to the analysis of categorical variables (at two levels, usually, though 3-level and 5-level fractional factorial designs are also possible). In other situations, however, the evaluator may seek to determine a quantitative relationship between driver performance and several quantitative independent factors, e.g., in-cab device parameters. Such a functional relationship is useful in that it allows for comparative predictions of various alternative system configurations, during, say, product development (Williges, 198 1). The empirical model form that has been recommended for human factors use is a second-order polynomial, i.e.,

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_1 + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \beta_{ij} x_i x_j + e$$

where y is a dependent measure (i.e., response variable),
Bi terms are the estimated model coefficients for pure linear effects,
Bii terms are the estimated model coefficients for pure quadratic effects,
Bij terms are the estimated model coefficients for two-factor interactions,
x_i is the main effect of the ith factor,
$x_i x_j$ is the interaction between the ith and jth variables,
x_i^2 is the pure quadratic term for the ith factor, and
e is the error term or difference between actual and estimated response values.

Least squares regression estimates of the beta parameters specified in the second-order polynomial response surface is given by the expression.

$$I3 = (X'X)^{-1}X'Y$$

where	X	is an n x p design matrix for a particular fractional factorial design (see Appendix N), augmented so that column 1 is a column of 1s so that the intercept may be estimated;
	X’	is the p x n transpose of the X design matrix (augmented);
	$(X' X)^{-1}$	is the p x p inverse of the sum of squares and cross products matrix; and
	Y	is the n x 1 column vector of responses for each condition run; and
	X’Y	is the p x 1 column vector that is the matrix product of the two matrices involved.

As with the fractional factorial design, the central composite design data is analyzed with regression methods and the ANOVA to assess the statistical signficance of each of the terms of the fitted regression equation. The selection of levels of each variable to economically collect data to build the empirical second-order polynomial equation has been discussed by Williges (1980) and Williges and Williges (1992).

In order to build such a model, data are needed to solve the least squares regression equations. Box and Wilson (195 1, cited in Williges, 1981) developed an experimental methodology that determines optimal combinations of various quantitative factors to define the response surface. The design approach is to use a composite three separate parts: a) 2^k factorial or 2^{k-1} fractional factorial design portion; b) 2k additional points that outline a star pattern in the factor space; and c) 1 center point from which the entire composite of points radiates. For this reason, this is referred to as a central composite design. Generally, any second order, central composite design can be specified with a total of T points:

$$T=F+2k+1$$

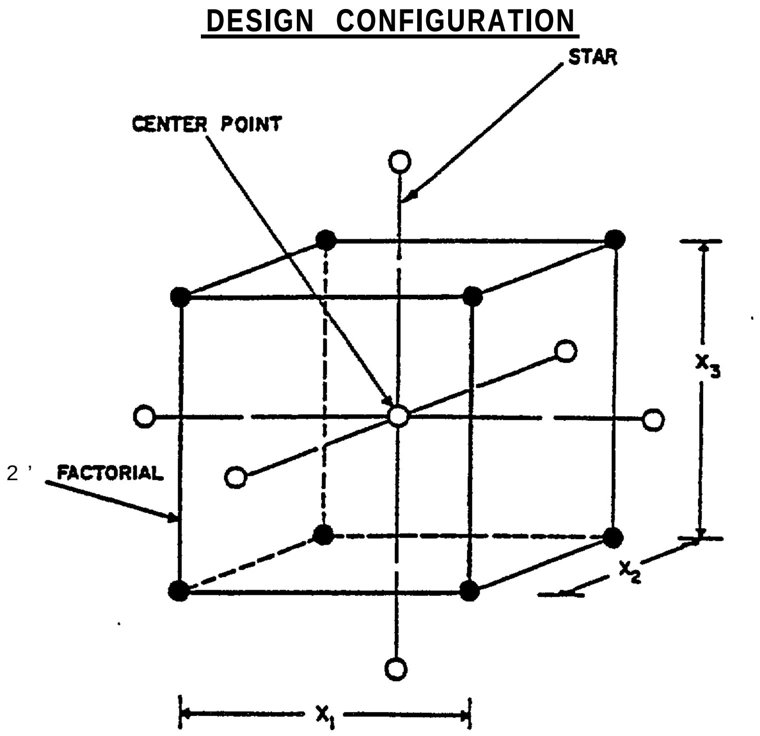
where $F =$ $2k$ (or 2^{k-p})
$k =$ the number of factors under investigation.
$p =$ a number that represents the degree of fractionation in the fractional factorial design (1 = one-half fraction; 2 = one-fourth fraction; 3 = one-eighth fraction; 4= one-sixteenth fraction)

For example, consider a three-factor evaluation. If one substitutes $k = 3$ in the equation above, the value of $T = 15$, i.e., fifteen unique combinations are required to specify the second-order polynomial response surface. If one used a complete factorial experimental design, then with each factor at 5 levels (the number of levels needed to map out a second-order polynomial), there would be 5^3 or 125 combinations. The economy of approach is apparent.

Like the fractional factorial approach, this economy comes at a price. Care must be taken such that the 2^k combinations of factors, if substituted by a fractional factorial design portion due to the large number of factors of interest, be chosen so that all first- and second-order components are present and are not aliases of each other so that a complete second-order response surface can be generated (Williges, 1981). Clearly, all factors must be quantitative, else it makes little sense to discuss linear and quadratic components. While minimally only 3 levels of a factor are needed to outline a curve, 5 levels, appropriately chosen, will provide sufficient detail to develop an entire surface.

Figure 3-2 shows a hypothetical example of a three-factor central composite design to evaluate automobile driving performance (y) as a function of wind gust characteristics (Williges, 198 1). Only 15 unique treatment combinations are required, as indicated in the coding scheme at the bottom of the figure. Replication is needed to allow for the analysis of variance mean square error term to be defined. This may involve running two (or more) subjects in each of the treatment conditions or perhaps having a single group of subjects in a repeated measures format drive in all treatment conditions and thus provide replication over the entire design surface. See Williges (198 1) for more details about replication decisions.

The value of a remains to be specified. While there are alternative ways to define the a level (see Williges, 198 I), one simple way is such that the first and second-order beta weights are orthogonal (which facilitates least squares regression and analysis of variance:



Treatment Combination	X_1 Wind Gust Frequency	X_2 Wind Gust Velocity	X_3 Wind Gust Direction
1	+1	+1	+1
2	+1	-1	+1
3	+1	+1	-1
4	+1	-1	-1
5	-1	+1	+1
6	-1	-1	+1
7	-1	+1	-1
8	-1	-1	-1
9	+a	0	0
10	-a	0	0
11	0	+a	0
12	0	-a	0
13	0	0	+a
14	0	0	-a
15	0	0	0

Figure F-2. Central Composite Design Illustration. Source: Williges (1980).

$$\alpha = (\frac{QF}{4})^{1/4}$$

where Q = $[F + 2k + C)^{1/2} - F^{1/2}]^2$
C = the total number of center points (1 if equal replication is used)
F 2^k (or 2^{k-p})
k: the number of factors under investigation.
P= a number that represents the degree of fractionation in the fractional factorial design (1 = one-half fraction; 2 = one-fourth fraction; 3 = one-eighth traction; 4= one-sixteenth fraction)

For the three factor design, Q = [(8 + 2(3) + 1)” - $8^{1/2}$]^2 = 1.092. This implies that a = $[(1.092)(8)/4]^{1/4} = 1.216$. The application of the coding scheme is then applied by assigning the a codes to the lower and upper limits of the factor range of interest, the code 0 is assigned to the midpoint of the range, and the -1 and +1 codes are assigned to factor values determined through linear interpolation. An example is provided in Table F-2. The design factors of interest are in-cab visual display luminance (x_1) (selectable over a range from 14 cd/m^2 to 140 cd/m^2), visual display contrast ratio (x_2) (selectable over a range from 2: 1 to 30:1), and symbol size (x_3) (selectable over a range from 10 to 28 arc-min). The response (y) is driver visual allocation time (number of glances x mean glance duration). The levels of each factor are included in Table F-2.

The central composite design, then, gets its name from the fact that it is a composite of a 2^k (or 2^{k-p}) design, augmented by a star pattern of data collection points that radiate from a center point. the Automated Experimental Design (AED) Assistant software (System Development Corporation, 1986) will also automatically generate central composite designs.

REPLICATION AND SINGLE-OBSERVATION FACTORIAL EXPERIMENTAL DESIGNS

In each of the preceding discussions, it was assumed that there were replications in each treatment combination of the experimental design in the form of two or more subjects in each treatment combination or having each subject perform in each treatment combination two or more times. In general, the greater the number of replicates in each treatment combination, the greater the precision in estimating error variance or experimental error. The price for this precision, however, is increased costs in data collection.

One simple approach to reduce the number of data observations to be collected is to eliminate replication altogether. This might be accomplished by assigning only one subject per treatment combination in a between-subjects experimental design. Alternatively, only one subject may be observed only once per treatment combination in a single-subject study. Many other schemes are also possible. Statistical tests of main effects and lower-order interactions are carried out by

Table F-2. Levels of three independent factors in Central Composite Experimental Design. (See text for explanation).

	Central Composite Design Codes and Associated Regressor Variable Values				
Regressor Variable	$-\alpha = -1.216$	-1	0	+1	$\alpha = 2.16$
x_1: Display Luminance	14 cd/m^2	23 cd/m^2	63 cd/m^2	127 cd/m^2	140 cd/m^2
x_2 : Contrast Ratio	2	4	14	27	30
x_3 : Character Size	10 arc-min	12 arc-mm	19 arc-min	26 arc-min	**28 arc-min**

Notes:

- See previous discussion for derivation of a.
- -a value is set to minimum of factor (regressor) range of interest
- +a value is set to maximum of factor (regressor) range of interest
- 0 value is set to mid-point of range from maximum to minimum of regressor range.
- -1 value is set to l/l .216 = .822 of the range between 0 and -a below the midpoint value
- +l value is set to l/l .216 = -822 of the range between 0 and -a above the midpoint value
- Note that all regressor values have been rounded.

pooling higher order interactions into a pooled-error term. This procedure assumes that the higher order terms are either non-significant or account for only a trivial proportion of the variability in the dependent measure. It is still possible that full factorial designs are infeasible because of the great number of treatment combinations. Under such conditions, the other economical data collection approaches should be considered.

SUMMARY

Efficient data collection supports efficient data analysis. For this reason, the data collection strategies presented here are beneficial to workload assessment and research. It is usually the case that the actual data collection and data capture vary somewhat from what was planned, especially in on-the-road evaluations or studies. The consultation of an experienced statistician or data analyst will be worthwhile to deal with the details of a statistical evaluation suitable for the data in hand.

REFERENCES

Box, G. E. P., Hunter, W. G., and Hunter, J. *S.* (1978). *Statistics for experimenters: An introduction to design, data analysis, and model building. New* York: Wiley.

Box, G. E. P., & Wilson, *K.* B. (1951). *On the experimental attainment of optimum conditions.* Journal of the Royal Statistical Society, Series B (Methodological), 13, l-45.
Keppel, G. (1991). *Design and analysis: A researcher 's handbook* Englewood Cliffs, NJ: Prentice Hall.

Kirk, R. E. *(1982). Experimental design: Procedures for the behavioral sciences* (Second edition), Belmont, CA: Brooks/Cole.

System Development Corporation (1986, December). *AED Volume II: The User-Assisted Automated Experimental (Test) design Program; Version* IV, Wright Patterson AFB, OH: Crew System Ergonomics Information Analysis Center (CSERIAS).

Tijerina, L., Kiger, S. M., Rockwell, T. H., & Tornow, C. E. (1995, September). *Final report: Workload assessment of in-cab text message system and cellular phone use by heavy vehicle drivers on the road* (Contract No. DTHN22-91-C-07003). Columbus, OH: Battelle.

Williges, R. C. (198 1). Development and use of research methodologies for complex system/simulation experimentation. In M. Moraal and *K. Kraiss* (Eds.), *Manned system design* (pp. 59-87). New York: Plenum.

Williges, R. C. (1984). Evaluating human-computer software interfaces. *Proceedings* **of** *the 1984 International Conference on Occupational Ergonomics,* Toronto, Canada, May, 1984.

Winer, B. J. (1971). *Statistical principles in experimental design* (2nd edition). New York: McGraw-Hill.

ATTACHMENT

EXAMPLE 2^K FRACTIONAL FACTORIAL DESIGNS

Design Matrix, 2^{3-1}, Resolution III Fractional Fractorial Design

Subject	Factor Level		
	1	2	3
1	-	-	-
2	+	-	+
3	-	+	+
4	+	+	-

Alternate Design Matrix:

Subject	Factor Level		
	1	2	3
1	-	-	+
2	+	-	-
3	-	+	-
4	+	+	+

Design Matrix, 2^{4-1}, Resolution IV Fractional Fractorial Design

Subject	Factor Level			
	1	2	3	4
1	-	-	-	-
2	+	-	-	+
3	-	+	-	+
4	+	+	-	-
5	-	-	+	+
6	+	-	+	-
7	-	+	+	-
8	+	+	+	+

Design Matrix, 2^{5-1}, Resolution V Fractional Fractorial Design

Subject	Factor Level				
	1	2	3	4	5
1	-	-	-	-	+
2	+	-	-	-	-
3	-	+	-	-	-
4	+	+	-	-	+
5	-	-	+	-	-
6	+	-	+	-	+
7	-	+	+	-	+
8	+	+	+	-	-
9	-	-	-	+	-
10	+	-	-	+	+
11	-	+	-	+	+
12	+	+	-	+	-
13	-	-	+	+	+
14	+	-	+	+	-
15	-	+	+	+	-
16	+	+	+	+	+

Design Matrix, 2^{6-1}, Resolution VI Fractional Fractorial Design

Subject	Factor Level					
	1	2	3	4	5	6
1	-	-	-	-	-	-
2	+	-	-	-	-	+
3	-	+	-	-	-	+
4	+	+	-	-	-	-
5	-	-	+	-	-	+
6	+	-	+	-	-	-
7	-	+	+	-	-	-
8	+	+	+	-	-	+
9	-	-	-	+	-	+
10	+	-	-	+	-	-
11	-	+	-	+	-	-
12	+	+	-	+	-	+
13	-	-	+	+	-	-
14	+	-	+	+	-	+
15	-	+	+	+	-	+
16	+	+	+	+	-	-

Design Matrix, 2^{6-1}, Resolution VI Fractional Factorial Design (Continued)

Subject	Factor Level					
	1	2	3	4	5	6
17	-	-	-	-	+	+
18	+	-	-	-	+	-
19	-	+	-	-	+	-
20	+	+	-	-	+	+
21	-	-	+	-	+	-
22	+	-	+	-	+	+
23	-	+	+	-	+	+
24	+	+	+	-	+	-
25	-	-	-	+	+	-
26	+	-	-	+	+	+
27	-	+	-	+	+	+
28	+	+	-	+	+	-
29	-	-	+	+	+	+
30	+	-	+	+	+	-
31	-	+	+	+	+	-
32	+	+	+	+	+	+

Design Matrix, 2^{7-1}, Resolution VII Fractional Factorial Design

Subject	Factor Level						
	1	2	3	4	5	6	7
1	-	-	-	-	-	-	+
2	+	-	-	-	-	-	-
3	-	+	-	-	-	-	-
4	+	+	-	-	-	-	+
5	-	-	+	-	-	-	-
6	+	-	+	-	-	-	+
7	-	+	+	-	-	-	+
8	+	+	+	-	-	-	-
9	-	-	-	+	-	-	-
10	+	-	-	+	-	-	+
11	-	+	-	+	-	-	+
12	+	+	-	+	-	-	-
13	-	-	+	+	-	-	+
14	+	-	+	+	-	-	-
15	-	+	+	+	-	-	-
16	+	+	+	+	-	-	+
17	-	-	-	-	+	-	-
18	+	-	-	-	+	-	+
19	-	+	-	-	+	-	+
20	+	+	-	-	+	-	-
21	-	-	+	-	+	-	+
22	+	-	+	-	+	-	-
23	-	+	+	-	+	-	-
24	+	+	+	-	+	-	+
25	-	-	-	+	+	-	+

Subject	Factor Level						
	1	2	3	4	5	6	7
26	+	-	-	+	+	-	-
27	-	+	-	+	+	-	-
28	+	+	-	+	+	-	+
29	-	-	+	+	+	-	-
30	+	-	+	+	+	-	+
31	-	+	+	+	+	-	+
32	+	+	+	+	+	-	-
33	-	-	-	-	-	+	-
34	+	-	-	-	-	+	+
35	-	+	-	-	-	+	+
36	+	+	-	-	-	+	-
37	-	-	+	-	-	+	+
38	+	-	+	-	-	+	-
39	-	+	+	-	-	+	-
40	+	+	+	-	-	+	+
41	-	-	-	+	-	+	+
42	+	-	-	+	-	+	-
43	-	+	-	+	-	+	-
44	+	+	-	+	-	+	+
45	-	-	+	+	-	+	-
46	+	-	+	+	-	+	+
47	-	+	+	+	-	+	+
48	+	+	+	+	-	+	-
49	-	-	-	-	+	+	+

Subject	Factor Level						
	1	2	3	4	5	6	7
50	+	-	-	-	+	+	-
51	-	+	-	-	+	+	-
52	+	+	-	-	+	+	+
53	-	-	+	-	+	+	-
54	+	-	+	-	+	+	+
55	-	+	+	-	+	+	+
56	+	+	+	-	+	+	-
57	-	-	-	+	+	+	-
58	+	-	-	+	+	+	+
59	-	+	-	+	+	+	+
60	+	+	-	+	+	+	-
61	-	-	+	+	+	+	+
62	+	-	+	+	+	+	-
63	-	+	+	+	+	+	-
64	+	+	+	+	+	+	+

Design Matrix, 2^{8-2}, Resolution V Fractional Fractorial Design

Subject	Factor Level							
	1	2	3	4	5	6	7	8
1	-	-	-	-	-	-	+	+
2	+	-	-	-	-	-	-	-
3	-	+	-	-	-	-	-	-
4	+	+	-	-	-	-	+	+
5	-	-	+	-	-	-	-	+
6	+	-	+	-	-	-	+	-
7	-	+	+	-	-	-	+	-
8	+	+	+	-	-	-	-	+
9	-	-	-	+	-	-	-	+
10	+	-	-	+	-	-	+	-
11	-	+	-	+	-	-	+	-
12	+	+	-	+	-	-	-	+
13	-	-	+	+	-	-	+	+
14	+	-	+	+	-	-	-	-
15	-	+	+	+	-	-	-	-
16	+	+	+	+	-	-	+	+
17	-	-	-	-	+	-	+	-
18	+	-	-	-	+	-	-	+
19	-	+	-	-	+	-	-	+
20	+	+	-	-	+	-	+	-
21	-	-	+	-	+	-	-	-
22	+	-	+	-	+	-	+	+
23	-	+	+	-	+	-	+	+
24	+	+	+	-	+	-	-	-

Design Matrix, 2^{8-2}, Resolution V Fractional Fractorial Design (Continued)

Subject	Factor Level							
	1	2	3	4	5	6	7	8
25	-	-	-	+	+	-	-	-
26	+	-	-	+	+	-	+	+
27	-	+	-	+	+	-	+	+
28	+	+	-	+	+	-	-	-
29	-	-	+	+	+	-	+	-
30	+	-	+	+	+	-	-	+
31	-	+	+	+	+	-	-	+
32	+	+	+	+	+	-	+	-
33	-	-	-	-	-	+	+	-
34	+	-	-	-	-	+	-	+
35	-	+	-	-	-	+	-	+
36	+	+	-	-	-	+	+	-
37	-	-	+	-	-	+	-	-
38	+	-	+	-	-	+	+	+
39	-	+	+	-	-	+	+	+
40	+	+	+	-	-	+	-	-
41	-	-	-	+	-	+	-	-
42	+	-	-	+	-	+	+	+
43	-	+	-	+	-	+	+	+
44	+	+	-	+	-	+	-	-
45	-	-	+	+	-	+	+	-
46	+	-	+	+	-	+	-	+
47	-	+	+	+	-	+	-	+
48	+	+	+	+	-	+	+	-
49	-	-	-	-	+	+	+	+

Design Matrix, 2^{8-2}, Resolution V Fractional Fractorial Design (Continued)

Subject	Factor Level							
	1	2	3	4	5	6	7	8
50	+	-	-	-	+	+	-	-
51	-	+	-	-	+	+	-	-
52	+	+	-	-	+	+	+	+
53	-	-	+	-	+	+	-	+
54	+	-	+	-	+	+	+	-
55	-	+	+	-	+	+	+	-
56	+	+	+	-	+	+	-	+
57	-	-	-	+	+	+	-	+
58	+	-	-	+	+	+	+	-
59	-	+	-	+	+	+	+	-
60	+	+	-	+	+	+	-	+
61	-	-	+	+	+	+	+	+
62	+	-	+	+	+	+	-	-
63	-	+	+	+	+	+	-	-
64	+	+	+	+	+	+	+	+

APPENDIX G

EXAMPLE SUBJECT INFORMED CONSENT FORM

SUBJECT CONSENT FORM

Title of Study: Heavy Vehicle Driver Workload Assessment Project

Study Description: Many high technology in-cab devices are being proposed for use in heavy trucks (e.g., route guidance systems, trip recorders, text displays, communications systems, etc.). These devices sometimes introduce additional tasks which might compete with the driver's primary job of safely controlling the vehicle at all times. Battelle is conducting a research project for the National Highway Traffic Safety Administration (NHTSA) to measure the effects on drivers of introducing high technology in-cab devices. We believe that this work will contribute significantly toward enhancing safety and promoting a driver-centered approach to the development of high technology in-cab devices.

As part of our work, Battelle (through our subcontractor R&R Research, Inc.) must collect data from drivers under various normal driving conditions. The purpose of this data collection is to better understand the various driving tasks drivers must perform today, the driving conditions under which they work, and the driver behaviors, performance, and attitudes which may reflect driver workload.

As a voluntary participant, you will drive a US Government tractor-semitrailer through public roadways selected for the study. During testing, a ride-along observer will be in the vehicle with you on your assigned route. This observer will operate measurement equipment, give instructions to you about where to drive, and ask you to operate equipment commonly found in modern heavy vehicles. On-the-road data collection will involve observation of driving behaviors and tasks performed, video taping of the road scene and the driver's visual scanning patterns, and various measures of driving performance such as lane keeping, speed control, headway maintenance, and so forth. The ride-along observer will ask you to visually scan the west coast mirrors and selected gauges on the instrument panel or to manipulate knobs or switches when driving permits. The ride-along observer will ask you to answer questions about heavy vehicle driving.

You, the ~~driver, are in control~~ and will be the final judge o[redacted]spond to any request. Do not blindly follow any request. Follow our requests and answer our questions only when it is safe and convenient to do so.

Risks: While driving for this study, you will be subject to all risks normally present during heavy truck driving. There are no known physical or psychological risks associated with participation in this study beyond those normally found in heavy truck driving. However, you must be aware that accidents can happen any time while driving.

You remain legally liable for your actions during this testing. You will not intentionally be asked to drive illegally. Should an action requested of you by the ride-along observer seem illegal, YOU are not to do it. Should you receive a speeding ticket or some other legal penalty for your driving during this testing, you understand that neither Battelle, R&R Research, Inc, nor the U.S. Government will compensate you for any fines or otherwise assist you in resolving legal problems arising out of any illegal action.

Benefits: The results of this study will provide valuable guidance for the development of an evaluation method to determine the safety of high technology in cab devices offered for use in heavy trucks.

By participation in this study, you will get some exposure to transportation research methods and will be lending your expertise and experience to support highway safety research regarding future use of in-cab devices.

Confidentiality: We are gathering information on heavy vehicle driving. We are not testing you. If you agree to participate in this study, your name will not be released to anyone other than the principal investigator. Individual performance will not have the subject's name associated with it in any interim or final reports. This confidentiality will be maintained.

Principal Investigator: If you have any questions or comments in relation to this study please contact the following person:

Louis Tijerina, Ph.D.
Battelle
505 King Ave.
Columbus, Ohio 43201
Ph: (614) 424-5406

Right to withdraw: YOU HAVE THE OPTION OF WITHDRAWING AT ANY TIME DURING THE COURSE OF THE STUDY WITHOUT PENALTY.

Compensation You will be paid a sum of $XXX.XX to participate in this study for an estimated one full day of your time. You are entitled to this pay even if you elect to withdraw at any time during the course of the study.

Cautions: As mentioned earlier, there are no known risks associated with participation in this study beyond those normally found in heavy truck driving. You will be the final judge of when or whether to respond to a question or request by the ride-along observer. You will be informed of new information which becomes available which might reasonably be expected to affect your willingness to continue to participate in the study.

Approximately six (6) to eight (8) drivers are expected to participate in this study.

It is not anticipated that you will be informed of the results of this study.

Disposition of Informed consent: The Principal Investigator will retain a copy of this Informed Consent Form. A copy of this form will also be provided to you upon completion of participation in this study.

INFORMED CONSENT:

I, ____________________________, UNDERSTAND THE TERMS OF THIS AGREEMENT
(Print your name)
AND CONSENT TO PARTICIPATE IN THIS STUDY.

____
SIGNATURE
DATE:

I, ______________________________, ACKNOWLEDGE RECEIPT OF $ XXX.XX FOR PARTICIPATING
(Print your name)
IN THIS STUDY.

____
SIGNATURE

DATE:

‘U.S. Government Printing Office: 1996 - 420-608