Final Report Supplement

Heavy Vehicle Driver Workload Assessment

Task 5: Workload Assessment Protocol

NOTE TO READER:

THIS IS A LARGE DOCUMENT

Due to its large size, this document has been segmented into multiple files. All files separate from this main document file are accessible from links (blue type) in the table of contents or the body of the document.

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.	3. Recipient's Catalog No.
	DOT HS 808 467 (5)	

4. Title and Subtitle Final Report Supplement --Task 5: Heavy Vehicle Driver Workload Assessment Protocol	5. Report Date June, 199.5
	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. DTNH22-9 1-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	18. Distribution Statement
human factors, heavy vehicle, driver workload measures, safety, assessment protocol	

19. Security Classif. (of this report)	20. Security Classif. (of this Page)	21. No. of Pages	22. Price
		184	

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

TABLE OF CONTENTS

Page

1.0 PERSPECTIVE . . . . , ... l-l

1.1 BACKGROUND.. . . . . l-l
1.2 PURPOSE l-2
1.3 A SIMPLE MODEL OF DRIVING l-3
1.4 RATIONALE FOR THE OMISSION OF ALTERNATIVE MEASURES OF WORKLOAD l-8
1.5 VALIDITY OF WORKLOAD MEASURES FOR PREDICTING SAFETY l-10
1.6 A SIMPLE THEORY OF CRASHES l-13
1.7 SCIENTIFIC BASES FOR THE SAFETY RELEVANCE OF WORKLOAD MEASURES 1-14

1.7.1 Visual Allocation 1-15
1.7.2 Lanekeeping 1-18
1.7.3 Speed Measures 1-19
1.7.4 Time Headway 1-19

1.8 ORGANIZATION OF THE PROTOCOL DOCUMENT . . . . l-l 9

2.0 WORKLOAD ASSESSMENT APPROACH DEVELOPMENT 2- 1

2.1 DETERMINE THE PURPOSE OF THE WORKLOAD EVALUATION 2-l

2.1.1 State the Objectives of the Workload Evaluation 2-l

2.2 DEFINE THE SYSTEM TO BE EVALUATED AND REFINE THE ASSESSMENT PROBLEM 2-3

2.2.1 Describe the System or Systems to Be Assessed 2-3
2.2.2 Identify Comparison Systems. Modes. or Functions to Be Assessed . . 2-4
2.2.3 Define the Situational Characteristics That Are Relevant to the Workload Evaluation 2-4
2.2.4 Define the Relevant Driver Ponulation 2-4
2.2.5 Define the Driver's Tasks to Be Assessed in the Workload Assessment 2-4

2.3 REVIEW RELEVANT LITERATURE . . . . , . 2-5

TABLE OF CONTENTS (CONTINUED)

Page

2.4 DETERMINE AVAILABLE AND REQUIRED RESOURCES 2-5

2.4.1 Logistical Resources . . . . , . . . 2-5
2.4.2 Establish the Workload Evaluation Team 2-6

3.0 WORKLOAD ASSESSMENT DETAILED EVALUATION PLAN 3-1

3.1 DEVELOP WORKLOAD ASSESSMENT TEST HYPOTHESES, MEASURES OF EFFECTIVENESS, MEASURES OF PERFORMANCE, AND DATA SOURCES 3-l
3.2 DEFINE THE VARIABLES TO BE CONSIDERED IN THE WORKLOAD EVALUATION 3-l

3.2.1 Select Indenendent Variables 3-6
3.2.2 Select Dependent Measures (Measures of Performance) 3-6
3.2.3 Identify and Determine Controls for Extraneous Variables 3-l 1

3.3 DESIGN THE EXPERIMENT/EVALUATION DATA COLLECTION STRATEGY 3-11
3.4 DEVELOP DATA COLLECTION HARDWARE/SOFTWARE INSTRUMENTATION NEEDS 3-12
3.5 DEFINE PROCEDURES TO BE FOLLOWED 3-12

3.5.1 Equipment Status and Startup Procedures 3-12
3.5.2 Test Participant Intake and Training Procedures 3-13
3.5.3 On-site Evaluator/Observer Procedures 3- 13
3.5.4 Data Evaluation and Management Procedures 3-13
3.5.5 Emergency Procedures 3-13

3.6 SELECT AND RECRUIT TEST PARTICIPANTS 3-14

3.6.1 Test Participants 3-14
3.6.2 Test Participant Recruitment 3-l 5
3.6.4 Test Participant Orientation and Training 3-15

3.7 DEVELOP WORKLOAD EVALUATION SCHEDULE 3-16

3.7.1 Develop Initial Schedule 3-16
3.7.2 Establish a Contingency Schedule 3-16

TABLE OF CONTENTS (CONTINUED)

Page

4.0 WORKLOAD ASSESSMENT TEST EXECUTION, ANALYSIS, AND REPORTING 4-l

4.1 PRE-PILOT TESTING 4-l
4.2 PILOT TESTING 4-2
4.3 FORMAL TESTING 4-2
4.4 PREPARE DATA FOR ANALYSIS 4-2

4.4.1 Reduce Data 4-2
4.4.2 Verify Reduced Data 4-3
4.4.3 Identify and Manage Missing Data 4-3

4.5 CONDUCT DATA ANALYSIS 4-4

4.5.1 Examine Reduced Data 4-4
4.5.2 Determine Appropriate Analysis Techniques 4-6
4.5.3 Apply Analysis Techniques 4-6

4.6 REPORT RESULTS 4-6

5.0 EPILOGUE 5-l

5.1 A STREAMLINED WORKLOAD ASSESSMENT PROTOCOL 5-l

5.1 .l Step 1. Analyze the Functions of the Device 5-l
5.1.2 Step 2. Apply an Ergonomics Checklist 5-l
5.1.3 Step 3. Conduct a Desktop Evaluation with a Video Game as the Primary "Driving" Task 5-2
5.1.4 Step 4. Carry Out a Check-Ride or Simplified On-The-Road Test . . . 5-3

TABLE OF CONTENTS (CONTINUED)

Page

LIST OF TABLES

Table l-l. 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) l-9

Table 3-l. TravTek Safety Evaluation Study Definition 3-2

Table 3-2. Does Safety-Relevant Driver-Vehicle Performance with In-Cab Device Use Significantly Differ from a Comparison Condition? 3-3

Table 3-3. Do Driver Behaviors with an In-Cab Device Differ Significantly from One or More Comparison Conditions? 3-4

Table 3-4. What are Driver Attitudes About the In-Cab Device? 3-6

Table 3-5. Independent Variables that may be Manipulated in a Driver Workload Evaluation 3-8

Table 5-l. Device Workload Checklist to be Completed by Driver After on-the-Road Use of In-Cab Device 5-4

LIST OF FIGURES

Figure l-l. A Generic Control-Theoretic Model of Driving l-4

Figure l-2. Venn Diagrams of Proportions of Variability Between an Arbitrary Human Factors Measure, Workload, and Safety 1-11

Figure l-3. Number of Crashes Distributed by Sources of Attentional Distraction, Broken Down Into Interior Source and Dash/Console/Steering Column Instrumentation Group 1-17

Figure l-4. Flow Diagram of Device Assessment Process From Driver Workload Perspective l-20

Figure 4-1. Distributional Transformations 4-5

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 Roadscene 5-5

TABLE OF CONTENTS (CONTINUED)

Page

LIST OF APPENDICES

APPENDIX A. VISUAL ALLOCATION MEASURES IN DRIVER WORKLOAD ASSESSMENT . . . . A-l

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

APPENDIX C. DRIVER-VEHICLE PERFORMANCE MEASURES IN WORKLOAD ASSESSMENT . . . . C-l

APPENDIX D. ALTERNATIVE SUBJECTIVE WORKLOAD ASSESSMENT TOOLS . . . . D-l

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

APPENDIX F. EXPERIMENTAL DESIGN STRATEGIES FOR DATA COLLECTION AND ANALYSIS . . . . F-l

APPENDIX G. EXAMPLE SUBJECT INFORMED CONSENT FORM . . . . G-l

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)

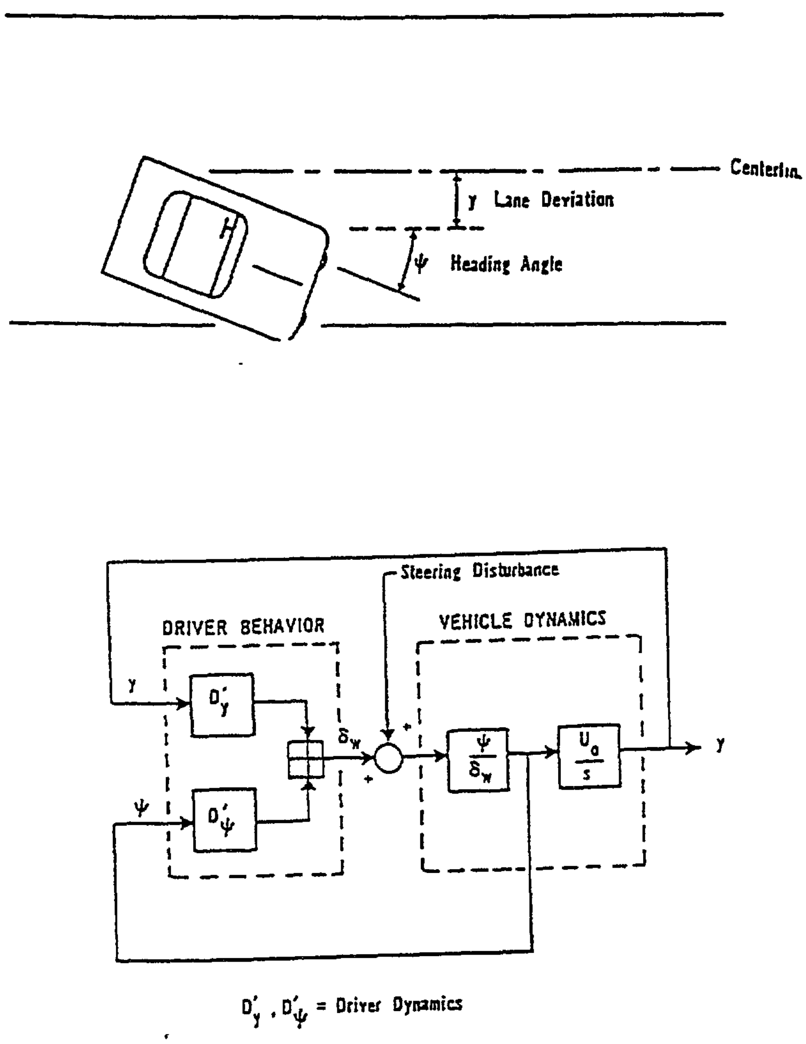

Figure 1-1. A Generic Control-Theoretic Model of Driving. Source: Hess, 1987.

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 (α_w) 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 (α_p) 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 l-l. 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.

IHuman 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
Svstems 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 $r_{HR\text{-}S}$ = 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 coeffkients 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

Venn Diagrams for Correlations Between Heart Rate, Workload and Safety Measures

$r_{HR*WL} = 0.7$, $r_{WL*S} = 0.7$, $r_{HR*S} = 0.0$

$r_{HR*WL} = 0.7$, $r_{WL*S} = 0.7$, $r_{HR*S} \approx 1.0$

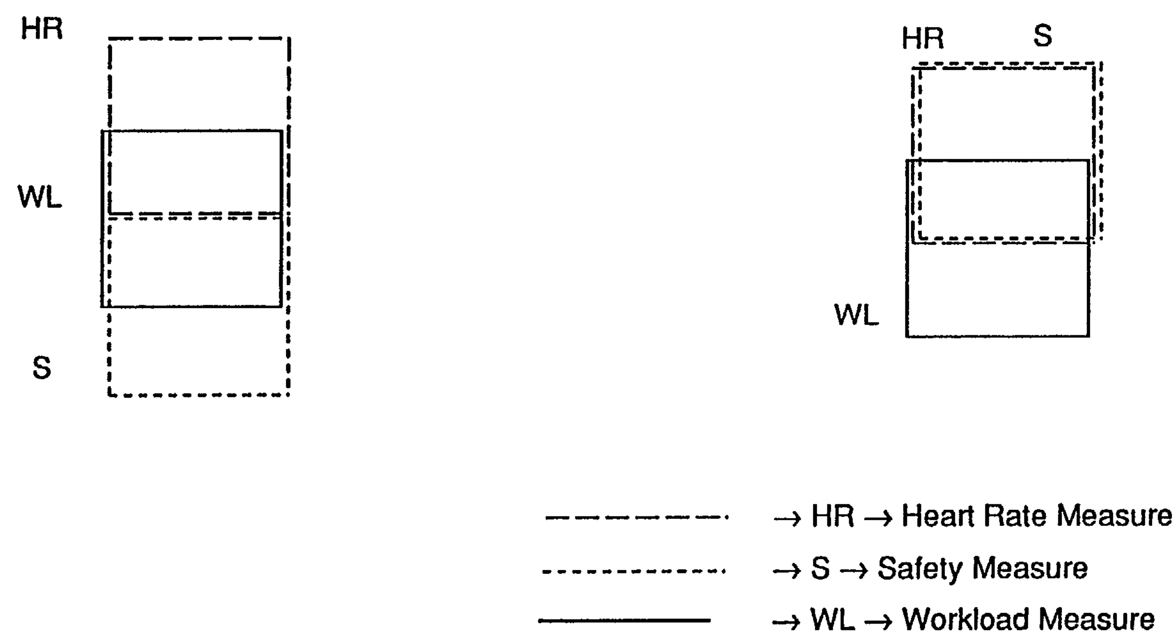

Figure 1-2. Venn Diagrams of Proportions of Variability between an arbitrary Human Factors Measure, Workload, and Safety.

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 Visual Allocation

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).

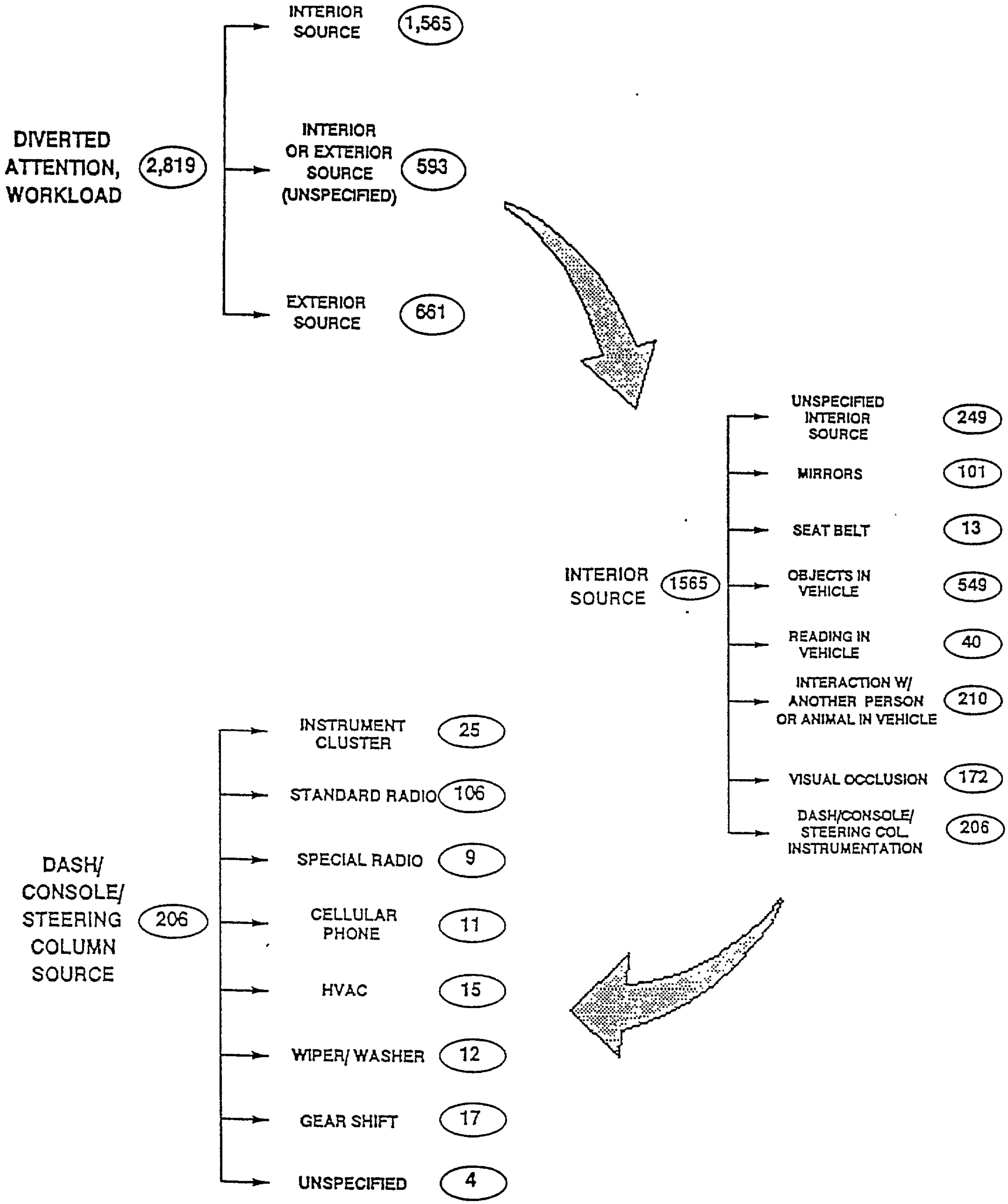

Figure 1-3. Number of Crashes Distributed by Sources of Attentional Distraction, Broken Down into Interior Source and Dash/Console/Steering Column Instrumentation Group (Source: Wierwille and Tijerina,1994).

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 <u>Speed Measures</u>

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 <u>Time Headway</u>

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 .O 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

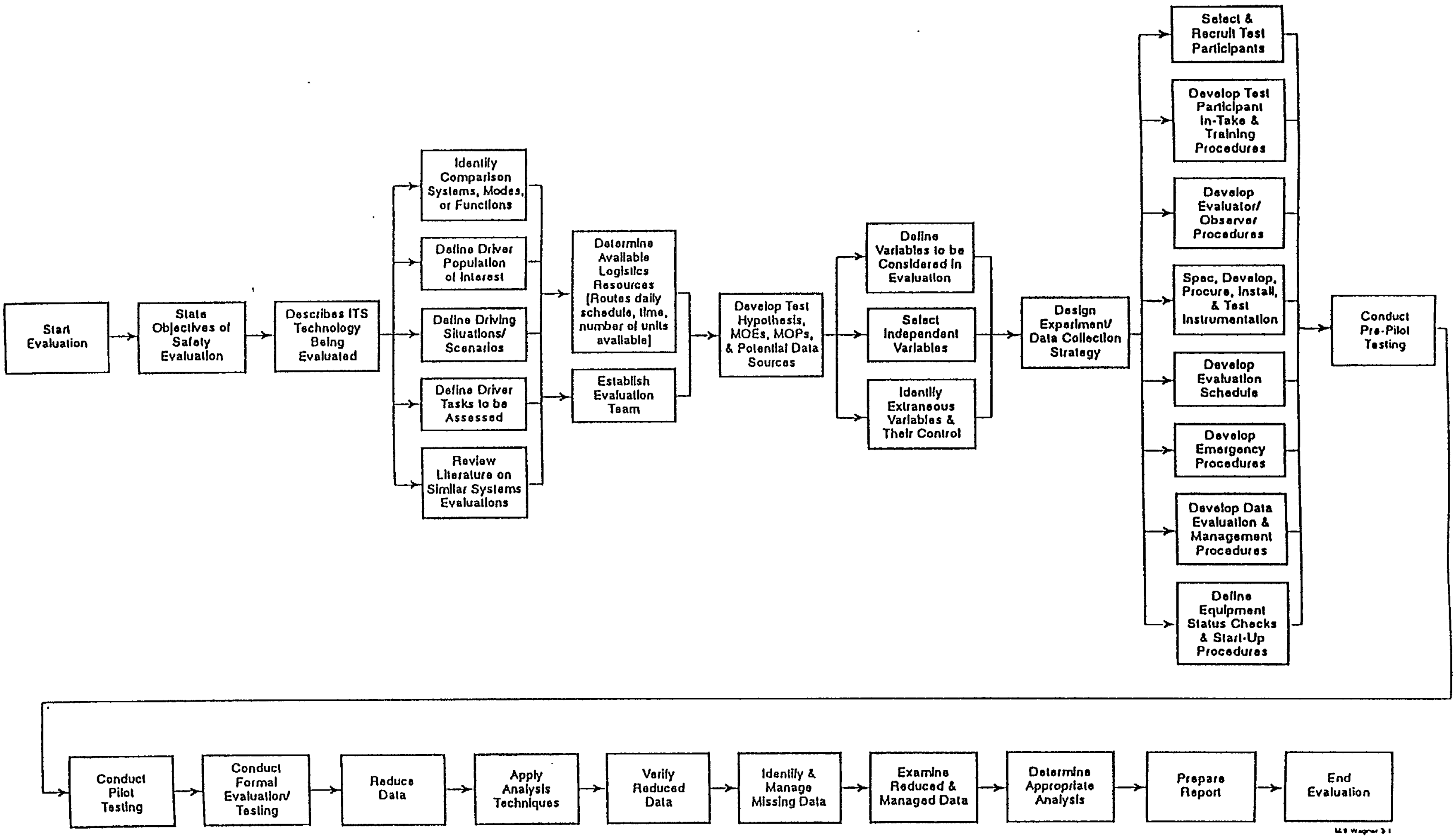

Figure 1-4. Flow Diagram of Device Assessment Process From Driver Workload Perspective.

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 exnerimenters 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 diffkult. 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 Gender Driving Experience Device Experience Skill or Abilities Permanent Handicaps Altered States (fatigue, intoxication, inattention) Motivation	Selection Selection Selection and Training Selection and Training Test & Selection Selection Various techniques (.e.,g., prolonged driving, administration of alcohol, visual occlusion, distractors) Instructions, payoffs, feedback	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.

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 SCI-IEDULE

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 <u>Verify reduced data</u>

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 <u>Identifv and Manage Missing Data</u>

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.

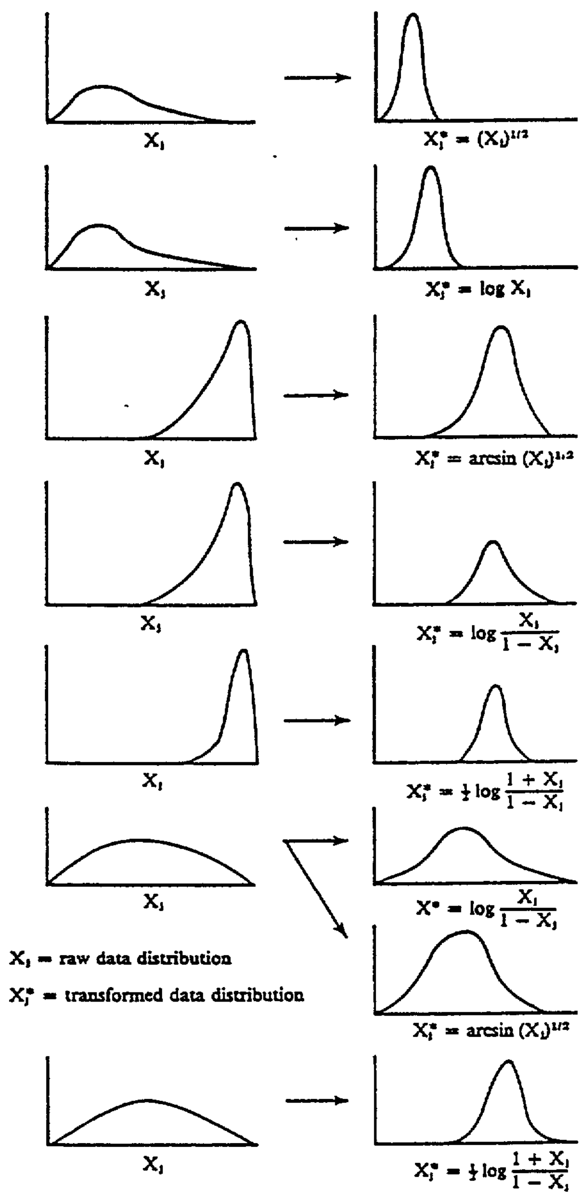

Figure 4-1. Distributional Transformations (Source: Rummel, 1970)

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 ahnost 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™), the Statistical Package for the Social Sciences (SPSS™), or the BioMedical Data Program (BMDP™) are quite sophisticated and have their own programming language. On the other hand, spreadsheets like Excel™ 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 difficu1t 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.	Ifthere is a visual display, does it display sets of numbers, text, map information, or other complex data?	**NO**	**YES**	________
5.	ln your opinion., is the device hard to read under normal conditions?	**NO**	**YES**	________
6.	Ifthe 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.	ln 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.	ls 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.

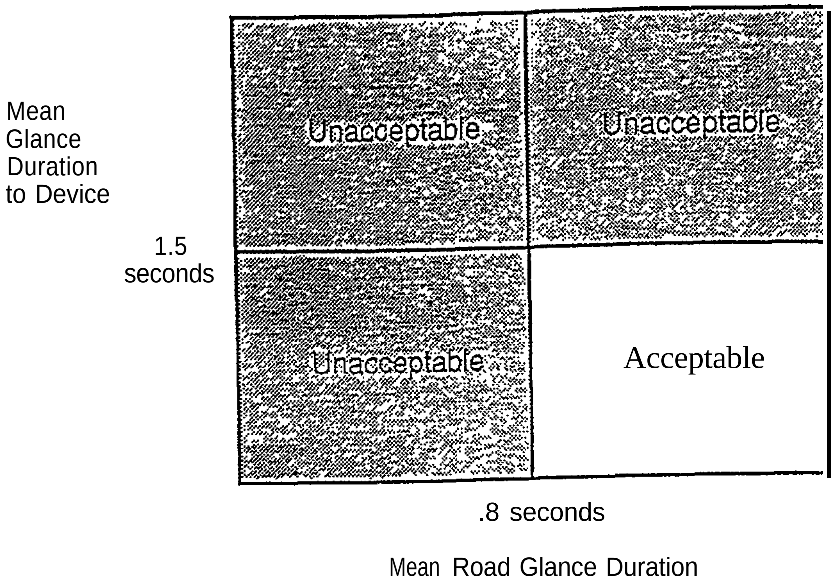

Figure 5-l. Hypothetical Workload Nomograph Based on the Visual Allocation Measures of Mean Glance Duration to the Device and Mean Glance Duration to the Road Scene. Shaded Areas Constitute Unacceptable Workload.

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,* 123 9 - 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.