



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



People Saving People
<http://www.nhtsadot.gov>

DOT HS 808 467

October 1996

Final Report Supplement

Heavy Vehicle Driver Workload Assessment

Task 4: Review of Workload and Related Research

This document is available to the public from the National Technical Information Service, Springfield, Virginia 22161.



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.

1. Report No.	2. Government Accession No. DOT HS 808 467 (4)	3. Recipient's Catalog No.	
4. Title and Subtitle Final Report Supplement -- Task 4: Review of Workload and Related Research		5. Report Date September, 1992	
		6. Performing Organization Code	
7. Author(s)W. Wierwille, L. Tijerina, S. Kiger, T. Rockwell, E. Lauber, and A. Bittner, Jr.		6. Performing Organization Report No.	
9. Performing Organization Name and Address Battelle Memorial Institute 505 Ring Avenue Columbus, OH 43201-2693		10. Work Unit No. (TRAIS)	
		11. Contract or Grant No. DTNH22-91-C-07003	
		13. Type of Report and Period Covered FinalReport July, 1991~September, 1995	
12. Sponsoring Agency Name and Address Office of crash Avoidance Research National Highway Traffic Safety Administration 400 7th Street, NRDJO Washington, DC 20590		14. Sponsoring Agency Code	
15. Supplementary Notes Contracting Officer's Technical Representative: Dr. Michael J. Goodman (NRD-50)			
16. Abstract This report reviews literature on workload measures and related research. It depicts the preliminary development of a theoretical basis for relating driving workload to highway safety and a selective review of driver performance evaluation, workload evaluation in driving contexts, and risk-taking and risk adaptation behavior. An actuarial approach to establish the safety relevance of workload measures in discussed in conjunction with a driver resources allocation model of in-cab device workload. The report concludes with the presentation of a driver resources-based taxonomy of in-cab tasks and candidate workload measures and their potential sensitivity to tasks described in the taxonomy.			
17. Key Words human factors, heavy vehicle, safety, workload measures, driver performance evaluation, risk-taking , risk adaptation, actuarial approaches, driver resource allocation model, driver taxonomy		16. Distribution Statement	
19. Security Classif. (of this report)	20. Security Classif. (of this Page)	21. No. of Pages 96	22. Price

TABLE OF CONTENTS

	<u>Page</u>
1.0 BACKGROUND	1-1
Introduction	1-1
Objective	1-1
Organization of the Report	1-1
A General Model of Driver Workload Management	1-2
Operating Practices and Delay	1-4
Driving Demands, In-cab Device Demands	1-4
Explicit Focus	1-5
Implicit Focus	1-5
Long-Term Memory (LTM)	1-6
Task Scheduling	1-6
Action: Resource Allocation and Execution	1-7
Implications of the Model	1-8
Approaches to Establishing the Safety Relevance of Driver Workload	1-10
Actuarial Approach	1-10
Inferential Approach	1-12
Use of Subjective Workload Measures & Secondary Task Techniques	1-12
References	1-17
Appendix 1A: Workload Measure Reviews: Tabulations for Selected Subjective and Secondary Task Measures	1A-1
2.0 A DRIVER RESOURCE ALLOCATION MODEL OF IN-CAB DEVICE WORKLOAD ..	2-1
Introduction	2-1
Review of Driver Resources	2-2
Visual Resource	2-2
Manual Resource	2-13
Cognitive Resource	2-14
Auditory Resource	2-15
Other Resources	2-15
Resource Ranking	2-16
A Driver Resources-based Taxonomy of In-cab Tasks	2-16
Quantification of Resource Usage	2-18
References	2-22
3.0 DRIVER-VEHICLE PERFORMANCE MEASURES AND PERTINENT TASKS	3-1
Introduction	3-1
Definitions of Driver-Vehicle Performance Measures Potentially Sensitive to In-Cab Device Workload	3-1
Likely Sensitivity of Driver-Vehicle Performance Measures to In-Cab Device Workload ..	3-8
Task 1: Visual only, Low Cognitive Load	3-8
Task 2: Visual only, High Cognitive Load	3-8
Task 3: Visual Primarily, Low Cognitive Load	3-8
Task 4: Visual Priiily, High Cognitive Load	3-9

TABLE OF CONTENTS
(Continued)

	<u>Page</u>
Task 5: Visual-Manual, Low Cognitive Load	3-9
Task 6: Visual-Manual, High Cognitive Load	3-9
Task 7: Manual Primary, Low Cognitive Load	3-9
Task 8: Manual Primary, High Cognitive Load	3-9
Task 9: Manual Only, Low Cognitive Load	3-9
Task 10: Auditory-Vocal	3-10
Development of a Sensitivity Matrix	3-10
References	3-15
Appendix 3A	3A-19
4.0 RISK TAKING AND DRIVER ADAPTATION IN THE USE OF HIGH-TECHNOLOGY IN-CAB DEVICES	4-1
Introduction	4-1
Driver Risk Taking Research.	4-2
Risk Adaptation Effects with the Use of New In-Cab Devices	4-3
A Proposed Methodology to Examine the Safety Criticality of In-Cab Devices	44
References	4-7
5.0 DISCUSSION.	5-1
The Importance of Approximate Science	5-1
Necessity for Realistic Tasks vs. “Laboratory” Tasks	5-2
First Order vs. Second Order Issues	5-2
The Use of Baseline Measurements With Conventional Tasks	5-2
Criticality of Primary Task Performance Covariation With In-Cab Device Use	5-3
Problems With Best-Case/Worst-Case Analysis	5-3
The Effects of Individual Differences	5-3
Tradeoffs Among Device-Induced Workload and More General Safety Benefits	5-3
Multiple Transactions and In-Cab Device Workload Assessments	5-3
Problem of Interactions Between Driving Tasks and In-Cab Tasks	5-4
Novelty Effects.	5-4
Refining the Measures of Merit	5-4

TABLE OF CONTENTS
(Continued)

List of Tables

	<u>Page</u>
Table 1-1 Driver Strategies for Coping with High Workload and the Use of In-cab Devices	1-9
Table 2-1 Summary of Glance Duration Data for Conventional Tasks (Rockwell, 1987)	2-6
Table 2-2 Summary of Results Presented by Bhise, Forbes, and Farber (1986)	2-7
Table 2-3 Total In-car Glance Times for a Variety of Conventional and Navigation Tasks (Dingus et al, 1989)	2-9
Table 2-4 Average Length and Number of In-car Glances for a Variety of Conventional and Navigation Tasks (Dingus et al, 1989)	2-10
Table 3-1. Driver-Vehicle Performance Measures	3-2
Table 3-2. Likely Sensitivity of Driver-Vehicle Performance Measures To Prototypic In-Cab Tasks	3-12

List of Figures

Figure 1-1 General Model of Driver Workload Management	1-3
Figure 1-2 Block Diam of the Steps Involved in Relating Accident Data to Workload	1-11
Figure 1-3 Relation Between Average Glance Frequency and Lane Crossings (Source: Berger, Smith, & Ziedman, 1989)	1-13
Figure 2-1 Driver Visual Sampling	2-5
Figure 2-2 Visual Sampling Model	2-11
Figure 2-3 Depiction of Competition for Driver Resources ,	2-17

1.0 BACKGROUND

Introduction

Several high technology in-cab devices have been proposed for use in heavy trucks. Many of these devices introduce subsidiary tasks which may compete with the driver's primary task of safely controlling the vehicle. Inattention of the driver to a crash hazard is assumed to play a role in 30 to 50 percent of accidents (Treat, et. al., 1977; Sussman, et. al., 1985). The challenge is to determine the efficiency, effectiveness, and safety of such devices from the driver's perspective.

In this research program, the Battelle team (which includes members from Battelle and R&R Research, Inc., a truck driver consultant, and various university consultants) will develop a workload assessment protocol. This protocol will be able to assess the safety implications of high technology systems that might be introduced into heavy trucks.

Objective

This report's major objective is to review pertinent workload and driver performance evaluation research and to develop a theoretical basis for relating driver workload to highway safety. The team conducted a selective review of literature that concerned driver performance evaluation, workload evaluation in a driving context, and risk taking and risk adaptation. An approach for assessment of in-cab devices was devised from these data and models to establish the safety relevance of workload measurement.

Organization of the Report

This report is organized in the following way:

- Section 1.0 provides a framework for considering driver workload induced by high-technology in-cab devices and sets out a number of key assumptions.
- Section 2.0 introduces a driver resources approach to in-cab device evaluation. It provides a review of key references on visual, manual, and cognitive loads induced by various classes of in-cab devices. It presents a model of visual allocation. It also includes a taxonomy of in-cab tasks, which is useful for organizing an evaluation protocol.
- Section 3.0 provides definitions of various classes of driver-vehicle performance measures which may be suitable for workload assessment. Prototypic in-cab tasks are presented and then related to various driver-vehicle performance measures with varying degrees of safety relevance. The likely sensitivity of candidate driver-vehicle performance measures is included in a matrix to guide selection of measures for preliminary testing.
- Section 4.0 discusses of risk-taking and risk adaptation associated with high technology in-cab device use. Section 4.0 also addresses the usefulness of the concept

of risk-taking and methods to assess risk-taking and risk-adaptation when using in-cab safety devices.

A General Model of Driver Workload Management

Traditional workload literature, such as Donchin and Gopher (1986); O'Donnell and Eggemeier, (1986); Hart and Wickens (1990), addresses the issues of "spare capacity" and the extent to which a person can do two things at once. There is still no generally agreed upon definition of workload, despite over 25 years of work. Different researchers find conflicting results; many studies remain theoretical or methodological in nature rather than oriented toward system development applications.

Hart (1989, cited in Adams, Tenney, and Pew, 1991) attributes this to the artificiality of the laboratory methods employed in such research. These methods include: pushing people to work to their limits without opportunity to adopt realistic coping strategies; using trials measured in minutes rather than realistic intervals of hours; forcing people to respond in particular ways (e.g., immediately and consistently); and imposing homogenous demands, as opposed to introducing more realistic sequential as well as overlapping demands of various magnitudes with various costs and benefits. Hart concludes that such workload assessments do not provide a realistic assessment of what people can actually do and how they cope with various demands that may sometimes exceed what can be done.

This led the team to consider the safety implications of driver workload differently. Instead of asking how many things the driver can do at once, the team became interested how a driver manages workload by prioritizing and scheduling both driving tasks and in-cab device tasks. This approach was influenced by the recent work of Adams, Tenny, and Pew (1991). Adams, et. al. presents a model of human performance based on a synthesis of literature that concerns human perception, cognition, text comprehension, memory, and skill. In this model, what the operators perceive is strongly driven by their understanding of the current situation (schema); their actions are prioritized and scheduled for execution because of this understanding and their goals. The driver's schema of the current situation, then, shapes what is perceived and attended to by a top-down process. Sensory cues also shapes the current schema by a bottom-up process. A central tenet is that human behavior is goal oriented and that task scheduling and execution is subordinated to the driver's goals and understanding of the current driving situation. Thus, whether a driver focuses on an in-cab task or on the driving task depends on an understanding of the driving situation, the importance to the driver's goals of working on one or the other at the next time interval, and the ease with which the driver can complete a task, interleave tasks, abruptly halt then smoothly resume an uncompleted task, or delay task initiation. The following text introduces this approach in order to draw performance predictions.

Figure 1-1 presents a general model of driver workload, introduces several important points that relate workload to highway safety. This model attempts to integrate Wickens and Flach's (1988) model of decision making with Adams, Tenny, and Pew's (1991) model of strategic workload and cognitive management of multiple tasks. It also incorporates findings from Task 3 of this project (Kiger, et al., 1992). In general, environmental demands (driving conditions, in-cab device demands, and the effects of operating practices and delays to a schedule) represent the world of the driver on the road. The driver processes cues from these elements of the world and makes sense of them based on an understanding of the current situation (schema). This understanding then provides inputs to a process of task scheduling and driver resource allocation that, when suitably modified by the driver's

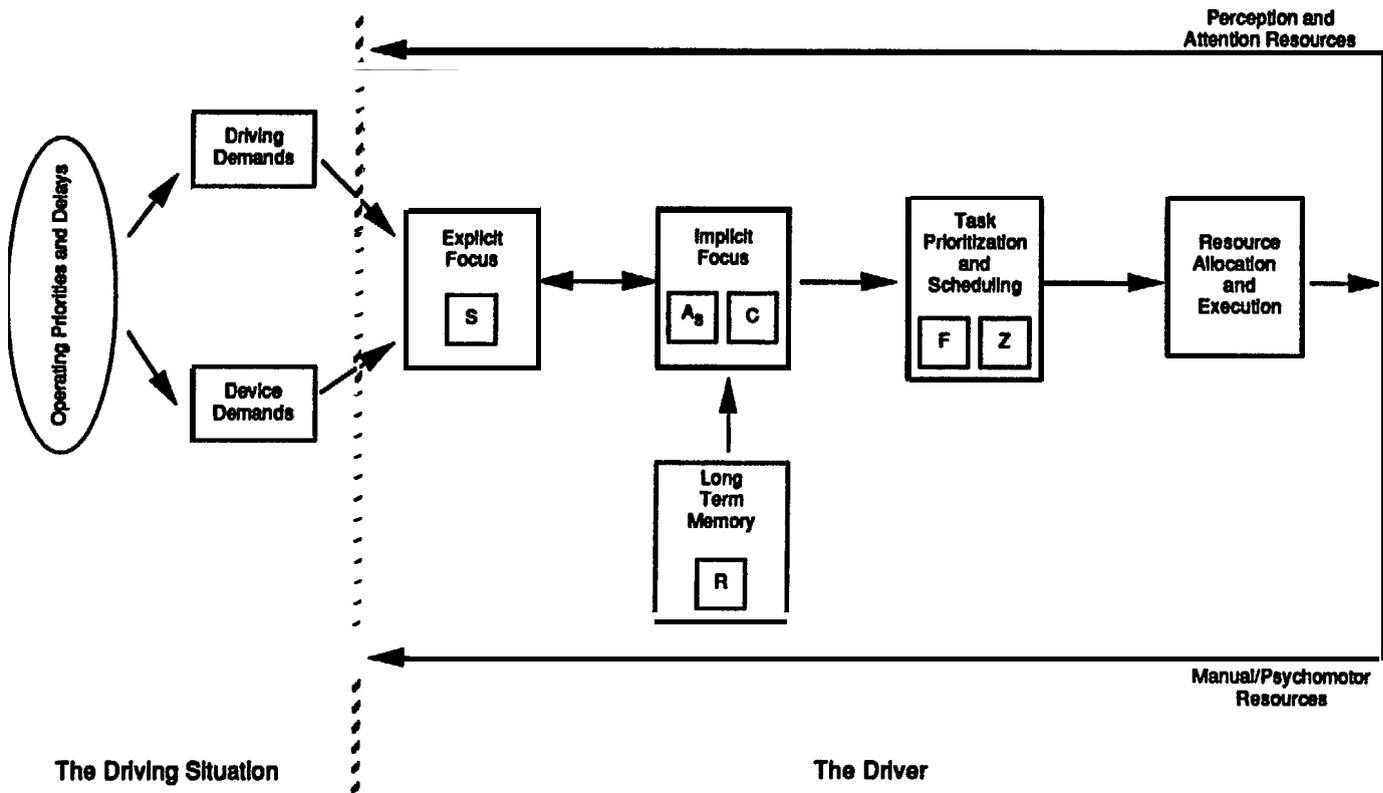


Figure 1-1
General Model of Driver Workload Management

goals, ultimately determines where the driver next attends and interacts. If all goes well, this process results in safe driving and safe in-cab device use. However, there are a number of biases or heuristics a driver may use that can lead to misjudgments about the current driving situation and the appropriateness of in-cab interactions. Each of the components in the model are presented next.

Operating Practices and Delay. In Task 3 of this project (Kiger, et al., 1992), it was found that heavy vehicle drivers do not think of workload in terms of the momentary demands placed on them by driving tasks, in-cab device transactions, or other driver activities while on the road. Instead, they consistently consider workload primarily in terms of time stress caused by delays to their planned schedule. This finding is consistent with the situational characteristic labeled “management/operating practices” reported in Clarke, Leasure, Radlinski, and Smith (1987).

The reasons why drivers consider delays to their schedule so stressful are understandable. If the driver arrives late at a destination, several costs may have to be paid. The driver may be censured by a dispatcher or warehouse manager. The driver’s pay may be penalized. There may be a further delay while the driver waits for an available space at the loading dock. The consignee may not unload a shipment until the next day, forcing the driver to wait and delay the loading of the next shipment. This is especially troublesome for irregular route drivers who must move from one destination to the next under tight time constraints. There may be additional physical work if the driver must do all loading or unloading because warehouse workers or personnel who would have done this are no longer available. All of these factors create stress in the driver who is behind schedule.

Delays to schedule can impact the driving condition demands. If the driver tries to make up lost time, greater demands may be encountered in the driving task perhaps because the driver drives faster, accepts shorter headways, makes more frequent lane changes, and so on. These adaptations in driving will reduce driving tolerances and minimize the resources available for in-cab activities. Thus, tight schedules or delays to schedule may increase the difficulty of the primary task of safely controlling the vehicle at all times.

Delays to schedule can impact in-cab device demands as well. Management practices may require more voice communications systems use for status reports when a delay occurs. The driver may take a different route, which perhaps increases the need to interact with a land navigation and routing system. The driver may be sent more information on a text on a text display about a new route or change in procedure that will occur at the destination. As these examples illustrate, delays to schedule can affect in-cab device workload by increasing the frequency or complexity of in-cab device transactions for communication systems, vehicle routing systems, text messaging systems, and the like.

Driving Demands. In-cab Device Demands. Driving demands are those placed on the driver from prevailing driving conditions such as traffic density, roadway type, lighting, roadway geometry, traction, and visibility. Vision is the prime element in safe driving and the road scene, therefore, makes unrelenting demands on visual attention. Driving places response demands on the driver as well. The driver usually has two hands that can be used to perform various tasks. For instance, the driver can keep one hand on the steering wheel while using the other to manipulate the controls of an in-cab device. Even so, there are instances where both manual hands must be allocated to driving, e.g., two-handed steering around a sharp curve or to perform an evasive maneuver.

Many manual tasks have a visual component. The driver may need to use a visually-guided movement to acquire a control, even though that control can then be manipulated without visual feedback (e.g., flipping a light switch on the instrument panel). Alternatively, a driver may need visual guidance for more or less continuous adjustment of a control until a desired setting is reached, such as tuning a radio to a specific frequency.

Other effector systems, such as foot pedal adjustments and voice commands, play a lesser role in determining driver workload. Finally, cognitive resources play a role. A driver must continually interpret visual and other sensory information in order to react to and anticipate driving events, and keep track of the stream of on-going tasks.

The central thesis of this project is that in-cab devices may demand the same set of driver resources as the driving task. Resources are limited, however, and the driver largely allocates these resources to one task at a time. Though the driver may do several things at once, thoughtful or conscious attention is focussed on one thing at a time. Safety is compromised when in-cab tasks demand resources and the driver allocates them at inopportune times, relative to the primary task of safety controlling the vehicle at all times.

Explicit Focus: At any moment, the driver is inundated with many sensory cues which come from both without and within the cab of the heavy vehicle. There is so much information potentially available that drivers attune themselves to those cues which are most appropriate to their interests and knowledge about the current situation (Gibson, 1966). Adams, et al., (1991), adapting theory from text comprehension, call this focus “explicit focus”. From explicit focus, a limited number of cues from the world and a rich set of associations that make other knowledge from the operator’s memory available, are passed on. These contribute to driver understanding, or schemas, of the current situation.

Explicit focus implies that cues that are relevant to driver’s current tasks and understanding of the driving situation will be processed readily. Since this focus is attuned to different cues based on driver’s goals and understanding, a salience bias (S in Figure 1-1) may arise (Wickens, 1992). Thus, only a small number of cues are focused on, even though others may be more critical or necessary to guide action. For example, a driver on an open highway may notice a slow-moving vehicle ahead (a salient cue) but fail to notice in the west coast mirror, that a motorcyclist has entered onto the roadway and is closing from behind (an important, but unnoticed cue). The driver may then make a lane change to overtake the slow-moving vehicle without sufficiently checking the adjacent lane. Safety is clearly compromised if the motorcyclist is in that adjacent lane.

Implicit Focus. Adams, et al., (1991) present implicit focus as a more or less activated part of long-term memory which overlaps with explicit focus. Together with explicit focus, this is the locus of a driver’s schemas. As mentioned earlier, environmental information is selectively processed with respect to this understanding; it also updates and determines task priorities and scheduling. Adams, et al. point out that visual attention is largely determined, not by physical characteristics of the stimulus, but by the perceivers’ intentions and goals. Exceptions in peripheral vision are given by sensitivity to movement and the abrupt onset of display elements (e.g., blinking lights at a railroad crossing). Nevertheless, busy people show shrinkage in the functional field of view such that even these physical features may be ignored.

Given that implicit focus (together with explicit focus) is where the driver's impression of the driving situation is formed, it is the source of as-if bias (A, in Figure 1-1). The as-if bias is a tendency to proceed "as if" the situation is what it seems to be. For example, the driver passes a parked car without slowing or making a lane change "as if it will remain parked and not start entering the traffic lane. The driver diverts attention to an in-cab device "as if" there is nothing ahead on the roadway to be concerned about. This is believed to be a significant part of the driver inattention problem which has been reported in crash statistics (e.g., Hendricks, et al., 1992). In this sense the crash causal factor "driver inattention" is a misnomer; the driver involved in such a crash was inattentive to the crash hazard, but was attentive to something else.

A related bias in human performance based on a person's understanding of or hypothesis about the driving situation, is known as the confirmation bias. This is the bias that causes people to seek out information that confirms their expectations and to ignore disconfirming evidence. While this bias was originally found in decision making and fault diagnosis (Wickens, 1992), it may be applicable to driver workload as well. That is, drivers will seek to affirm, not falsify, their understanding of driving conditions and allocate resources according to that expectation, whether it is true or not. For example, a driver who is convinced that the roadway ahead is clear for highway speeds may seek out additional visual data to confirm that it is safe to continue to drive faster than 45 mph during snow (e.g., other vehicles are traveling at this speed) and fail to check outside ambient temperature (which might indicate freezing and ice).

Long-Term Memory (LTM). Long-term memory may be thought of as being either episodic or semantic in nature. Episodic memory represents knowledge about specific events: for example, what happened on the last haul through this intersection or this route. Semantic memory represents memory for meaning, e.g., how to steer out of a skid. Both kinds of memory are important for driver situation assessment and Adams, et al., (1991) explain that knowledge in long term memory is largely latent and can be accessed only with considerable effort or particularly strong cues.

In general, people make use of a 'representativeness' heuristic (R in Figure 1-1) in which people try to understand a current situation by finding a typical or representative pattern in episodic memory (Tversky and Kahneman, 1974). For example, the driver who recalls that people have a tendency to sometimes run a particular stop sign will probably proceed through that intersection with caution.

Task Scheduling. Drivers schedule tasks based on their understanding of the current situation and their goals. Drivers' goals and plans determine their actions, direct their attention in certain directions, define the relevance of certain cues, and so on (Gibson, 1966; Neisser, 1976; Adams, et al., 1991). It is not possible to discern the nature of the scheduling "table" that the driver uses to start, stop, interleave, delay, and drop tasks. However, there are two effects associated with scheduling which may be of particular importance to the evaluation of in-cab device workload. The first is known as the Zeigarnik effect (Zeigarnik, 1965); the second is known as the problem of framing (Tversky and Kahneman, 1981).

The Zeigarnik effect (Z in the figure) states that, when people are interrupted before task completion, they are strongly motivated to come back and finish it. In addition, the memory for the task is quite strong and persists even when the individual knows that the task will not be resumed (Zeigarnik, 1965). This suggests that one goal drivers may have is to finish in-cab tasks which they started. This may prompt them to continue to interact with the device and when it is perhaps ill-advised and the incompleted in-cab task will occupy space in working memory when they are not working on it,

The problem of framing (F in the figure) relates to deciding between two alternatives, one a risk and the other a sure thing, and the way in which those alternatives are perceived, i.e., as losses vs. gains. Insurance buying notwithstanding, Tversky and Kahneman (1981) found that people are biased to choose the risky loss rather than the certain loss, even when the expected loss from the former is greater. In heavy vehicle operation, drivers might choose between turning back in the face of bad weather (with the certain delay that will cause) and continuing on (with the risk that they may have an accident vs. the chance they may get through safely). These options are framed as a choice between two losses. The framing bias predicts the risky choice (carry on).

On the other hand, when the problem is framed as a choice between gains, the sure-thing is preferred. Thus, if the driver perceives the same choice as one of ensuring that property and his life are safe vs. possibly staying on schedule or possibly having an accident, the bias would be toward the former option. The implication of this for in-cab device workload and highway safety is that, if device interaction is perceived as a risk with loss consequences (maybe lose data, time and effort) if the transaction isn't completed) in the face of certain loss (e.g., have to slow down, have to stop and pull off the road to be sure the transaction is safely completed), the driver will likely be attracted to the riskier option.

Recent research into how people schedule tasks has possible applicability to in-cab device use. Tulga and Sheridan (1980, reported in Sheridan, 1991) had people work on multiple tasks which varied in importance and duration. Under high stress, subjects tended to do whatever needed to be done in a hurry regardless of the fact that some other long-term very important task could be worked on 'ahead of schedule'. This suggests that, if the driving task is perceived as the most time-limited process, drivers should attend to it regardless of what in-cab device tasks need to be done. This is evident by the limits drivers tend to put on glance durations to in-cab tasks (see Section 2-O of this report). On the other hand, if the in-cab task is considered time-limited and is important to the driver's goals, the driver may attempt to finish it and give the driving demands fewer resources. In either event, time-limiting in-cab devices (or managerial policies on their use) pose a special risk for highway safety which deserve careful scrutiny.

Action: Resource Allocation and Execution. Once drivers decide on what to do next, they allocate their resources accordingly. They devote perception and attention to particular aspects of driving or in-cab device demands, and manual or psychomotor inputs to those demands as well. The perception-assessment-scheduling-execution cycle then begins again. Subsequent sections of this report discuss the nature of these resources and how they may be allocated. It is self-evident that visual resources are of primary importance to driving.

Hart (1989) has correctly pointed out that pilots (and presumably drivers) will perform according to the following guidelines:

- Complete tasks within a time window or by a deadline rather than immediately.
- Complete tasks with an acceptable (rather than an optimal) level of performance, and
- Trade-off less important for more important tasks.

Thus, measures of timeliness are more appropriate than measures of speed, and measures of acceptable performance are appropriate and specific deviations within that range are irrelevant. Understanding how tasks are traded off and scheduled becomes highly relevant.

Implications of the Model. The framework presented above leads to the following implications:

- Drivers will act consistent with their goals and their understanding of the driving situation (Adams, Tenny, and Pew, 1991). This means that most in-cab technologies may be used safely most of the time if they are quick and easy to work with. Otherwise they may not be used at all (Rockwell, 1987), unless operational practices do not make such device use optional.
- Safety is most likely to be compromised when a driver believes it is appropriate to work within in-cab device and, in fact, safety hazards exist on the roadway.
- Delays to schedule represent a pervasive stressor which can increase driving condition demand and in-cab device demand.
- Drivers may be able to do many things at once, but they can focus consciously on only one task at a time. Generally, this focus is visual in nature for driving (Wierwille, 1992).
- Once a driver begins a task, if it is not completed, there is a tendency to want to go back and finish it (Zeigarnik, 1965). This impulse to return to the interrupted task grows with the relevance of the task to current goals and subgoals of the driver. Incomplete tasks take up space in working memory.
- If device interaction is perceived as a risk with loss consequences (possibly lose data, time, or effort in the transaction isn't completed) in the face of certain loss (e.g., have to slow down, pull off the road, or stop to be sure transaction is safely completed), the driver may tend toward the riskier option which keeps the vehicle on the road and moving to the next destination.
- Switching one's attentional focus from task to task can take time and effort. As Adams, et al., (1991) stated, what a person is thinking about determines the contents of focal memory and the contents of focal memory determine what the person is prepared to think about.
- Measures of timeliness, acceptable performance, and tradeoffs among driving and in-cab tasks are preferred for workload assessment. Some examples of tradeoff strategies a driver may use are listed in Table 1-1.

Table 1-1
Driver Strategies for Coping with High Workload
and the Use of In-cab Devices

When confronted with high workload, drivers are likely to do one or more of the following:

- delay the start of an in-cab interaction
- finish an in-cab interaction before entering a 'high' workload driving condition
- stop an in-cab interaction

- increase attention to one of the concurrent tasks
- increase the headway to the vehicle ahead
- slow down the vehicle to decrease driving demand

- turn off onto a side road, access road, or other less congested roadway
- pull off the road to avoid driving workload altogether
- delegate some tasks, perhaps to a team driver or to automation (e.g., turn on the headway detection system, cruise control)

- accept poorer performance on one or perhaps both tasks [e.g., miss some information coming in on voice communications, increase lane deviation)
- narrow the field of attention (e.g., ignore more driving condition factors in the periphery)
- rely more on expectations and assumptions than feedback processing
- monitor the road scene less often because of attention to the in-cab device (or vice versa)
- pursue possible explanations less thoroughly (e.g., while making decisions about unexpected events).

Approaches to Establishing the Safety Relevance of Driver Workload

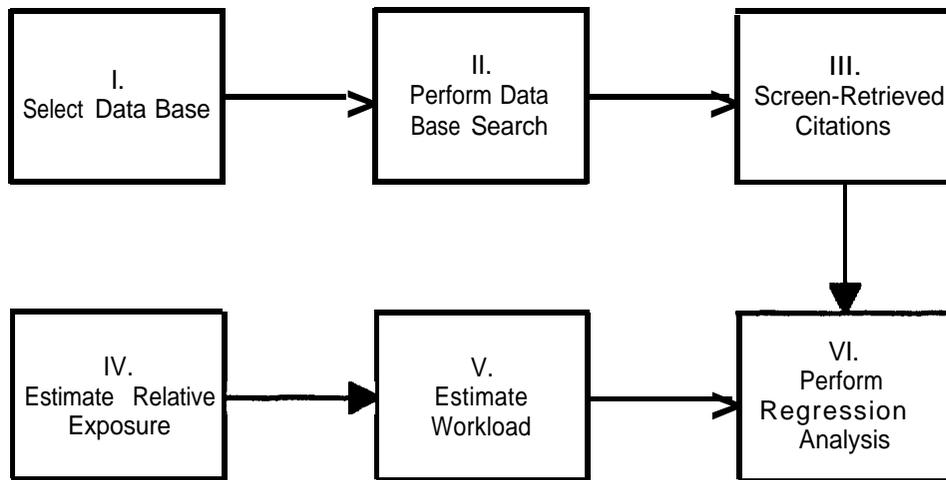
Workload measures must be correlated to safety-critical changes in driver-vehicle performance in order to relate driver workload demanded by in-cab devices to highway safety. There appear to be two general approaches to establishing this correlation. The first is termed an actuarial approach, which attempts to relate in-cab loads to crash incidence. The second is an inferential approach; in-cab measures are correlated with driver-vehicle performance measures that have prima facie safety relevance. A discussion of each of these approaches follows.

Actuarial Approach. This approach to problems occurring with in-vehicle controls (first applied by Perel (1976), attempts to relate crash statistics to in-vehicle workloads. Perel found, through a review of detailed accident reports, the proportion of crashes in a North Carolina data base caused, for example, by lack of attention while tuning the in-car radio. One might use this approach to determine how several classes of in-cab distractions with different levels of workload relate to crash incidence.

Figure 1-2 presents a block diagram of a refinement to Perel's approach. The refinement relates workload to the increased likelihood of crash occurrence:

- First, a data base must be selected with records of sufficient detail to determine the probable cause of the crash event. The data base must also allow for searches by key words and phrases related to in-cab tasks (e.g., "attending to radio", "reading a map", "taking eyes off the road", etc.).
- Second, a list of search terms must be developed and refined.
- Third, retrieved citations must be screened for those that are not workload-related. The workload-related citations are then placed in categories, which will likely be device specific, such as CB, cellular phone, radio, HVAC, lighter, instrument panel, map reading, etc.
- Fourth, the relative exposure of device categories (i.e., the base rates of use) must be estimated. For example, the instrument cluster is used frequently whereas map reading is performed infrequently. The workload-related citations will reflect both frequency of use and driver workload. To complicate matters further, some devices in the vehicle fleet are common (e.g, radio), while others are less so (e.g., cellular phone). It may be that rank ordering of frequency of use and availability in the fleet will, over time, be feasible even if more exact estimates are not.
- Fifth, the device workloads must be estimated (e.g., determine visual, manual, or cognitive load). Given that a range of in-cab devices with a range of workload demands are uncovered in the data base, average values of workload (perhaps drawn from the scientific literature) may be sufficient to characterize the different demands associated with different crash incidence rates.
- Sixth, a regression analysis must be performed to relate the product of a given workload level and exposure rating to its associated crash incidence.

Figure 1-2
Block Diagram of the Steps Involved in Relating Accident Data to Workload



This actuarial approach has a number of challenges associated with it. These include the availability of sufficient crash reports which span a range of in-cab device types, the ability to estimate exposure rates, reporting biases in the database citations, and so on. However, such a model, if constructed, would represent a substantial stride forward in developing the relevance of workload measures to highway safety. In fact, this approach is being pursued as part of this contract. More details of the method employed and results obtained will be provided in a special report once that effort is completed.

Inferential Approach The second approach tries to correlate device workload measures to safety-relevant driver-vehicle performance. Rather than relate workload to crash incidence, workload is correlated with lanekeeping, gap acceptance, headway maintenance, obstacle detection, or other measures which are prima facie safety-relevant. The inference is that degraded performance does not always lead to a crash, but does increase the likelihood of a crash by reducing safe driving tolerances and the ability to recover in the event of unexpected events.

An example of the approach is provided here. It can be assumed that lane crossings are safety-relevant. Some lane crossings occur without incident. However, crossing a centerline increases the risk of a collision with a car traveling in the adjacent travel lane. Crossing the fog line increases the risk of roadway departure and/or colliding with roadside appurtenances. Therefore, the inferential approach may lead to an attempt to correlate a workload measure such as glance frequency demanded of in-cab devices with the occurrence of lane crossings while working with that device. An example of such a relationship is given in Figure 1-3, taken from Burger, Smith, and Ziedman (1989).

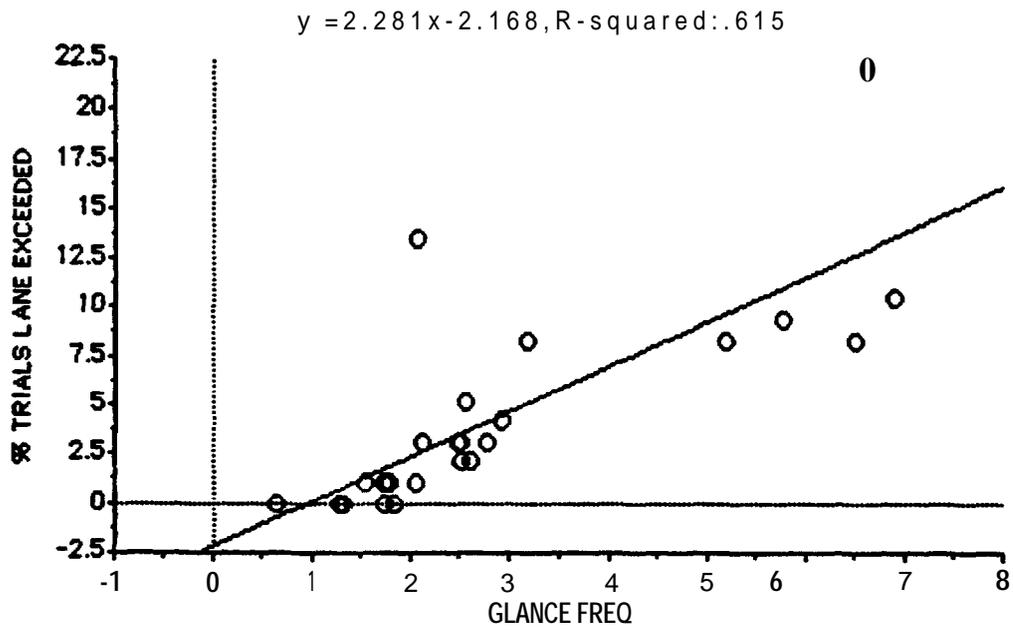
Another version of this approach dispenses with correlating device workload measures and driver-vehicle performance, and looks at performance directly. For example, McKnight and McKnight (1991), using an open-loop simulator, looked at the effects of cellular phone use on drivers reaction to traffic conditions such as stopped cars, traffic signal changes, lane terminations, and pedestrians stepping into the travel lane. They unobtrusively measured the frequency and correctness of driver brake, accelerator, and steering inputs. They found that hands-free and eyes-free voice communication led to slower response and more missed events than did driving without the distraction of the conversation. The safety implications of this method are obvious.

There are several benefits to using device-related workload measures to correlate with driver-vehicle performance measures. First, measures like glance duration or glance frequency may be more diagnostic than driver-vehicle performance in uncovering a workload-related device problem. Second, such measures may be more reliable (in the statistical sense), sensitive, or responsive to transient workload. Third, device-specific measures might be combined into a measure of merit (see Section 2-O) which can be used in a predictive model. Note, however, that if in-cab device workload is to be considered deleterious to highway safety, driving performance degradation with device use must be demonstrated under realistic conditions. Therefore, the inferential approach requires both device-specific measures of workload as well as driver-vehicle performance measures taken during in-cab device use.

Use of Subjective Workload Measures & Secondary Task Techniques

In subsequent sections of this report, various classes of driver-vehicle performance measures and device specific measures will be introduced. This subsection briefly mentions the uses to which

Figure 1-3
Relation Between Average Glance Frequency and Lane Crossings
(Source: Berger, Smith, & Ziedman, 1989)



subjective assessments might be put. It also presents some reservations associated with the use of secondary tasks for this project.

Appendix 1A presents a review of selected subjective assessment techniques available in the workload literature. Subjective methods may be taken during an in-cab device interaction or afterward (perhaps with the use of a memory aid such as a videotape of the transaction to prompt the driver.). Furthermore, they can be either absolute assessments or relative assessments (e.g., comparing two alternative designs, comparing two different transactions with the same device). These measures are relatively easy to obtain. However, subjective assessments are not necessarily easy to interpret as reflections of the demands incurred by use of a device. This is because subjective workload assessments sometimes dissociate with performance and the relationship between the two is not well understood (Gopher and Donchin, 1986).

One area where subjective assessments are expected to have high value as well as high validity is in the area of driver acceptance of an in-cab device. Some driver responses to an unacceptable system include (from Ramsey and Atwood, 1979):

- Dis-use: defeating, ignoring, or otherwise failing to employ the device as its designers intended;
- Mis-use: shortcutting operational difficulties by means which may undermine system integrity or intended effectiveness and induce unexpected device demands;
- Partial use: use of only a subset of the full system capabilities;
- Modification of the task: changes in (driving) task behavior to match the needs of the system.
- Frustration and Apathy: if a device's use is mandated by operating policy, drivers who experience frustration and apathy may evidence poor performance (e.g., in device interaction or driving performance) which can negatively affect efficiency of operations and highway safety. Frustration and apathy can also be manifested in acts of sabotage or damage to poorly designed equipment by frustrated drivers.

Subjective methods may best be considered a structured means to gauge how the driver who uses it feels about the device being evaluated. This can lead to valuable marketing inputs and design improvements of such devices. Driver subjective workload ratings may also lend some insights into the driver's mood or sense of confidence in working with a device. There is considerable recent empirical evidence in the literature that suggests that one's affective state can influence such things as organization in memory, risk-taking tendency, utilities of losses in gambling, and likelihood judgments for "good" and "bad" outcomes independently (see Isen, Nygren, and Ashby, 1988; Nygren and Isen, 1985; Nygren and Morera, 1987). Subjective workload measures could be explored in this project as indicators of the potential affective biases that may actually influence judgments of mental workload or changes in the dynamic decision making strategies used by drivers. Thus, subjective assessments merit consideration for inclusion in a workload assessment protocol; they provide complementary data to other data on driver resource loads and driver-vehicle performance.

Traditional dual- and multi-task tools (e.g., Damos, 1991) were created to measure workload in a relatively simple laboratory situation. Subjects perform a clearly-defined primary task. Performance on a secondary task is treated as an index of primary-task demands. (Only objective methods are herein discussed-subjective and physiological tools have other problems.) There are at least two critical assumptions behind this methodology:

- (1) Performing the secondary task does not alter primary-task performance (Kantowitz, 1974).
- (2) The subject devotes sufficient attention or capacity to the primary task.

The validity of the first assumption can be tested by incorporating appropriate single-task and other control conditions (Kantowitz, 1985), although there are many potential pitfalls associated with this methodology (e.g., Kantowitz & Weldon, 1985). In a laboratory environment it is assumed that subjects performing a dual-task are capable of following instructions (especially with feedback and an appropriate pay-off matrix) and attending first to the primary task. The likelihood is further increased by training subjects on the primary task. Performance is always scored on the primary task so that any failure of this second assumption would be revealed by a performance decrement.

This logic, while still valid for real-world and simulator tasks, depends upon the two assumptions listed above. These assumptions are less likely to be satisfied in the real world or even in a medium fidelity simulation of reality, in which case the methodology fails. The issue is even more complicated because there is no universal secondary task. Any selected secondary task can only be partially diagnostic, even if the assumptions above are valid.

There is yet another difficulty associated with the use of secondary task workload measures for in-cab device assessment. This difficulty is namely that use of the to-be-evaluated device is itself a secondary task relative to the primary driving task! Thus, introducing a workload protocol secondary task to assess the demand of the in-cab device reflects putting a third level of concurrent tasking on the driver. The confusion and interpretive difficulties associated with such an approach are easy to imagine. Secondary task methodology was originally developed from an attempt to determine the "spare capacity" of the operator. Yet for this project, how many more things the driver could do is not so important as how many things the driver must be able to do at any given instant. With a given in-cab device, a safety relevant workload protocol would ideally measure whether primary driver-vehicle performance and safety is degraded during device interaction under the demands of realistic driving conditions and driving tasks.

One possible means to productively make use of workload secondary task methods is to distinguish between unembedded and embedded tasks (see Appendix 1A) An unembedded secondary task is an extraneous task which has no natural place in the driver's normal task ensemble. On the other hand, an embedded task is defined here as one which is a naturally occurring component of the primary driving task. For example, while the driving task involves lateral and longitudinal control, legitimate components of the driving task also include object and event detection and wayfinding. Evaluation of an in-cab device might therefore be augmented in the following way:

- Assess driver performance on such tasks as object and event detection or wayfinding in a baseline condition, i.e., under normal conditions (no high technology in-cab device interaction). Examples of the object and event detection might be driver response to sudden lead vehicle braking, an object in the roadway, pedestrian stepping onto the travel lane, traffic signal changes, etc. Examples of the wayfinding might be measured by driver performance on road sign reading, correct vs. incorrect turns, lateness in merging into an exit lane, time-of-arrival to the destination, etc. These types of measures would have safety relevance, operational effectiveness relevance, or both;
- Collect the same types of measures but during device interaction while driving. Presumably protocol scenarios would synchronize at least some of the events and in-cab transactions to insure that comprehensive measurements were taken;
- Compare the baseline data with the in-cab device data to assess the extent to which the in-cab device introduces subsidiary tasks which compete with driver resources which otherwise would be devoted to driving tasks.

The details of this embedded task approach have not yet been developed. However, the difficulties associated with introducing an unembedded workload-secondary task during in-cab device interaction while driving have been introduced. As will be discussed in subsequent sections, there is value in introducing embedded tasks which are a natural part of driving into a safety relevant workload assessment protocol for in-cab devices.

APPENDIX 1 A

WORKLOAD MEASURE REVIEWS:

TABULATIONS FOR SELECTED SUBJECTIVE AND SECONDARY TASK MEASURES

Table 2.1: Modified Cooper-Harper (MCH)

Criteria for Workload Measurement Selection	Explanation
Sensitivity: MODERATELY HIGH	Lysaght et al. (1989) rated this "HIGH". But, Hill et al. (1992) have shown that other subjective measures appear to have greater sensitivities (based upon larger factor validities).
Diagnosticity: LOW	Rating from Lysaght et al. (1989) as MCH is a global rating.
Timing of Measurement: VARIED	MCH has been collected at regular intervals during a task when this was not invasive, but often must be taken during lulls between tasks or after task completion.
Intrusiveness: VARIED	MCH and other subjective and secondary task measures might compromise safety when taken at closely spaced regular intervals (vs. intermittent during lulls or after task completion).
Invasiveness: LOW	Operators frequently "like" to give their subjective ratings. TLX, OW, and SWAT are more liked more as reported by Hill et al. (1992).
Operator Acceptance: HIGH	This will be generally high when intrusiveness is kept low, because invasiveness is low.
Reliability: MODERATELY HIGH	Rating supported by factor validity findings of Hill et al. (1992).
Validity: MODERATELY HIGH	Hill et al. (1992) support this rating as TLX and other subjective measures have somewhat higher factor validities.
Relevance for Safety: MODERATELY HIGH	Predicated upon (i) the workload-performance relationship (e.g., Lysaght et al. 1989) and (ii) moderately high sensitivity.
Implementation: HIGH	MCH and other subjective ratings can be recorded by a variety of non instrumentation intensive electronic and paper means.
Cost: LOW	Rating based on Lysaght et al. (1989).
Ease of Data Reduction/Analysis: LOW	Rating derived from Lysaght et al. (1989).
Flexibility: MODERATE	Hill et al. (1992) found that some modestly verbal users had difficulty using the flow chart to guide their MCH ratings.
Theoretical Basis: HIGH	Predicated upon (i) the hypothetical workload-performance relationship (e.g., Lysaght et al., 1989) and (ii) the argument that workload is best accessed subjectively (Sheridan 1980, p.1).

Table 2.2: Overall Workload (OW)

Criteria for Workload Measurement Selection	Explanation
Sensitivity: HIGH	Hill et al. (1992) have shown OW to have greater sensitivity than other subjective measures that Lysaght et al (1989) generally rate as 'HIGH'.
Diagnosticity: LOW	Rating follows Lysaght et al. (1989) for global ratings.
Timing of Measurement: VARIED	Can be taken more regularly than other subjective measures as it is quickest to obtain (Hill et al., 1992). But often must be taken during lulls or after task completion to avoid intrusiveness.
Intrusiveness: VARIED	OW and other subjective and secondary task measures might compromise safety when taken at closely spaced regular intervals (vs. intermittent during lulls or after task completion).
Invasiveness: LOW	Operators frequently 'like' to give their subjective ratings. OW is more liked than MCH and SWAT as reported by Hill et al. (1992).
Operator Acceptance: HIGH	This will be generally high when intrusiveness is kept low, because invasiveness is low.
Reliability: HIGH	Rating supported by factor validity findings of Hill et al. (1992).
Validity: HIGH	Hill et al. (1992) support this rating (only TLX generally has a consistently higher factor validity).
Relevance for Safety: HIGH	Predicated upon (i) the hypothetical workload-performance relationship (e.g., Lysaght et al. 1989) and (ii) high sensitivity as indicated by Hill et al. (1992).
Implementation: HIGH	OW and other subjective ratings can be recorded by a variety of non instrumentation intensive electronic and paper means.
cost: LOW	Rating is consistent with assessments of Lysaght et al. (1989).
Ease of Data Reduction/Analysis: HIGH	Generally easier than other subjective scales so rating by Lysaght et al. (1989).
Flexibility: HIGH	Hill et al. (1992) found that users had least difficulty making OW ratings.
Theoretical Basis: HIGH	Predicated upon (i) the hypothetical workload-performance relationship (e.g., Lysaght et al., 1989) and (ii) the argument that workload is best accessed subjectively (Sheridan 1980, P.11).

Table 2.3: Task load Index (TLX) and Raw Variant (RTLX)

Criteria for Workload Measurement Selection	Explanation
Sensitivity: HIGH	Hill et al. (1992) have shown that TLX appears to have greater sensitivity than most other subjective measures that Lysaght et al (1989) generally rate as HIGH. RTLX is highly correlated (r=.97) with TLX (Byers et al., 1989).
Diagnosticity: HIGH	Rating follows Lysaght et al. (1989) who note that it accesses six scales.
Timing of Measurement: VARIED	Can be taken less regularly than most other subjective measures as it takes longer to obtain (Hill et al., 1992). Often must be taken during task lulls (or afterward) to avoid intrusiveness.
Intrusiveness: VARIED	These other subjective and secondary task measures might compromise safety when taken at closely spaced regular intervals (vs. intermittent during lulls or after task completion).
Invasiveness: LOW	Operators frequently 'like' to give their subjective ratings. TLX was more liked than OW, MCH and SWAT as reported by Hill et al. 1992.
Operator Acceptance: HIGH	This will be generally high when intrusiveness is kept low, because invasiveness is low. Operators consistently rated this as 'Best description of workload' (Hill et al., 1992).
Reliability: HIGH	Rating supported by high TLX factor validity findings of Hill et al. (1992).
Validity: HIGH	Hill et al. (1992) support this rating with TLX generally has having consistently highest factor validity. RTLX was highly correlated with TLX, supporting this rating for it (Byers et al., 1989).
Relevance for Safety: HIGH	Predicated upon (i) the hypothetical workload-performance relationship (e.g., Lysaght et 1989) (ii) high sensitivity as indicated by Hill et al. 1992, and (iii) high diagnosticity (see above).
Implementation: HIGH	TLX and other subjective ratings can be recorded by a variety of non instrumentation intensive electronic and paper means.
cost: LOW	Rating is consistent with assessments of Lysaght et al. (1989). RTLX will be less costly than TLX as it avoids a paired comparison step (Byers et al., 1989).
Ease of Data Reduction/Analysis: HIGH	TLX rating by Lysaght et al. (1989). RTLX will be easier than TLX as it avoids handling paired comparisons weights (Byers et al., 1989).
Flexibility: HIGH	Hill et al. (1992) found that users found only OW ratings easier.
Theoretical Basis: HIGH	Predicated upon (i) the hypothetical workload-performance relationship (e.g., Lysaght et al., 1989) and (ii) the argument that workload is best accessed subjectively (Sheridan 1980, 0.1). TLX and RTLX are representatives of a six dimensional model of operator workload.

Table 2.4: Subjective Workload Technique (SWAT)

Criteria for Workload Measurement Selection	Explanation
Sensitivity: MODERATELY HIGH	Lysaght et al. (1989) rated this "HIGH". But, Hill et al. (1992) have shown that other subjective measures appear to have greater sensitivities (based upon larger factor validities).
Diagnosticity: HIGH	Rating from Lysaght et al. (1989) as SWAT accesses time-load, mental effort load, and psychological stress load.
Timing of Measurement: VARIED	SWAT has been collected at regular intervals during a task when this was not invasive, but often must be taken during lulls between tasks or after task completion (c.f., Gidcumb, 1985).
Intrusiveness: VARIED	SWAT and other subjective and secondary task measures might compromise safety when taken at closely spaced regular intervals (vs. intermittent during lulls or after task completion).
Invasiveness: LOW	Operators frequently 'like' to give their subjective ratings. SWAT less liked than OW and TLX as reported by Hill et al. (1992).
Operator Acceptance: HIGH	This will be generally high when intrusiveness is kept low, because invasiveness is low.,
Reliability: MODERATELY HIGH	Rating supported by factor validity findings of Hill et al. (1992).
Validity: MODERATELY HIGH	Hill et al. (1992) support this rating as TLX and other subjective measures often have somewhat higher factor validities.
Relevance for Safety: MODERATELY HIGH	Predicated upon (i) the workload-performance relationship (e.g., Lysaght et al. 1989) and (ii) moderately high sensitivity.
Implementation: HIGH	SWAT and other subjective ratings can be recorded by a variety of non instrumentation intensive electronic and paper means.
cost: LOW	Rating based on Lysaght et al. (1989).
Ease of Data Reduction/Analysis: LOW	Rating derived from Lysaght et al. (1989). Scaling of ratings adds only marginally when computer-based analyses are used.
Flexibility: MODERATE	Hill et al. (1992) found that many low verbal users had difficulty producing satisfactory card sorts (a SWAT scaling step). Lysaght et al. (1989) report earlier anecdotal evidence.
Theoretical Basis: HIGH	Predicated upon (i) the hypothetical workload-performance relationship (e.g., Lysaght et al., 1989) and (ii) the argument that workload is best accessed subjectively (Sheridan 1980, p. 1). SWAT is based on a formal measurement model (Nygren, 1991).

Table 3.1: Embedded Secondary Task (EST)

Criteria for Workload Measurement Selection	Explanation
Sensitivity: GENERALLY HIGH (BUT CAN VARY).	Lysaght et al. (1989) rated this as "HIGH". But, ESTs could prove insensitive measures of capacity when statistically unreliable (when spare capacity is focus, e.g., Kahneman, 1973). Assessment of EST impacts on the primary task could be compromised by the primary's unreliability (when of alternate interest, e.g., Gopher & Donchin, 1986).
Diagnosticity: LOW TO MODERATE	Rating is in contrast with "moderate to high" given by Lysaght et al. (1989) -- who focused on use of several ESTs and their components to establish diagnosticity.
Timing of Measurement: HIGH	Can be taken more regularly than other secondary or subjective measures as it is part of the task (Lysaght et al., 1989 p. 128).
Intrusiveness: LOW-MODERATE	Safety is less a question when the EST is part of the larger task (driving). ESTs can interfere with primary tasks (e.g., Shingledecker & Crabtree, 1982).
Invasiveness: LOW	Operators see ESTs as part of the job (Lysaght et al., 1989 p.128).
Operator Acceptance: HIGH	This will be generally high when intrusiveness is relatively low as invasiveness is also low (cf., Lysaght et al. (1989).
Reliability: GENERALLY HIGH (BUT COULD BE VARIED)	See sensitivity discussion above.
Validity: GENERALLY HIGH (BUT COULD BE VARIED)	Lysaght et al.(1989) rate this as high when used in the context of its purpose. But, predictive validity would ultimately depend on the reliability of the measures of interest.
Relevance for Safety: GENERALLY HIGH	Predicated upon (i) the hypothetical workload-performance relationship (e.g., Lysaght et al. 1989), (ii) integral nature of the EST (to the driving task).
Implementation: MODERATE TO HIGH	EST procedures are the most accommodated of the class of secondary tasks as they are most integral to the primary task. Instrumentation should also be most accommodated for the same reasons, but may be may be moderately challenging to implement (vs. "HIGH" rating by Lysaght et al., 1989).
Cost: LOW TO MODERATE	Rating differs from assessment of Lysaght et al. (1989) as they did not consider the sometimes moderate challenge of instrumentation.
Ease of Data Reduction/Analysis: MODERATE TO HIGH	Rating is based upon the possible complexities of differential tradeoffs between the primary and secondary tasks (analogous to speed-accuracy tradeoffs).
Flexibility: HIGH	The EST is most flexible when it is built on a device to be evaluated (e.g., where the effects of warnings on aspects of other primary driving tasks are of concern).
Theoretical Basis: HIGH	Predicated upon (Lysaght et al., 1989): (i) the hypothetical workload-performance relationship and (ii) arguments that ESTs provide a direct measure of relevant reserve capacities.

Table 3.2: Choice Reaction Time (CRT)

Criteria for Workload Measurement Selection	Explanation
Sensitivity: MODERATE	Lysaght et al. (1989) rated this as 'moderate*' based upon Bortolussi et al. (1986; 1987), who did not use cast in slope-intercept form. CRT (1 and 4 choices) exhibit moderately high reliability efficiencies (Bittner et al., 1986), supporting this rating.
Diagnosticity: LOW	Rating is in contrast with 'moderate' rating of Lysaght et al. (1989), who based their's on diagnostic use of the slope and intercept measures (following Wickens et al , 1986). But, questions of the statistical reliability of CRT (and other) slope measures leads to the modified rating (cf., Carter et al., 1988, Bittner et al., 1986).
Timing of Measurement: VARIED	CRT has been collected at regular intervals during a task when this did not interfere with safety but, as Schiflett 11980) found with the Sternberg item recognition task, may be abandoned if this is the case.
Intrusiveness: VARIED	Safety not a question when CRT is used in simulator. But may interfere with primary task safety in operational (e.g. experienced by Schiflett, 1980 with Sternberg).
Invasiveness: VARIED	Operators may see CRT as part of the job in a laboratory setting, but feel encroached upon in operational setting (per informal comments following Schiflett's analogous task).
Operator Acceptance: VARIED	Lysaght et al., rate this as 'MODERATE'. But this may be either "HIGH" when intrusiveness and invasiveness are low (lab settings), or 'LOW' when either or both are low.
Reliability: MODERATELY HIGH	Biiner et al. (1986) support this rating.
Validity: MODERATE	Supported by historical definition (Donders, 1868) but instability of slope scores 1986) under extended practice is troublesome (ca., Teichner and Krebs, 1974).
Relevance for Safety: MODERATE TO HIGH	Predicated upon (i) the hypothetical workload-performance relationship (e.g., Lysaght et al. 1989), and (ii) application to access remaining capacity.
Implementation: MODERATE	CRT procedures and instrumentation should be accommodated in most settings, but could be troublesome in some high display environments.
Cost: MODERATE	Rating consistent with assessment of Lysaght et al. (1989).
Ease of Data Reduction/Analysis: MODERATE TO HIGH	Rating is based upon the possible complexities of differential tradeoffs between the CRT and primary (analogous to speed-accuracy tradeoffs).
Flexibility: MODERATE	CRT may be troublesome to use an on-road studies.
Theoretical Basis: MODERATELY HIGH	Predicated upon (Lysaght et al., 1989): (i) the hypothetical workload-performance relationship, and (ii) arguments that the CRT would provide a measure of reserve capacities (but not as relevantly as with ESTs).

Table 3.3: Sternberg Memory Task (SMC)

Criteria for Workload Measurement Selection	Explanation
Sensitivity: MODERATE	Lysaght et al. (1989) rated this as 'moderate' based largely on the review of Wickens et al. (1988). SCM 1, 2, and 4 item tasks exhibit moderately high reliability efficiencies (Bittner et al., 1986), supporting this rating for separate measures.
Diagnosticity: LOW	Rating is in contrast with 'moderate' rating of Lysaght et al. (1989), who based their's on diagnostic use of the slope and intercept measures (cf., Wickens et al , 1986). But, questions of the statistical reliability of SCM (and other) slope measures leads to the modified rating (cf., Carter et al., 1986, Bittner et al., 1986).
Timing of Measurement: VARIED	SMC has been collected at regular intervals during a task when this did not interfere with safety but, as Schiflett (1980) has found, may be abandoned if this is the case (Wickens et al. 1986 also note problem of operators "shunning* task in high workload settings).
Intrusiveness: VARIED	'Safety not a question when SMC is used in simulator. But may interfere with primary task safety in operational settings (cf., Schiflett, 1980).
Invasiveness: VARIED	Operators may see SCM as part of the job in a lab setting, but feel encroached upon in operational setting (per informal operator comments following Schiflett, 1980).
Operator Acceptance: VARIED	Lysaght et al. (1989) rate SCM as 'MODERATE". But this may be either 'HIGH" when intrusiveness and invasiveness are low gab settings), or 'LOW' when either or both are low.
Reliability: MODERATELY HIGH	Bittner et al. (1986) support this rating for individual item scores, but not either for slopes (paralleling Carter et al., 1986) or intercepts (not discussed by Carter et al.).
Validity: MODERATE	Supported by historical definition (Sternberg, 1966) but flattening and instability of slope scores with extended practice is troublesome (Carter, et al. 1986).
Relevance for Safety: MODERATE TO HIGH	Predicated upon (i) the hypothetical workload-performance relationship (e.g., Lysaght et al. 1989), and (ii) application to access remaining capacity.
Implementation: MODERATE	SCM procedures and instrumentation should be accommodated in most settings, but could be troublesome in some high display environments where space is limited.
Cost: MODERATE	Rating consistent with assessment of Lysaght et al. 11989).
Ease of Data Reduction/Analysis: MODERATE TO HIGH	Rating is based upon the possible complexities of differential tradeoffs between the SCM and primary tasks (analogous to speed-accuracy tradeoffs).
Flexibility: MODERATE	SCM may be troublesome to use an on-road studies.
Theoretical Basis: MODERATELY HIGH	Predicated upon (Lysaght et al., 1989): (i) the hypothetical workload-performance relationship, and (ii) arguments that the SCM would provide a measure of reserve capacities (but not as relevantly as with ESTs).

Table 3.4: Time Estimation Task (TET)

Criteria for Workload Measurement Selection	Explanation
Sensitivity: MODERATE (BUT CAN VARY)	Lysaght et al. (1989) rated this as 'MODERATE' on combined results of Wierwille et al. (1985) and Bortolussi et al. (1986). But, Bittner et al. (1986) summarize results that indicate cross-day instability that can reduce sensitivity in cross-day studies.
Diagnosticity: LOW	Rating is in contrast with 'MODERATE' rating of Lysaght et al. (1989) who did not consider measures limited content.
Timing of Measurement: VARIED	TET has been collected at regular intervals during a task when this did not interfere with safety but, similar to the Sternberg experiences of Schiflett (1980) and Wickens et al. (1986), may be 'shunned' during respectively hazardous or high workload conditions.
Intrusiveness: VARIED	Safety not a question when TET is used in simulator. But may interfere with primary task safety in operational settings (as indicated above).
Invasiveness: VARIED	Operators may see TET as part of the job in a lab setting, but may feel encroached upon in operational setting (albeit perhaps less than with other secondary tasks as indirectly suggested by Hart, 1986).
Operator Acceptance: VARIED	This may be relatively higher than all but embedded secondary tasks. But the potential exists to be either 'HIGH', when intrusiveness and invasiveness are low (lab settings), or 'LOW' when either or both are high.
Reliability: MODERATE (BUT CAN VARY)	Bittner et al. (1986) support this rating based upon modest reliabilities compounded by cross day instabilities.
Validity: MODERATE	Supported by historical definition (cf., Vroon, 1976) but cross-day instability is troublesome.
Relevance for Safety: MODERATE TO HIGH	Predicated upon (i) the hypothetical workload-performance relationship (e.g., Lysaght et 1989), and (ii) application to access remaining capacity.
Implementation: HIGH	TET's simple procedure and low instrumentation requirements will result in it all being accommodated in almost all settings.
Cost: MODERATE	Rating consistent with assessment of Lysaght et al. (1989).
Ease of Data Reduction/Analysis: MODERATE TO HIGH	Rating is based upon the possible complexities of differential tradeoffs between the TET and primary task (analogous to speed-accuracy tradeoffs).
Flexibility: MODERATE	TET may be troublesome to use in on-road studies.
Theoretical Basis: MODERATELY HIGH	Predicated upon (Lysaght et al., 1989): (i) the hypothetical workload-performance relationship, and (ii) arguments that the TET would provide a measure of reserve capacities (but not as relevantly as with ESTs).

References

- Adams, M. J., Tenny, Y.J., & Pew, R.W. (1991, December). State of the Art Report, Strategic work lead and the cognitive management of advanced multi-task systems (Reports No. CSERIAC 91-6). Wright-Patterson AFB, OH: Crew Systems Ergonomics Information Analysis Center (CSERIAC).
- Bittner, Jr., A.C., Carter, R.C., Kennedy, R.S., Harbison, M.M and Krause, M. (1986). Performance Evaluation Tests for Environmental Research (PETER): Evaluation of 114 measures. *Perceptual and Motor Skills*, 63, pp. 683-708.
- Bortolussi, M.R., Hart, S.G.. & Shively, R.J. (1987). Measuring moment-to-moment pilot workload using synchronous presentations of secondary tasks in a motion-base trainer. In *Processings of the Fourth Symposium on Aviation Psychology* Columbus, OH: Ohio State University.
- Bortolussi, M.R., Kantowitz, B.H., & Hart, S.G. (1986). Measuring pilot workload in a motion base trainer. *Applied Ergonomics* 17, pp. 278-283.
- Burger, W.J., Smith, R.J., & Ziedman, K. (1989). Supplemental electronic in-cab truck displays : An inventory of devices and approaches to their evaluation (Report No. DOT HS 807 411). Santa Monica: Vector Enterprises, Inc.
- Byers, J.C., Bittner, Jr., A.C., & Hills, S.G.(1989). Traditional and Raw Task Load Index (TLX) correlation: Are paired comparisons necessary? In *Advance in Industrial Ergonomics and Safety I*. London: Taylor & Francis, pp.481-485.
- Carter, R.C., Krause, M. & Harbeson, M.M. (1986). Beware the reliability of slope scores for individuals. *Human Factors*, 28, pp. 673-684.
- Clarke, R.M., Leasure, Jr., W.A., Radlinski, R.W., & Smith, M. (1987, March). Heavy truck safety study (Report No. DOT HS 807 109). Washington, DC: National Highway Safety Traffic Administration.
- Demos, D.L. (1991). Multiple-task performance Bristol, PA: Taylor & Francis, Inc., 363-386.
- Donders, F.C., (1868). Die Schnelligkeit psychischer Processe. *Archiv for Anatomie und Pysiologie und Wissenschaftliche Medizin*, pp. 657-681. [Also, donders, F.C. (1969). On the speed of mental processes. (Translated by W.G. Koster) *Aceta Psychologica*, 30, 412-431.
- Gibson, J.J. (1966). *The senses considered as perceptual systems*. New York: Cornell University.
- Gidoumb, C. (1985). Survey of SWAT use in flight test (BCM/A-85-0630-TR). Albuquerque, NM: BDM Corporation.
- Gopher, D., & Donchin, E. (1986). Workload: An examination of the concept. In K.R. Boff, L. Kaufman & J. Thomas (Eds.), *Handbook of perception and human performance: Vol II Cognitive Processes and Performance* (pp. 41-49). New York: John Wiley and Sons.
- Hart, S.G. (1986) Theory and measurement of human workload. In J. Zeider (Ed.), *Human Productivity Enhancemant Vol 1* pp. 396-455. New York: Praeger.
- Hart, S.G. (1983). Workload : A new perspective. Unpublished working paper cited in Admans, Tenny, and Pew (1991).
- Hart, S.G. (1989). Crew workload management strategies: A critical factor in system performance. *Proceedings of the Fifth International Symposium on Aviation Psychology (Vol. I)*. 22-27.
- Hart, S.G. & Wichens, C.D. (1990) Workload assessment and prediction. In H.R. Booher (Ed.). *Manprint: an approach to systems integration* (pp. 257-296). New York: Van Nostrand Reinhold.
- Hendricks, D., Allen, J., Tijerina, L., Knipling, R., & Wilson, C. (1992). VNTSC IVHS program topical report #1; rear end crashes. Columbus, OH : Battelle.
- Hill, S.G., Iavecchia, H.P., Byers, J.C., Bittner Jr., A.C., Zaklad, A.L., and Christ, R.E. (1992). Comparison of four subjective workload rating scales. *Human Factors*, in press.
- Isen, A.M. Nyren, T.E, & Ashby, F.G. (1988). The influence of positive affect on the subjective utility of gains and losses : It's not worth the risk. *Journal of Personality and Social Psychology*.
- Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, N.J.: Prentice-Hall.

- Kantowitz, B.H.,(1974). Double stimulation. In B.H. Kantowitz (Ed>) Human Information Processing Tutorials in Performing and Cognition. Hillsdale, NJ: Erlbaum.
- Kantowitz, B.H., & Weldon, M.S. (1985). On scaling performance operating characteristics: Caveat emptor. *Human Factors*, 27 531-547. (Reprinted in M. Venturino (Ed.) Selected Readings in Human Factors, Scanta Monica, CA: Human Factors Society, 1890).
- Kiger, S., Rockwell, T.H., Niswonger, S., Tijerina, L., Myers, L.B., & Nygren, T (1992). Report on NHSTA heavy vehicle driver workload assessment task 3 interim report: task analysis data collection. Columbus, OH: Battelle.
- Lysaght, R.J., (1989). Operator Workload: Comprehensive Review and Evaluation of Operator Workload. Methodologies (ARI TR 851), alexandria, VA: Army Research Institute for the Behavioral Sciences.
- McKnight, J.A., McKnight, A.S. (1991). The Effect of Cellular Phone use upon driver attention. Landover, MD: National Public Services Research Institute.
- Neisser, U. (1976). Cognition and reality, San Francisco: Freeman and Co.
- Nygren, T.E., & Isen, A. M. (1985, May). Effect of positive affect on judgments of likelihood of events and on gambling behavior. Paper presented at the meeting of the Midwestern Psychological Association, Chicago, IL.
- Nygren, T.E., & Morera, O. (198, July). Davidson, Suppes, and Siegel revisited: Evidence for a dual bilinear model. Paper presented at the Society for Mathematical Psychology Meeting, Northwestern University.
- Nygren, T.E.,(1991) Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload. *Human Factors*, 33, 17-33.
- O'Donnell, R.D., & Eggemeir, F.T (1986). Workload assessment methodology. InK.R. Boff, L. Kaufman, & J.P. Thomas (Eds.) Handbook perception and human performance. Vol II cognitive processes and performance (pp.42-48). John Wiley.
- Perel, M. (1976). Analyzing the role of driver/vehicle incompatibilities in accident causation using police reports. (Report No. DOT HS 801-858). Washington, DC: U.S. Department of transportation.
- Ramsey, H.R., & Atwood, M.E, (1979, September). Human Factors in Computer Systems: A review of the literature Englewood, CO: Science Applications Inc.
- Rockwell, T.H. (1988). Spare visual capacity in driving revisited: New empirical results for an old idea. In A.G, Gale, et al. (Eds.) Vision in Vehicles II Proceedings of the Second International Conference on Vision in Vehicles., UK 14-17 Sept 1997 (pp.317-324). Amsterdam, Netherlands: Elsevier.
- Sohiflett (1980). Evaluation of a pilot workload assessment device to test alternate display formats and control handling qualities (Report SY-33R-80). Patuxent River, MD: Naval Air Test Center.
- Sheridan, T.B. (1980). Mental workload –What is it? Why bother with it? *Human Factors Society Bulletin* 23, pp. 1-2.
- Sheridan, T.B. (1991). Human Factors of driver vehicle interactions in the IVHS environment (Report No. DOT HS 807-837). Cambridge, MA: Center for Transportation Studies.
- Shingledecker, C.A., & Crabtrees, M.S. (1982). Subsidiary radio communications tasks for workload assessment in R&D simulation: ask sensitivity evaluation (AFAMRI-TR-82-57). Wright-Patterson AFB, OH: Aerospace Medical Research Laboratory.
- Sternberg, S.(1966). High speed scanning in human memory. *Science*, 153, pp.652-654.
- Sussman, E.D., Bishop, H. Madnick, B., & Walter, R. (1985). Driver inattention and highway safety. *Transportation Research Record*, 40-48.
- Teiohner, W.H., & Krebs, M.J. (1974). Laws of vidual choice reaction time. *Psychological Review* 81 pp.75-98.
- Treat, J.R., Tumbas, N.S., Mcdonald, S.T., Shiner, D., Hume, R.D., Mayer, R.E., Stansifer, R.L., & Castellan, N.J. (1977< March). Tri-level study the cause of traffic accidents (report No. DOT-HS-0340305035077). Indiana University.

- Tulga, M.K., & Sheridan, T.B. Dynamic decisions and workload in multitask supervisory control *IEEE Transactions on Systems, Man and Cybernetics*, 1980, SMC-10, 217-232.
- Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and Biases. *Science*, 185, 124-131.
- Tversky, A., & Kahneman, D., (1984). The Framing of decisions and the psychology of choice. *Science*, 211, 453-458.
- Vroom, P.A. (1976). Sequential estimations of time. *Acta Psychologica* 40, pp. 457-487.
- Wickens, C.D., Hyman, F. Dellinger, J., Taylor, H. & Meador, M. (1988). The Sternburg memory search task as an index of pilot workload, *Ergonomics*, 28, pp.1371-1383.
- Wickens, C.D., & Flach, J. (1988). Human Information processing. In e. Wiener & D. Nagel(Eds.), *Human factors in aviation* (pp.111-155). New York: Academic Press.
- Wickens, C.D. (1992). *Engineering psychology and human performance* (2nd ed.) New York: Harper Collins.
- Wierwille, W.W., Casali, J.G., Connor, S.A., Rahimi, M. (1985). Evaluation of the sensitivity and intrusion of mental workload estimation techniques. In W. Rorer (Ed.). *Advances in Man-Machine System Research* Vol. 2 (pp. 510 1271. Greenwich, CT.: J.A.I Press.
- Wierwille, W.W. (1992). Visual and manual demands of in-car controls and displays. In J.B. Peacock & W. Karwowski (Eds.) *Automotive ergonomics: human factors in the design and use of the automobile*. London: Taylor & Francis. (In Press)
- Zeigarnik. B.V. (1965). *The pathology of thinking*. New York: Plenum.

2.0 A DRIVER RESOURCE ALLOCATION MODEL OF IN-CAB DEVICE WORKLOAD

Introduction

The problem to be solved in an inferential approach to safety relevance is to develop the relationship between workload and safety (if it exists). The difficulty with developing such an approach is that the highway transportation system of the U.S. (as well as elsewhere) is usually a forgiving one. There are many examples of this. A few are indicated:

- Interstates and improved U.S. highways are much wider than cars and usually wider than heavy vehicles (HVs). Thus, some weaving or lane deviation can be tolerated.
- Other drivers can evade or otherwise compensate for a problem created by a specific driver by using the direct visibility and mirror systems that are currently available.
- Vehicles are equipped with safety devices such as horns, stop-lights, center-high-mounted tail lights and marker-lights which increase the likelihood of a driver detecting an impending conflict with another vehicle. Presumably, Intelligent Vehicle Highway System (IVHS) crash avoidance countermeasures (e.g., headway detection and warning systems, near object detection systems) will increase this hazard detection potential further.
- Roadway shoulders are designed so that, if a driver does deviate from the roadway, the driver can usually recover control.

Because of this forgiving design, there are probably many workload-induced problem situations which occur that do not result in accidents. For example, a driver who diverts attention to a cellular phone may not maintain adequate visual scan of the forward scene or may not maintain adequate lateral-directional control of the vehicle. Nevertheless, the occurrence of an accident remains unlikely because of the forgiving nature of the system. However, the margin of safety is reduced in such situations. If the appropriate accident records were available, they would probably reflect greater accident frequency in the accident population at large. Thus, the inferential approach does not link workload measures with accidents but rather with increased hazard exposure (the increased likelihood or severity of a crash).

Because the highway system is fault-tolerant, it is difficult to develop safety-relevant criteria for workload. The highway system absorbs driver errors so that workload-related difficulties do not necessarily map into a one-to-one relationship with accident statistics, particularly for small samples. This suggests that some other method must be developed for generating safety-relevant criteria. A new approach must necessarily be based on assumptions about the relationship between safety and the use of driver resources. These assumptions would probably be considered acceptable if they were reasonably conceived.

The resource allocation approach is a new method of developing safety-relevant criteria. This concept specifies the general categories of resources that HV drivers require, excluding in-vehicle tasks, and then determines or assesses the degree to which the in-vehicle tasks-of-interest encroach on the driving task. In other words, the driver performs both the primary task of driving the vehicle and a

secondary task that is the specific in-cab task being analyzed. The in-cab task uses some of the driver resources needed to perform the primary task, and thereby reduces the margin of safety but only if excessive resources are used.

The driver is represented by a pool of resources. When the driver performs only the primary task of driving, these resources can be dedicated to the driving task. On the other hand, when the driver performs an in-cab task, some of the driver's resources must be diverted to the in-cab task, thus leaving a less-than-full complement for driving. It can be assumed that, the greater the incursion into driver resources caused by the in-cab task, the greater the risk associated with performing the task. Risk can be minimized and safety can be maximized by reducing incursions into driver resources caused by in-vehicle tasks. Through a process of evaluating device demands for driver resources, improvements to device design can be made in a more informed manner.

Review of Driver Resources

The resources that the driver brings to the driving task are relatively easily stated and can be reasonably well specified. This section reviews these resources and provides a ranking in terms of importance.

Visual Resource. It has been estimated that about 90 percent of all information that a driver uses for the primary task of driving is obtained visually (Rockwell, 1972). Auditory and proprioceptive/tactual cues play a small role, but are not essential for driving. Deaf drivers generally have little difficulty in performing the driving task. Similarly, ordii drivers are able to drive fixed-base simulators that have physical motion cues deleted. However, it is impossible to drive a vehicle with eyes closed because the driver performs the loop closure operations, hazard detection, and navigation using visual inputs. Thus, visual loading caused by in-vehicle tasks must be given careful consideration.

There are, in fact, two forms of visual resources: foveal and peripheral. Foveal vision, which provides high resolution capabilities, allows the driver to gather detailed information. Peripheral vision is important because it provides motion impressions (flow cues) and can be used to detect potential hazards, particularly those that are in motion relative to the driver.

It has been demonstrated in different settings that, as the load imposed by the foveally fixated task increases, the functional field of view shrinks (Ma&worth, 1976). Therefore, although the driver must often foveally fixate towards a source at the edge of the peripheral field to attend to it, the likelihood that this source will attract attention is reduced the more the driver uses foveal attention (Rockwell et. al., 1977).

Normal drivers have only one foveal resource. Human eyes work as a system and cannot be decoupled from one another and used separately (for example, to observe the forward view and an in-vehicle display at the same time). The only way that the driver can gather detailed visual information from sources at different positions is to move the foveal resource about in time, that is, to sample or time-share. This fact has profound ramifications for the design of in-vehicle displays. If the displays require that the driver extract detailed information, then they will create a demand on the driver's foveal visual resource.

One means to reduce the distance between the road scene and instrument panel visual displays is to project that same information via a Head-Up Display (HUD). A HUD is a virtual-image display in which the symbology typically appears located at some distance beyond the cockpit, cab, or workstation (Weintraub and Ensing, 1992). HUD applications for cars or trucks are currently being developed. From the standpoint of driver workload, there are at least two major issues surrounding HUD applications which deserve mention here.

One potential problem with Head-Up Displays is known as cognitive capture. Studies of dynamic visual attention have indicated that selection between two spatially overlapping objects (e.g., HUD image vs. background road scene) is often complete. For example, Neisser and Becklen (1975) simultaneously presented films of a handgame and a ballgame on the same screen. Subjects instructed to count the number of exchanges in, say the handgame showed very little awareness of the irrelevant ball game, or of unexpected events (e.g., a girl with a parasol promenading behind the ballgame players). That this is not just a laboratory artifact comes from an analogous study by Fischer, Haines, and Price (1980) of a HUD in a flight simulator. Pilots making a landing approach by means of the HUD were slow to detect an object (e.g., simulated aircraft which appeared to taxi onto the same runway for takeoff) because they did not attend to the object, even though they "saw" it. Given the much greater visual demand of driving vs. flying, any non-trivial application of HUD technology to cars and trucks deserves significant evaluation. It is possible that cognitive capture occurs only with complex or compelling displays. If so, this is especially of concern for HUD applications which might project visually complex or compelling images such as terrain maps or frequently moving gauges or icons. Furthermore, edges and displayed images on LCD displays and HUDS may obscure vision though they are otherwise transparent (Chong and Triggs, 1989; Iavecchia, Iavecchia, and Roscoe, 1988). Thus, HUD technology may induce perceptual loads of its own.

A second potential source of driver load with HUDs is primarily visual rather than perceptual in nature. Weintraub and Ensing (1992) discuss at length HUD symbology parameters and recommendations for symbol contrast, symbol size, font design, spacing, resolution, and the like. It is important to insure that HUD symbols or images do not provide a visual mask to critical road scene information. An example might be a pedestrian stepping onto the driving lane who is not noticed early on by the driver because the HUD symbology effectively obscured that portion of the road scene. Clearly, there are many issues associated with this aspect of HUD evaluation as well.

The technical challenges to evaluating HUD displays are formidable. It is quite possible that attentional affects might be uncovered using more sophisticated eye movement techniques than visual allocation. Examples of possibly sensitive workload measures might be vergence eye movements or reduced visual scanning. There are substantial difficulties associated with collecting such measures in a heavy truck on the road. Thus, such an assessment will be of lower priority than visual allocation in the face of time or budget constraints. This is in keeping with the greater applicability of other assessment procedures to be developed in this project and the currently low incidence of HUD applications in heavy vehicle applications. However, as HUD use grows, the importance of developing appropriate measures of the loads imposed on the driver grow as well.

Given the importance of visual allocation to safety and interface evaluation, it is not surprising that a good deal of work has been done on assessing the visual demands of conventional and advanced technology in-vehicle devices. Research has shown that the driver ordinarily employs a visual sampling strategy to complete in-vehicle tasks. The driver samples between the forward view and the in-vehicle device as shown in Figure 2-1 (Wierwille, 1987; Wienville, 1992; Kiger, et. al., 1992).

in-vehicle device as shown in Figure 2-1 (Wierwille, 1987; Wierwille, 1992; Kiger, et. al., 1992). The driver samples the in-cab task, returns to the driving scene, samples the in-cab task once again, returns to the driving scene, and so on until the in-cab task is completed.

Senders, Kristofferson, Levison, Dietrich, and Ward (1967) showed that, on a closed section of new highway, a driver could drive a vehicle with samples only, that is, using a helmet with a visor that periodically occluded the forward view.

They concluded that:

- A driver can drive with samples of the forward view.
- Between samples, the driver becomes uncertain.
- When the uncertainty reaches a threshold, the driver needs new samples of the forward view.

Of course, it can be assumed (and later work has indeed shown) that more closely-spaced samples are necessary when roadways are sharply curved or there is traffic. Time-to-Line Crossing (TLC) is a measure developed by Godthelp, Milgram, and Blaauw (1984) that is closely related to the Senders, et al. work. TLC assumes that uncertainty is directly related to the amount of time it would take for the vehicle to cross the center line or fog line assuming the steering wheel is held constant. TLC does not yet account for traffic, emergencies, crosswinds, or transverse roadway disturbances (Tijerina, et al., 1991) so the Senders et al. concept might be more robust. This work shows that in-vehicle devices must be designed so that the driver can get enough samples per unit time of the forward view to insure safety. In other words, the length and number of samples per unit time that a driver must spend on an in-vehicle device cannot be excessive.

Research characterizing the nature of driver visual sampling began quite recently. One of the earliest studies that examined sampling of in-car displays and controls was by Mourant, Moussa-Hamouda, and Howard (1977) (also Mourant, Herman, and Moussa-Hamouda, 1980). Their study compared five different types of stalk controls and three different types of panel controls. The results of the study indicated that the frequency of direct looks increased as the reach distance increased, and that direct looks were slightly longer for stalk controls than for panel controls.

In pioneering studies, Rockwell (1988) examined the in-car glance durations and number of glances for radio and mirror tasks. He used a cross section of drivers in traffic. He found that individual glance times into the car clustered around 1.25 seconds, and, that for radio tasks such as tuning, four or five glances were required (See Table 2-1). Based on the results of the experiments, Rockwell concluded that there was a fairly consistent time-sharing strategy for performing instrument panel (IP) tasks while driving, that individual glance length was relatively consistent, and, that for complex tasks, more glances were required.

Bhise, Forbes, and Farber (1986) extended Rockwell's work to a greater variety of tasks. In summarizing their results, they indicated that single glance times to the IP do vary somewhat with the type of task, and number of glances varies greatly with the type of task (see Table 2-2).

Figure 2-1
Driver Visual & Manual Activity Sampling
(Source: Wierwille, 1992)

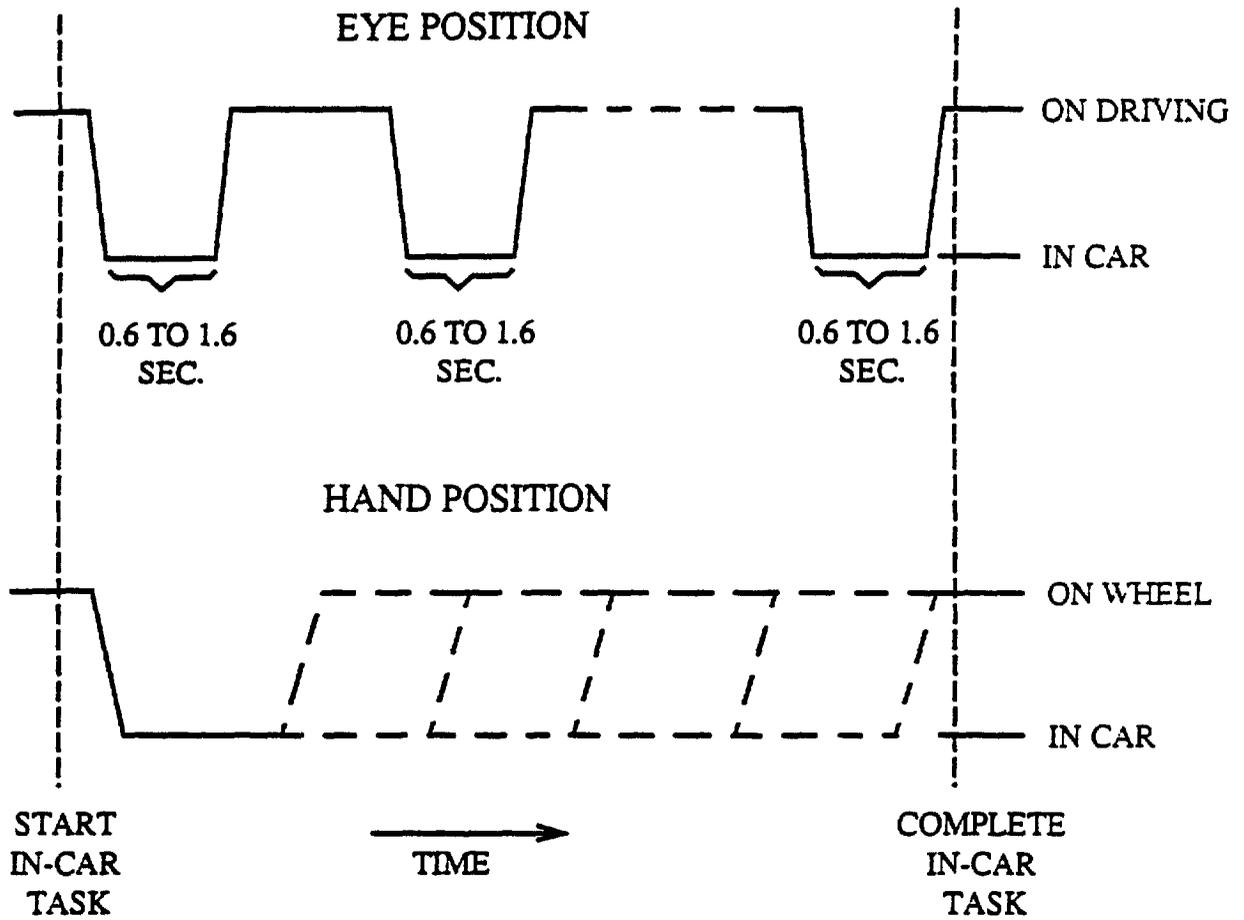


Table 2-1
Summary of Glance Duration Data for Conventional Tasks (Rockwell, 1987)

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

Commanded mirror looks of discrimination

Note: All data given in seconds.

Table 2-2
Summary of Results Presented by Bhise, Forbes, and Farber (1986)

Summary of Results Presented by Bhise, Forbes, and Farber (1986).

Tasks Requiring a Single Glance		
Task	Mean Glance Duration (Seconds)	
Read Analog Speedometer		
• Normal	0.4 to 0.7	
• check	0.8	
• Exact Value	1.2	
Read Analog Fuel Guage	1.3	
Read Digital Clock	1.0 to 1.2	
Tasks Requiring Several Glances		
Task	Number of Glances	Mean Glance Duration (Seconds)
Turn on Radio, Find Station, Adjust Volume	2to7	1.1
Read all Labels on a 12-Button Panel	7to15	1.0

Recently, Dingus, Antin, Hulse, and Wierwille (1989) performed a study in which in-car conventional and navigational tasks were compared in terms of visual glance times and number of glances. This research was performed in an instrumented vehicle on public roads with an operational, computerized, moving-map navigation display. The results of their study (Tables 2-3 and 2-4) showed very clearly that total visual demand (the sum of individual glance lengths into the car) varies markedly with the task.

It ranged from 0.78 second for reading speed up to 10.63 seconds for determining the name of the roadway where the next turn was to be made. Single glance lengths varied somewhat more than previous studies had indicated, with a range of 0.62 to 1.66 seconds. The average number of glances ranged from 1.26 to 6.64. This latter result agrees with earlier work, but is more comprehensive.

The results of these studies demonstrate several important relationships. First, they show the relatively narrow range of single glance times and the relatively broad range of number of glances. The results also show that gathering information for several navigation tasks and certain other conventional tasks such as radio tuning, demand much visual attention. Most importantly, the driver does not, on the average, allow single glance times to exceed about 1.6 seconds in normal highway driving conditions, even for complex information gathering tasks. Instead, the driver returns to the forward scene, attends to the driving task, and then returns to gather additional in-car information. This process continues until the task is completed.

A model for visual sampling can be developed as shown in Figure 2-2 (Wierwille, 1992). This is a logic model and is based on the experimental data just cited. The model is normative and deterministic. As shown in Figure 2-2, when a driver begins to perform an in-vehicle task, the driver does so by glancing to the appropriate location. Information extraction (chunking) begins as time passes. If the information can be chunked at about 1 second or less, the driver will do so and will then return glance to the forward scene. On the other hand, if chunking takes longer, the driver will continue to glance at the location for a bit longer. However, in doing so, the driver senses time pressure to return to the forward scene. If the glance to the in-car location continues up to about 1.5 seconds and the information cannot be chunked, the driver will return glance to the forward scene anyway, and will try again later. On the other hand, if chunked information can be obtained within about 1.5 seconds, the driver will extract the information and return to the forward scene. Additional samples would be handled in exactly the same way, until the driver obtains required visual information. Note, this process uses working memory of the previous glance to guide a subsequent glance. Furthermore, working memory must be allocated to concentrate all chunks into a single coherent "message".

The model depicted in Figure 2-2 is admittedly too deterministic. Drivers do not measure elapsed time accurately and they do not consciously decide whether the available information can be chunked. Furthermore, they do not use the logic flow as specifically as it is shown. Nevertheless, the model does explain the processes that must be taking place that limits the lengths of glances to in-car displays and controls.

Table 2-3
Total In-car Glance Times for a Variety of Conventional and Navigation Tasks
(Dingus et al, 1989)

<i>Task</i>	<i>Mean*</i>	<i>Standard Deviation</i>
Speed	0.78	0.65
Following Traffic	0.98	0.60
Time	1.04	0.56
Vent	1.13	0.99
Destination Direction	1.57	0.94
Remaining Fuel	1.56	0.95
Tone Controls	1.59	1.03
Info. Lights	1.75	0.93
Destination Distance	1.83	1.09
Fan	1.95	1.29
Balance	2.23	1.50
Sentinal	2.36	1.71
Defrost	2.86	1.59
Fuel Economy	2.87	1.09
Correct Direction	2.96	1.86
Fuel Range	3.0	1.43
Cassette Tape	1.59 + 1.64**	0.96 (0.59)**
Temperature	3.50	1.73
Heading	3.58	2.23
Zoom Level	4.00	2.17
Cruise Control	4.82	3.80
Power Mirror	5.71	2.78
Tune Radio	7.6	3.41
Cross Street	8.63	4.86
Roadway Distance	8.84	5.20
Roadway Name	10.63	5.80

* Time required to search for and orient cassette tape.

** Mean total glance time is the product of mean glance duration times mean glance frequency.

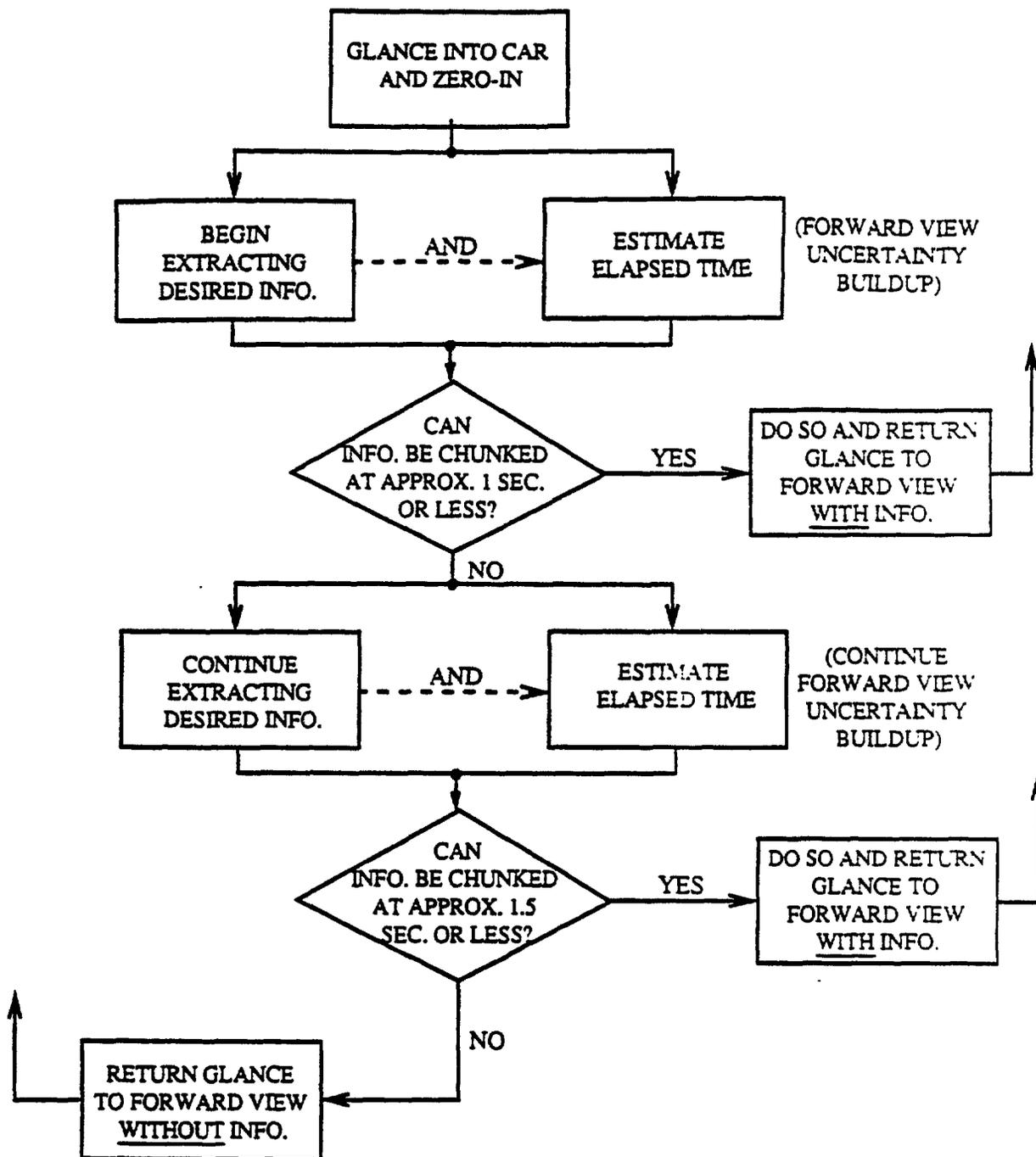
Note: All times given in seconds.

Table 2-4
Average Length and Number of In-car Glances for a Variety of Conventional
and Navigation Tasks (Dingus et al, 1989)

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

Note: Glance length given in seconds.

Figure 2-2
 Driver Visual Sampling Model
 (Source: Wierwille, 1992)



One of the major questions for in-cab device-use concerns the limits on sample lengths and the number of samples that should be considered acceptable. Zwahlen, Adams, and DeBald (1988) in performing work on simulated touch entry devices (TEDs) have suggested that in-vehicle devices should not require mean single glance lengths greater than about 1 second and number of glances greater than about three. Wierwille (1992) believes that mean single glance lengths of 1.25 seconds should be considered acceptable, although values shorter than this are preferred. In terms of number of glances, a limit of about six is probably acceptable, particularly when mean single glance times do not exceed 1.25 seconds. Yet longer glance durations and greater glance frequencies were observed for the subjects in Task 3, suggesting that perhaps under certain low demand driving conditions these kinds of looking behaviors do not compromise safety. Clearly, what is needed at present is better or more objective information on which to base such judgments. Additionally, information is needed on how to design advanced-technology devices so that they do not require mean glance durations or number of glances above specified criteria. Such information does not presently exist.

Given a free choice, drivers would prefer not to glance at the appropriate location for more than about 1 second. They will do so, however, under many circumstances if they must. They will not usually glance away from the road for more than about 1.6 seconds. While it varies somewhat with the individual driver and the driving conditions, each driver has an upper bound. Time pressure and forward scene uncertainty build to the point that the driver is compelled to return glance to the forward scene. Wierwille, Hulse, Fischer, and Dingus (1988) and Kurokawa and Wierwille (1990) have found that drivers behave as might be expected from the model presented in Section 1.0 of this report:

- As roadways become more difficult to drive, the probability that the driver's eyes will be on the roadway increases and the probability that the driver's eyes will be on an in-car navigation display decreases by about the same amount.
- Glance length to the forward driving scene increases with increased traffic density and the possibility of an impending conflict. Under similar conditions, the probability that the driver's eyes will be on an in-car navigation display decreases.
- As crosswind disturbance levels increase while the driver performs various conventional m-car tasks, single glance length to the roadway increases and single glance length into the car decreases.

All of these results and those reported by Rockwell and others indicate that, as predicted by the model given in Section 1.0, drivers adapt their task scheduling strategies according to their assessment of the driving situation and emphasize the driving task as appropriate. This important result is well captured by Rockwell (1987):

Our driver is generally resistant to the electronic temptations placed before him on the instrument panel and limits his excursion from the road to less than 1 1/2 seconds. As long as these temptations are optional . . . , he will pay the price in more glances but not longer glances- Hopefully, his visual frustration will be paid for in the marketplace where poor ergonomic designs will meet rapid obsolescence. (p. 324).

A significant caveat for the commercial vehicle operator is that in-cab device use may not be perceived as optional (e.g., replies to a dispatcher).

Manual Resource. Ordinarily drivers have two hands with which to perform the driving task. One hand is generally sufficient to maintain steering input and make small corrections. Viano, Patel, and Ciccone (1989) analyzed arm position from films of nearly 2,000 vehicles and reported that the average driver used an armrest 34.4 percent of the time on the open road and the driver's left hand was on the steering wheel 69.6 percent of the time when stopping at intersections. The manual activity analysis reported in the Task 3 interim report for the current project (Kiger, et al, 1992) indicated that, in a field study of heavy vehicle drivers, both hands were on the wheel less than half the time. This suggests that drivers often can drive with only one hand, as appears to be the case, based on intuition.

Larger corrections and turns generally require the use of both hands. If the vehicle has a manual transmission or requires the use of additional levers or controls, the driver must time-share one hand between that extra task and the steering wheel task. Experienced drivers generally learn to perform such time-sharing proficiently, and seldom think about the process. There are nevertheless occasions when part of the manual task must be delayed slightly. When turning a sharp corner, for example, a driver will try to delay shifting gears while the steering wheel is in motion, so that both hands can be used on the wheel to obtain more complete control.

For in-vehicle tasks, most of which are performed on straight or only slightly curved sections of road, the driver can ordinarily make a hand available for in-cab device use. Hand-off-wheel time has been measured in at least two studies (Kurokawa and Wierwille, 1990; Hayes, Kurokawa, and Wierwille, 1989). Their results show that, for conventional tasks, mean hand-off-wheel time ranges from about 1.25 seconds (press one of three buttons on a radio) to about 12.5 seconds (press AM/FM, then tune to a specific frequency using a digital dial and rate change tuning control). However, it is also important to note that the steering wheel is not always "a home base" for the hands (Kiger, et. al., 1992).

In many cases, hand-off-wheel time "tracks" (follows the same time course) as task completion time since more complex in-car tasks combined visual-and-manual tasks. For target acquisition tasks (primarily manual tasks in which the driver must position his hand to a discrete control such as a two state button or switch), the relationship between hand-off-wheel and visual allocation is likely to be predicted well by Fitts' Law (Jagacinski, 1989). That is, the smaller the control and/or the further the travel distance to reach that control, the longer the hand is off the wheel and the eyes are off the road. For continuous adjustments which are visually guided (e.g., tuning a radio, tabbing to a desired menu option) both of these values increase substantially. An exception to the rule that visual allocation and manual allocation track together may be for highly overlearned adjustments where non-visual feedback is available (e.g., shifting gears, engaging turn signals).

It is probable that while the hand is attending to the in-vehicle task, the driver has somewhat less capability in maneuvering the vehicle in the lateral-directional mode. However, assuming detection of a hazard has occurred, it does not take long to bring the hand back to the wheel to perform an emergency maneuver. Thus, manual demand of in-vehicle tasks has a cost in terms of driver capability, but it is relatively small. It must be recognized, however, that the manual demand of an in-vehicle task cannot be treated in isolation. As discussed above, eye-hand coordination is necessary for most manual inputs to the IP (instrument panel) which creates additional visual load. This is an important consideration in the design of electronic in-vehicle controls, since control usage usually includes a visual cost.

Cognitive Resource. There is a cognitive component associated with the primary task of driving as well as with most in-vehicle tasks. This can be relatively small, as in the case of driving on a good road with low traffic density, or in adjusting the volume of a radio. There are other cases where cognitive load can be high; such as when the driver attempts to interpret confusing route or direction signs in the forward view, or complex visual displays in the cab.

It might appear that cognitive load is not very important because a driver can easily shift attention to the primary task of driving if the driver detects a hazard either by visual or auditory means. If this is so, then it would appear that cognitive load, though high, has little bearing on safety or accident rate. Closer examination indicates otherwise. Cognitive load can result in perceptual narrowing and temporary rejection of apparently irrelevant stimulus information (Broadbent, 1958; Moray, 1981). Cognitive load is also likely to draw foveal vision to a specific area or device, thereby temporarily suppressing visual scanning, which may be needed for hazard detection. Thus, sharing cognitive resources can reduce the margin of safety.

Cognitive load is also difficult to assess, because, unlike visual and manual load, it cannot be directly measured. Ordinarily a researcher must resort to workload estimation procedures, such as opinion or response measures, which are subject to inaccuracies. In general, one can usefully relate cognitive load or attention with foveal vision (Wierwille, 1992). However, there are instances where the in-cab device has no visual display and requires no manual input, e.g., hands-free voice communications systems. Here, three studies have been conducted whose methods show promise for indicating the level of cognitive load. These three studies are similar in that no visual or manual loads are present for the telephoning task, i.e., manual dialing or handset holding are excluded (cf. Stein, Parsegian, and Allen, 1987).

Brown, Tickner, and Simmonds (1969) conducted one of the earliest on-the-road assessments of cognitive load in the absence of device-induced visual or manual loads, i.e., the use of hands-free mobile telephones. In this study, drivers actively listened to logical propositions (A follows B) and applied the propositions to test cases (BA), after which the driver had to vocally respond "true" or "false". Results indicated that when engaged in such a "telephoning" task, the drivers errors in gap acceptance judgements increased (they more frequently tried to drive through gaps smaller than the car and slightly fewer gaps that were larger than the car compared to a no-device control condition). Drivers also slowed down while engaged in the telephoning task. Thus, gap acceptance might be considered an indicator of cognitive load. It would be important to extend this work to real-world distance estimation needed for freeway entrance maneuvers.

Alm and Nilsson (1990) more recently conducted a simulator study of hands-free mobile telephones using a combined logical reasoning and memory test as the telephoning task. Sensical or non-sensical sentences were presented over an intercom (e.g., "The boy brushed his teeth", "The train bought a newspaper") and a driver had to respond "yes" if the sentence was sensible, "no" if not, within a three second deadline. In addition, after five sentences had been presented, the driver had to recall the last word in each sentence. This telephoning task led to longer driver reaction times to an unexpected visual stimulus (a red square simulating a car parked on the roadside) when the driving task was easy (i.e., relatively straight route); drivers also drove more slowly in this condition. The telephoning task also increased variability in lane position, but only when the driving task was hard (i.e., very curvy route). Subjective workload was uniformly higher when telephoning than when not (using TLX scales). Interestingly, no differences in telephoning performance (either in correct decisions for the logical reasoning task, or for recall) were found across hard vs. easy driving tasks.

While this lack of a primary task difficulty effect may indicate a flaw in the study design (Kantowitz & Knight, 1978), overall results suggest that object detection and lanekeeping performance may be both sensitive and safety-relevant indicators of cognitive load.

Most recently, McKnight and McKnight (1991) completed a simulator study on cellular phone use for the AAA Foundation for Traffic Safety. In a novel approach, drivers were presented with a 25 minute video driving sequence containing 47 situations to which drivers would normally respond either by braking, slowing, or turning. Instead of asking drivers if they saw what they were supposed to respond to, the occurrence and latency of appropriate responses were recorded. Simple conversations were created by chit-chat with the experimenter while complex conversations were created by presenting mental arithmetic or short term memory span problems to which the driver responded. Considering hands-free telephoning only, results indicated that complex conversations resulted in the longest average latencies and the highest proportion of non-responses; simple conversation had the smallest effect on responding. These effects were most pronounced among older drivers, with the likelihood that some driving situation will go unnoticed roughly twice that of younger drivers. These results indicate the viability of yet another means to assess cognitive load; object and event detection in driving conditions such as lane termination, vehicles stopping or turning, route changes (turns), stop signs or lights, and the like. It also indicates the viability of using driver-vehicle performance measures with highway safety implications to assess cognitive load directly. Note that in this case, the level of cognitive resources given to the telephoning task by the drivers is, on average, somewhat greater than that given to the driving task. Thus, the framework given in Section 1.0 of this report predicts that when drivers perceive the value of in-cab transactions to be greater than that for the primary driving task, object and event detection will decrease.

In general, there is a need to extend these results to more realistic communications tasks. One approach might be to record and analyze a set of dispatcher communications for average length, content, degree of driver interaction, and so forth. These could be used to form more realistic messages for use in protocol scenario development.

Auditory Resource. Drivers do ordinarily use auditory capabilities to help them drive, even though such capabilities are not absolutely essential. In-vehicle devices may require auditory resources as well. Examples include most communications systems and many types of warnings. Auditory inputs to the driver must generally be received serially to be easily understood. While time-overlap can be tolerated, in some cases it may create increased cognitive load in deciphering messages and may increase errors in reception. There is a possibility also that, while receiving auditory messages, the driver may exhibit perceptual narrowing, as with cognitive load. This resource can be equated with cognitive resources as described in the previous discussion.

Other Resources. There are other resources the driver uses that may be needed for in-vehicle devices. Primary among these is speech, which is usually needed for communications and which may become important for commanding in-vehicle devices, once speech recognition systems become reliable for use in vehicles. Some use of this technology has already been implemented in cellular telephone dialing.

Pedal (foot) usage is another resource that might possibly be used. However, in-vehicle technology does not appear to be moving in the direction of foot usage, probably because the feet are usually needed for longitudinal control.

Finally, the proprioceptive (tactile) resource may be used. Research has been done on using a hand-sensed display as an aid in car-following (Fenton, 1966). The driver uses the right hand (in left-hand drive vehicles) to determine whether car following distance is safe. Again, recent trends do not include the use of this resource.

. **Resource Ranking.** A review of these resources provides an intuitive insight into their importance. It is quite clear that the visual resource must be given the highest weighting and that it must receive the most careful consideration. The manual resource has some importance because of its need in steering. Similarly, cognitive load is important, because of perceptual narrowing and the need to draw on foveal visual resources for information gathering. Finally, auditory and speech resources have some importance because the driver does gather some driving information auditorily and because both auditory and speech resources may include a cognitive component. It is unlikely that pedal and proprioceptive resources would be used to any appreciable extent **for** in-cab devices, so they may be eliminated from further consideration. Conceptually, the deployment of driver resources can be depicted as shown in Figure 2-3. When the driver performs an in-cab task, that task must compete for resources that could be used for the primary task.

A Driver Resources-based Taxonomy of In-cab Tasks

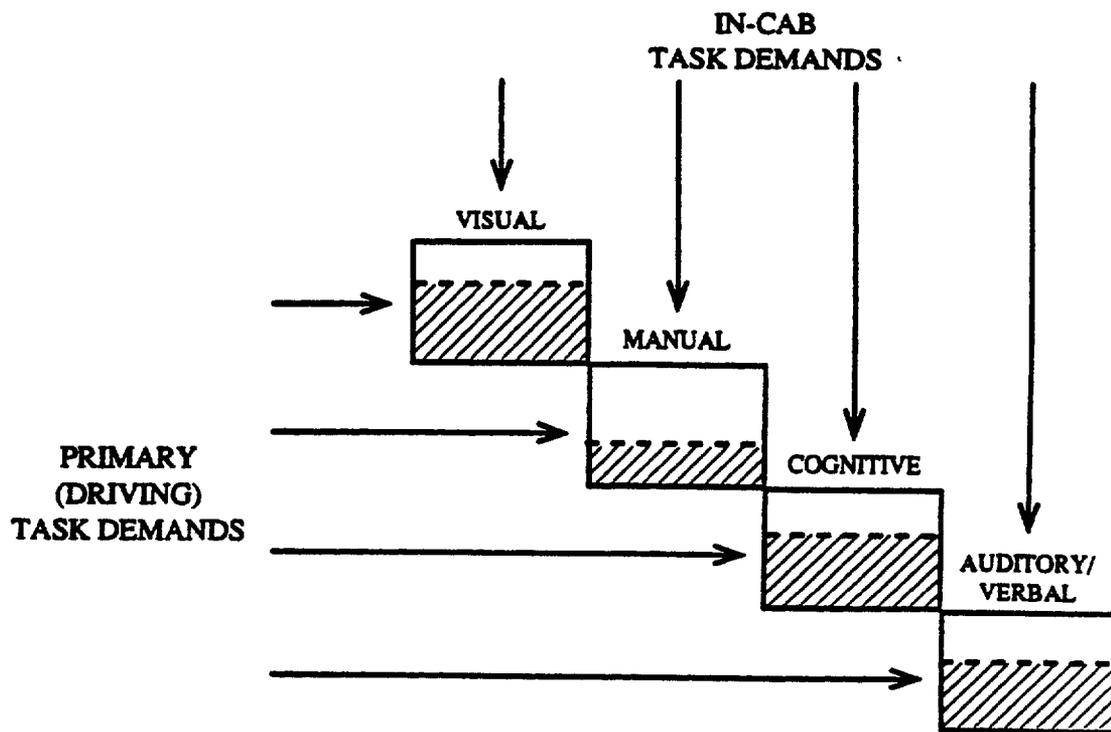
Based on the driver resources needed, a taxonomy of in-cab tasks can be defined. This taxonomy is taken from Wierwille (1992). Its usefulness is elaborated upon in Section 3.0 of this report.

The various classes of in-cab tasks can be defined in terms of their visual and manual aspects:

- **Manual only** (tasks performed without visual reference, e.g., pressing set or resume on **cruise** control).
- **Manual Primarily** (vision used to find control, then task performed without further visual reference, e.g., turning on radio, adjusting volume, changing fan speed on air conditioner).
- **visual only** (tasks that are completely visual, e.g., reading speedometer, determining current mode of HVAC, determining from a map display whether vehicle is headed in correct direction).
- **Visual Primarily** (some manual input required, e.g., determining the station frequency on the radio when display is initially in time mode, changing modes and verifying change and then completing the task).
- **Visual-manual** (tasks with interactive visual and manual demands, e.g., manually tuning a radio to a specific frequency, operating a cellular telephone, making mirror adjustments, zooming in or out on a map display).

Note that this taxonomy was taken from a paper which emphasized visual and manual loading. other task types are possible (e.g., auditory feedback when adjusting radio volume). The issue of cognitive demand (above and beyond visual or manual demand) deserves special mention. Generally, the driver can shift cognitive resources quickly, provided that vision is in the appropriate direction, i.e., events

Figure 2-3
Depiction of Competition for Driver Resources



are captured in the visual periphery. Thus, the driver has an excellent chance to detect an emergency and take corrective actions. If the cognitive demands of an in-cab task are high, visual resource measures (e.g., glance frequency) will likely reflect this (for visual displays), and object and event detection will likely reveal this also. If the in-cab device is not visual in nature, then object and event detection may be the most likely measures to be sensitive to this kind of cognitive load. Further discussion of the relationship of task types to cognitive load is presented in Section 3.0.

Quantification of Resource Usage

Probably the most difficult portion of any model of resource allocation for an in-vehicle task is specifying the model quantitatively. This problem is difficult to deal with because weightings and operating characteristics must be supplied before any numerical values can be computed. Such specifications must be determined intuitively, in much the same way that a mathematician must specify a criterion prior to performing an optimization. The criterion must necessarily be determined intuitively, even though it may be based on logical assumptions.

Kurokawa and Wierwille (1990) have attempted to provide appropriate characteristics for visual and manual demands by developing a so-called figure of merit for instrument panel (IP) tasks. The characteristics they used were as follows:

$$V_1 = 40 A^2$$

$$V_2 = 6B^{3/2}$$

$$V = \alpha V_1 + (1-\alpha) V_2$$

$$M = 11C$$

$$FM = \beta V + (1-\beta) M$$

$$\overline{FM} = \frac{n}{10} \cdot FM$$

In these equations, the following definitions are used:

- A = average length of a glance to the IP in seconds
- B = number of glances to the IP associated with the IP task
- α = a weighting constant between zero and one, used for weighting the two visual metrics
- V = the visual demand assessment
- C = hand-off wheel time associated with the IP task in seconds
- M = the manual demand assessment

- β = a weighting constant between zero and one, used for weighting visual and manual demands
- FM = figure of merit, or assessment of combined visual and manual demands, per task attempt
- n = number of times per hour the task is performed
- = the overall figure of merit weighted by frequency of use (or performance)

These equations take visual demand as the sum of the two components, glance duration (A) and glance frequency (B), and then combine them with a manual demand component, hand-off wheel time (C). The constant α determines how the two visual components are weighted relative to one another, and the constant β determines how the visual and manual demands are weighted relative to one another. Also in the equations, A , B , and C have all been scaled (via their numeric multipliers) so that their values range between zero and approximately 100, based on the type of task and the range of values found in empirical research.

To use the approach, an experimenter obtains the values A , B , and C experimentally using driver subjects performing the IP tasks while driving (either in simulation or in full scale.) The experimenter then specifies α , β , and n . The constant α would ordinarily be set at a value of 0.5, for example, hereby weighting the two visual components equally. The constant β would ordinarily be set at a value of about 0.7 to represent the much greater importance of visual demands as compared with manual demands.

Finally, it should be noted that the two visual components have different operating characteristics. The average single glance time operates on a squared characteristic, which emphasizes the hazard of long glances into the vehicle. The number of glances operates on a $3/2$ power operating characteristic because individual short glances separated by glances back to the forward view are not quite as hazardous as long single glances.

The overall figure of merit is thus a combined assessment of demand per use made on the visual and manual resources, weighted according to importance, and also weighted according to frequency of use. Larger values indicate larger incursions into resources.

The approach just presented includes consideration of visual and manual demands. It does not include cognitive and auditory/verbal demands. However, the approach could easily be expanded to include them. In this section, equations are provided for specification of an overall figure of merit which includes the cognitive and auditory/verbal demands.

Assume that cognitive demands are to be determined by a single metric. This metric ranges from say 0 to 10, where larger values are representative of higher cognitive demands. Then let

$$K = 3D^{3/2}$$

where K is the cognitive demand assessment, and D is the cognitive demand metric.

The cognitive demand assessment as been weighted using a 3/2 power operating characteristic because high cognitive demands are more likely to create perceptual narrowing and similar problems.

Similarly, assume that auditory/verbal demands are to be assessed by a single metric where the metric ranges from, say 0 to 1, based on the proportion of IP task time when there is message traffic. Then

$$L = 100 E$$

where L is the auditory/verbal demand assessment, and E is the proportion of IP task time in which communications occur.

The figure of merit can now be redefined to include these latter two additional resources as follows:

$$FM_1 = \beta_1 V + \beta_2 M + \beta_3 K + \beta_4 L$$

where $\beta_1 + \beta_2 + \beta_3 + \beta_4 = 1$ and where each β_i represents a relative weighting of importance in the figure of merit.

The overall figure of merit can be defined as it was previously, namely,

$$\overline{FM}_1 = \frac{n}{10} \cdot FM_1$$

The selected values for the weighting constants in a specific assessment might be as follows:

$$\begin{aligned} \beta_1 &= 0.65 \\ \beta_2 &= 0.15 \\ \beta_3 &= 0.15 \\ \beta_4 &= 0.05 \end{aligned}$$

Here, visual demand has been given the heaviest weighting, manual and cognitive load equal and much smaller weighting, and communication an even lighter weighting. Note that, as a first approximation, a simple linear model is given. Cross-product or interaction terms are omitted but could be included if warranted.

This approach assumes that both the figure of merit and overall figure of merit are related to safety. The larger these values become, the greater are the driver's prime resources that are being diverted toward performance of the in-vehicle task.

These figures of merit are intended to be sensitive to task redesign in which lower-importance resources are substituted for higher-importance ones. For example, if an in-vehicle navigation system is converted from visual input to auditory input (on an equivalent basis), the figure of merit would decrease, indicating a lower value of incursion into the driver's weighted resource pool. This is exactly the type of evaluation that should be made for new tasks added to the HV driving environment.

While specific equations have been provided here, they should be considered as examples of how to factor in individual differences. Fundamentally, this approach involves assessing the driver resources needed to service the in-vehicle task, then weighting them by importance and subsequently summing them. The approach also allows for inclusion of frequency of use data, that is, frequency of performing the in-vehicle task.

One of the questions that remains to be answered is that of cutoff values for FM_1 and \overline{FM}_1 . Is there a way to specify maximum tolerable values for new devices? It appears that once again one must return to accident data bases to determine if when certain \overline{FM}_1 values are reached, accident rates become excessive. However, it could be counter-argued that even the smallest incursion can result in an accident, for example, resetting the volume control on the radio. Thus, there are no fully justifiable cutoff values. Ultimately some type of judgment would have to be made before cutoff values could be specified.

Of course, setting cutoff values is not required if relative comparisons are to be made. For relative comparisons it is only necessary to determine which in-vehicle design has the lower figure-of-merit values. Workload figures of merit for relative comparisons are likely to be quite useful yet less contentious than setting absolute thresholds for acceptable workload.

References

- Alm, H. & Nilsson, L. (1990). Changes in driver behaviour as a function of handsfree mobile telephones. Linköping, Sweden: Swedish Road and Traffic Research Institute.
- Antin, J. F., Dingus, T. A., Hulse, M. C., & Wierwille, W. W. (1990). An evaluation of the effectiveness and efficiency of an automobile moving-map navigational display. International Journal of Man-Machine Studies, 33, 581-594.
- Bhise, V. D., Forbes, L. M., & Farber, E. I. (1986). Driver behavioral data and considerations in evaluating in-vehicle controls and displays. Paper presented at the Transportation Research Board, National Academy of Sciences, 65th Annual Meeting, Washington, D.C.
- Broadbent, D. (1958). Perception and Communication. New York, Pergamon.
- Brown, I. D., Tickner, A. H., & Simmonds, D. C. (1969). Interference between concurrent tasks of driving and telephone. Journal of Applied Psychology, 53(5), pp. 419-424.
- Carr, T. H. (1986). Perceiving visual language. In K. Boff, L. Kaufman, & J. Lincoln (Eds.), Handbook of perception and human performance. Volume II (pp. 29-1 to 29-92). New York: Wiley.
- Chong, J., & Triggs, T. (1989). Visual accommodation and target detection in the vicinity of a window post. Human Factors, 31(1), 63-75.
- Dingus, T. A., Antin, J. F., Hulse, M. C., & Wierwille, W. W. (1989). Attentional demand requirements of an automobile moving-map navigation system. Transportation Research-A, 23A(4), 301-315.
- Fenton, R. E. (1966). An improved man-machine interface for the driver-vehicle system. IEEE Transactions on Human Factors in Electronics, HFE-7(4), 150-157.
- Fischer, E., Haines, R., & Price, T. (1980). Cognitive issues in head-up displays (NASA Technical Paper 1711). Washington, DC: NASA.
- Hayes, B. C., Kurokawa, K., & Wierwille, W. W. (1989). Age related decrements in automobile instrument panel task performance. Proceedings of the Human Factors Society 33rd Annual Meeting. Vol. 1 (pp. 159-163). Santa Monica, CA: Human Factors Society.
- Iavecchia, J., Iavecchia, H., & Roscoe, S. (1988). Eye accommodation to head-up virtual images. Human Factors, 30(6), 689-702.
- Jagacinski, R. J. (1989). Target Acquisition: Performance Measures, Process Models, and Design Implications. In Applications of Human Performance Models To System Design (pp. 135-149). New York, NY.
- Kantowitz, B. H., & Knight, J. L. (1978). When is an easy task difficult and vice versa? Acta Psychologica, 42, 162-170.

- Kiger, S., Rockwell, T. H., Niswonger, S., Tijerina, L., Myers, L. B., & Nygren, T. (1992). Report on NHTSA heavy vehicle driver workload assessment task 3 interim report: task analysis data collection. Columbus, OH: Battelle.
- Kurokawa, K. & Wierwille, W. W. (1991). Effects of instrument panel clutter and control labeling on visual demand and task performance. 1991 SID Digest of Technical Papers (pp. 99-102). Plaza del Rey, CA: Society for Information Display.
- Kurokawa, K. & Wierwille, W. W. (1990). Validation of a driving simulation facility for instrument panel task performance. Proceedings of the Human Factors Society 34th Annual Meeting pp. 1299-1303. Santa Monica, CA: Human Factors Society.
- Mackworth, N. H. (1976). Useful field of view. In R.A. Monty & J. W. Senders (Eds.), Eye Movements & Psychological Processes (pp. 173-178). Hillsdale, N. J.: Erlbaum.
- McKnight, J. A., McKnight, A. S. (1991). The effect of cellular phone use upon driver attention. Landover, MD: National Public Services Research Institute.
- Moray, N. (1981). The role of attention in the detection of errors and the diagnosis of errors in man-machine systems. In J. Rasmussen & W. Rouse (Eds.), Human Detection and Diagnosis of Failures. New York, Plenum.
- Mourant, R. R., Herman, M., & Moussa-Hamouda, E. (1980). Direct looks and control location in automobiles. Human Factors, 22(5), pp. 417-425.
- Mourant, R. R., Moussa-Hamouda, E., & Howard, J. M. (1977). Human factors requirements for fingertip reach controls. (Report No. DOT HS 803 267-. Washington, DC: U.S. Department of Transportation.
- Neisser, U. (1976). Cognition and Reality. San Francisco: Freeman and Co.
- Neisser, U., & Becklin, R. (1975). Selective looking: Attending to visually specified events. Cognitive Psychology, 7, 480-494.
- Rockwell, T. H. (1988). Spare visual capacity in driving revisited: New empirical results for an old idea. In A. G. Gale, et al. (Eds.) Vision in Vehicles II - Proceedings of the Second International Conference on Vision in Vehicles. Nottingham, UK. 14-17 Sept. 1987 (pp. 317-324). Amsterdam, Netherlands: Elsevier.
- Rockwell, T. H. (1972). Skills, judgment, and information acquisition in driving. In T. W. Forbes, (Ed.), Human Factors in Highway Traffic Safety Research, New York, Wiley Interscience.
- Rockwell, T. H., Balasubramanian, K. N., Kretovics, T., & Wilfong, E. J. (1977, August). The Utility of Peripheral Vision To Motor Vehicle Drivers (Report No. 784224, Contract No. DOT-HS-5-01203). Columbus, OH: Department of Industrial & Systems Engineering, The Ohio State University.

- Senders, J. W., Kristofferson, A. B., Levison, W. H., Dietrich, C. W. & Ward, 3. L. (1967). The attentional demand of automobile driving. Highway Research Record, 195, 15-32.
- Stein, A. C., Parseghian, Z., Allen, R. W. (1987). A simulator study of the safety implications of cellular mobile telephone use. In: American Association for Automotive Medicine, Proceedings of the 31st Annual Conference. Desplaines: AAAM.
- Streeter, L. A. Vitello, D., and Wonsiewicz, S. A. (1985). How to tell people where to go: comparing navigational aids. International Journal of Man-Machine Studies, 22, 549-562.
- Tijerina, L., Kiger, S., Wierwille, W., Rockwell, T. H., Kantowitz, B., Bittner, A., Nygren, T., Myers, L. B., Tolbert, C.A., McCallum, M. (1992) Report on NHTSA heavy vehicle driver workload assessment task 1 interim report: Task analysis data & protocols review. Columbus, OH: Battelle.
- Viano, D., Patel, M., & Ciccone, M. (1989). Patterns of Arm Position During Normal Driving. Human Factors, 31(6), 715-720.
- Weintraub, D. J., & Ensing, M. (1992 May). Human factors issues in head-tin display design: The book of HUD (Technical Report No. CSERIAC 92-2). Wright-Patterson AFB, OH: Crew Station Ergonomics Information Analysis Center. Godthelp, H., Milgram, P., & Blaauw, G. J. (1984). The development of a time-related measure to describe driving strategy. Human Factors, 26(3), 257-268.
- Wierwille, W. W. (1992). Visual and manual demands of in-car controls and displays. In J. B. Peacock & W. Karwowski (Eds.), Automotive ergonomics: human factors in the design and use of the automobile. London: Taylor & Francis. (In press).
- Wierwille, W. W., Hulse, M. C., Fischer, T. J., & Dingus, T. A. (1988). Strategic use of visual resources by the driver while navigating with an in-car navigation display system. XXII FISITA Congress Technical Papers: Automotive Systems Technology: The Future, Vol. II. SAE-P-211, 2.661-2.675.
- Wierwille, W. W. (1987). Can dash instrumentation visual attentional demand be predicted using the design driver concept? Paper presented at the annual meeting of the Transportation Research Board, Washington, DC.
- Wierwille, W. W. (1981). Statistical techniques for instrument panel arrangement. In J. Moraal & K. F. Kraiss (Eds.), Manned Systems Design: Methods, Equipment, and Applications, pp. 201-218, New York, Plenum.
- Zwahlen, H. T., Adams, C. C., Jr., & DeBald, D. P. (1988). Safety aspects of CRT touch panel controls in automobiles. In A. G. Gale, et al. (Eds.) Vision in Vehicles II - Proceedings of the Second International Conference on Vision in Vehicles. Nottingham. UK. 14-17 Sept. 1987 (pp. 335-344). Amsterdam, Netherlands: Elsevier.

3.0 DRIVER-VEHICLE PERFORMANCE MEASURES AND PERTINENT TASKS

Introduction

A key to establishing the safety relevance of in-cab device load measures is to correlate them to driver-vehicle performance. In this way, one may draw a link between a device-driven measure such as glance frequency with phenomena that affect highway safety (e.g., lane crossings). In this section, a number of different driver-vehicle performance measures are introduced which are believed to be of interest in protocol development. A selective review of many of these measures from the literature is also provided. Finally, prototypical tasks from the in-cab task taxonomy are presented and these relate to driver-vehicle performance measures likely to be sensitive to concurrent in-cab device loads.

Definitions of Driver-Vehicle Performance Measures Potentially Sensitive to In-Cab Device Workload

Table 3-1 lists driver-vehicle performance measures. All measures are computed over an arbitrary, specified time interval. These measures span several broad categories including:

- Lateral control measures: lane-related measures, steering-related measures, and heading and lateral-acceleration related measures.
- Longitudinal control measures: accelerator-related measures, brake/deceleration-related measures, speed-related measures, and vehicle following-related measures.
- Obstacle and event detection: probability of detection measures, detection latency measures.
- Driver response measures: stimulus-response measures.
- Vision-related measures: visual allocation to roadway, visual allocation to in-cab controls and displays.
- Manual-related measures: hands-on-wheel frequency, duration, and total time.

These measures introduce a broad range of measurement options for possible use in driver-vehicle performance assessment. No details about the instrumentation needed to collect such measures are provided at this time, but some general comments can be made. Many measures (e.g., speed-related measures) should be relatively easy to obtain using the electronic instrument package present in the cab of a modern vehicle. Some measures (e.g., vehicle-following measures) may be captured with off-the-shelf technology added to a test vehicle (e.g., VORAD headway detection system). Collection of others (e.g., lane-related measures) will possibly depend on the availability of emerging technologies (e.g., LaneLok) or may be most readily captured in a simulator setting.

TABLE 3-1. DRIVER-VEHICLE PERFORMANCE MEASURES

Measurement Category	Measurement	Category
Lane-Related Measures	Mean Lane Position	Mean value of lateral lane position as measured from a point on the centerline of the vehicle projected downward to the pavement and subtracted from lane centerline or other fixed lane reference. (Note that mean lane position is dependent on selection of the lane reference.)
	Lane Standard Deviation	The standard deviation of lateral lane position as measured from a point on the centerline of the vehicle projected downward to the pavement and subtracted from lane centerline or other fixed reference.
	Lane RMS Deviation	The root-mean-square deviation of lateral lane position as measured from a point on the centerline of the vehicle projected downward to the pavement and subtracted from lane centerline or other fixed lane reference. (Note that since RMS computations include the mean value, this measure will change value with lane reference position.)
	Peak Lane Deviation	The maximum magnitude of lateral lane position as measured from a point on the centerline of the vehicle projected downward to the pavement and subtracted from lane centerline or other fixed reference. (Note that peak lane deviation is dependent on selection of the lane reference.)
	Lane Exceedences	The number of times that any portion of the vehicle exceeds the designated lane boundaries.
	Mean Lane Exceedence Duration	The mean time length of lane exceedences. (Note that this measure can only be computed if at least one lane exceedence occurs.)
Heading/Heading Rate/Lateral Acceleration--Related Measures	Yaw Standard Deviation	The standard deviation of angular difference between the vehicle longitudinal axis and the instantaneous tangent to the roadway (or lane) centerline. In articulated vehicles, the longitudinal axis of the section in which the driver is seated (the tractor) is to be used.
	Yaw-Rate Standard Deviation	The standard deviation of the derivative (with respect to time) of the angular difference between the vehicle longitudinal axis and the instantaneous tangent to the roadway (or lane) centerline. In articulated vehicles, the longitudinal axis of the section in which the driver is seated (the tractor) is to be used.
	Lateral Acceleration Standard Deviation	The standard deviation of lateral acceleration as measured along the vehicle's transverse axis. In articulated vehicles, the transverse axis of the section in which the driver is seated (the tractor) is to be used.
	Steering-Related Measures	Steering Standard Deviation

TABLE 3-1. DRIVER-VEHICLE PERFORMANCE MEASURES (Continued)

	Peak Steering Deflection	The maximum magnitude of steering deflection as measured from a fixed (or zero) deflection position.
	Steering Velocity Standard Deviation	The standard deviation of steering velocity.
	Steering Reversals	The number of times that steering velocity changes from a magnitude of ϵ or greater in one direction to a magnitude of ϵ or greater in the other direction.
	Small/Large Steering Reversals	The number of intervals between steering reversals in which the magnitude of incremental steering displacement equals or exceeds β . For small steering reversals, β is selected as a small value. For large reversals, β is selected as a large value. (Note that the number of steering reversals equals or exceeds the number of small steering reversals, and the number of small steering reversals equals or exceeds the number of large steering reversals.)
	Steering Holds	The number of times that the magnitude of steering velocity remains at or below ϵ for a duration of T seconds or longer.
	Steering Zero-Crossings	The number of times that steering displacement passes from a magnitude of δ_0 or greater in one direction, through zero, and then to a magnitude of δ_0 or greater in the other direction.
	Steering Response Time	The time from presentation of a specified stimulus (with specified start time) to the time that <ul style="list-style-type: none"> a. incremental steering displacement magnitude equals or exceeds β_1 or b. incremental steering velocity magnitude equals or exceeds β_2 or both of the above occur. (Note that since steering is a normal part of driving, additional defining aspects may be required to separate stimulus response from ongoing steering.)
Accelerator-Related Measures	Accelerator Standard Deviation	The standard deviation of accelerator displacement.
	Accelerator Reversals	The number of times that accelerator velocity changes from a magnitude of ϵ or greater in one direction to a magnitude of ϵ or greater in the other direction.

TABLE 3-1. DRIVER-VEHICLE PERFORMANCE MEASURES (Continued)

	Small/Large Accelerator Reversals	The number of intervals between accelerator reversals in which the magnitude of incremental accelerator displacement equals or exceeds β . For small accelerator reversals, β is selected as a small value. For large reversals, β is selected as a large value. (Note that the number of accelerator reversals equals or exceeds the number of small accelerator reversals, and the number of small accelerator reversals equals or exceeds the number of large steering reversals.)
	Accelerator Holds	The number of times that the magnitude of accelerator velocity remains at or below ϵ for a duration of T seconds or longer.
	Mean Accelerator Hold Duration	The mean time length of accelerator holds. (Note that this measure can only be computed if at least one accelerator hold occurs.)
	Total Accelerator Hold Time	The total length of time that the magnitude of accelerator velocity remains below ϵ .
	Accelerator Releases	The number of times that the accelerator is returned to the undeflected position for T seconds or longer.
	Mean Accelerator Release Duration	The mean time length of accelerator releases. (Note that this measure can only be computed if at least one accelerator release occurs.)
	Total Accelerator Release Time	The total time that the accelerator is in the undeflected position.
Brake/Deceleration-Related Measures	Brake Applications	The number of times the brakes are applied with sufficient force to energize the vehicle's brake lights.
	Mean Brake Application Duration	The average length of time per application that the brakes are applied with sufficient force to energize the vehicle's brake lights. (Note that this measure can only be computed if at least one brake application occurs.)
	Total Brake Time	The total length of time that the brakes are applied with sufficient force to energize the vehicle's brake lights.
	Brake Response Time	The time from presentation of a specified stimulus (with specified start time) to the time that the brakes are applied with sufficient force to energize the vehicle's brake lights. (Note that since braking occurs in normal driving, additional defining aspects may be required to separate stimulus response from ongoing braking response.)
	Peak Longitudinal Deceleration	The maximum deceleration as measured along the vehicle's longitudinal axis.
Speed-Related Measures	Mean Speed Error	The mean of the difference between vehicle forward (longitudinal) velocity and instructed or reference speed.

TABLE 3-1. DRIVER-VEHICLE PERFORMANCE MEASURES (Continued)

	Speed Standard Deviation	The standard deviation of forward (longitudinal) velocity of the vehicle.
	Speed RMS Deviation	The root-mean-square value of the difference between forward (longitudinal) velocity of the vehicle and instructed or reference speed.
Vision-Related Measures	Mean Roadway Glance Duration	The mean of glance lengths to the vicinity of the forward roadway. A glance is defined as one or more consecutive fixations to an object or small area. (Very small changes in angular eye position are included in the same glance.)
	Number of Roadway Glances	The number of glances to the vicinity of the forward roadway. A glance is defined as one or more consecutive fixations to an object or small area. (Very small changes in angular eye position are included in the same glance.)
	Total Roadway Glance Time	The total time that the eyes are directed in the vicinity of the forward roadway.
Manual-Related Measures	Hands-on-Wheel Occurrences	The number of times that the driver places both hands on the wheel without changing hand positions. Hands-on-wheel is defined as having both of the driver's hands in contact with the rim or spokes of the steering wheel.
	Mean Hands-on-Wheel Duration	The mean length of time that the driver places both hands on the wheel without changing hand positions. Hands-on-wheel is defined as having both of the driver's hands in contact with the rim or spokes of the steering wheel. (Note that this measure can only be computed if hands-on-wheel time is nonzero.)
	Total Hands-on-Wheel Time	The total time that both of the driver's hands are in contact with the rim or spokes of the steering wheel.
Vehicle Following-Related Measures'	Following Distance	Distance between a lead vehicle and the driver's vehicle
	Time Head way	The following distance (see above) divided by the rate of change of the following distance (i.e., the closing velocity between the vehicle).
	Following Distance Mean Error	The mean value of error in following distance, where error in following distance is defined as the difference between actual following distance and instructed or specified following distance.
	Following Distance Standard Deviation	Standard deviation of the distance between the lead vehicle and the driver's vehicle.
	Following Distance RMS Error	The root-mean-square of error in following distance, where error in following distance is defined as the difference between actual following distance and instructed or specified following distance.
	Minimum Following Distance (Closest Longitudinal Approach)	The smallest value of distance between the lead vehicle and the driver's vehicle. (Zero or negative values indicate collision.)

TABLE 3-1. DRIVER-VEHICLE PERFORMANCE MEASURES (Continued)

	Following Distance Peak Closing Velocity	The maximum value of the negative of the rate of change of distance between the lead vehicle and the driver's vehicle.
Object/Obstacle Related Measures²	Minimum Miss Distance (Closest Approach)	The smallest resultant distance between any point on the driver's vehicle and a specified object or obstacle. (A zero or negative value indicates collision.)
	Peak Closing Velocity	The maximum value of the negative of closing velocity, where closing velocity is defined as the rate of change of resultant distance between a specified point on the driver's vehicle and a specific object or obstacle.
Stimulus Response Measures³	Response Time (Type I)	The time from presentation of a specified stimulus (with specified start time) to the time that the driver responds correctly, either verbally or with appropriate hand or foot motion.
	Response Time (Type II)	The stimulus is assumed to be at a fixed location in the forward scene.) The time from correct driver response (verbal, manual, or pedal) until the driver's vehicle is alongside the stimulus.
	Response Distance	The distance of the driver from the stimulus when the driver responds correctly, either verbally or with appropriate hand or foot motion.
	Errors of Omission	The number of times that a driver fails to respond to specified stimulus presentation.
	Errors of Commission	The number of times that a driver responds incorrectly to specified stimulus presentations.

1 NOTE THAT THE MEASURES LISTED BELOW AND DEFINED IN OTHER SECTIONS MAY ALSO BE USED AS VEHICLE FOLLOWING MEASURES:

ACCELERATOR REVERSALS	BRAKE RESPONSE TIME
SMALL/LARGE ACCELERATOR REVERSALS	PEAK LONGITUDINAL DECELERATION
ACCELERATOR RELEASES	SPEED STANDARD DEVIATION
BRAKE APPLICATIONS	MEAN ROADWAY GLANCE DURATION
MEAN BRAKE APPLICATION DURATION	

2 NOTE THAT THE MEASURES LISTED BELOW AND DEFINED IN OTHER SECTIONS MAY ALSO BE USED AS OBJECT/OBSTACLE MEASURES:

PEAK LANE DEVIATION	MEAN ACCELERATOR RELEASE DURATION
LANE EXCEEDENCE DURATION	MEAN BRAKE APPLICATION DURATION
PEAK LATERAL ACCELERATION	BRAKE RESPONSE TIME
PEAK STEERING DEFLECTION	PEAK LONGITUDINAL DECELERATION
STEERING RESPONSE TIME	MEAN ROADWAY GLANCE DURATION

3 NOTE THAT THE MEASURES LISTED BELOW AND DEFINED IN OTHER SECTIONS MAY ALSO BE USED AS STIMULUS RESPONSE MEASURES:

STEERING RESPONSE TIME	MEAN ROADWAY GLANCE DURATION
TOTAL ACCELERATOR RELEASE TIME	ACCELERATOR HOLDS
BRAKE RESPONSE TIME	ACCELERATOR RELEASES

Other measures (e.g., object and event detection measures) will depend on special procedures and instrumentation to be developed specifically under this program. Many measures will benefit by software which supports data reduction of continuous data sampled at reasonably high data rates (e.g., steering wheel velocity). More specifics about measurement procedures will be provided in Task 5 of this project.

Appendix 3A provides some indication of the use of these measures in the literature. In general, some measures are more sensitive than others to at least some driver loads:

- With the exception of mean lane position, lateral control measures show promising sensitivity to variations in driving demand and driver interaction with in-cab tasks. Longitudinal control measures provide mixed opportunities.
- Accelerator-related measures do not seem to show much sensitivity (based on the two studies reviewed in Appendix 3A), and their safety relevance is not clear.
- Brake/deceleration measures have both high safety relevance and demonstrated sensitivity to various experimental manipulations.
- Speed-related measures are more safety relevant than accelerator pedal measures and have shown sensitivity to at least some variations in driving conditions (e.g., driving through shopping area versus residential area; wind gusts in a simulator setting). In a related vein, vehicle following measures have high safety relevance (e.g., maintaining safe headway) and have shown speed-related and subsidiary task-related sensitivity.
- Obstacle and event detection has prima facie safety relevance and has been shown to be a sensitive measure of in-cab device loads.
- Driver response measures such as secondary task responses have shown sensitivity in some instances but suffer the same problems associated with the methodology in general. That is, it is an unrealistic task if it represents a load which may not be clearly interpretable either in terms of the in-cab device characteristics or in terms of highway safety and it is presented in the presence of an on-going in-cab transaction simultaneous with driving. On the other hand, embedded tasks which are a legitimate part of driving (e.g., obstacle detection) are preferred.
- The importance of vision and the sensitivity of vision-related measures to in-cab devices is discussed in Section 2.0. Manual-related measures have not been studied extensively but, for at least some complex interactions, might be expected to decrease lanekeeping performance.

The next step is to characterize prototypical in-cab tasks, as discussed in the next subsection. After this, the applicability of the various driver-vehicle performance measures is assessed in terms of these tasks.

Likely Sensitivity of Driver-Vehicle Performance Measures to In-Cab Device Workload

One of the difficulties associated with estimation of likely sensitivity of driver-vehicle performance measures to in-cab device demands is that in-vehicle tasks always require a combination of driver resources. For example, there is no such thing as an in-vehicle task that requires visual resources only. Even the simplest task having a visual component, such as reading the speedometer, has a cognitive component. Thus, it is unrealistic to assess likely sensitivity in terms of the major resources (visual, manual, cognitive, and auditory/verbal) separately. They do not occur separately, and there are no corresponding previous data on which to base sensitivity.

A better procedure is to specify tasks that draw upon differing amounts of the resources and then subsequently to estimate the likely sensitivity of various driver/vehicle/roadway measures to such tasks. Such a procedure provides a more realistic assessment and also provides a means of matching future tasks to current ones for which there are data. In this section, nine tasks are defined that can serve as benchmark tasks. They represent typical tasks that “exercise” the various driver/vehicle/roadway measures and thus provide an indication of likely sensitivity. Once the tasks have been defined, a sensitivity matrix can be developed for the various measures.

These benchmark tasks were selected on the basis of their visual, manual, and cognitive components. The auditory/verbal components have been included only indirectly, because including them directly would have resulted in too many tasks and because the auditory/verbal components are considered to be least important in terms of competing with resources needed for driving.

The tasks consist of five classes of visual/manual components and two classes of cognitive components (high and low). This would result in ten tasks; however, one of the ten (manual-only, high cognitive load) is a null condition in which no known task could be found. Thus, nine benchmark tasks are proposed.

Task 1: Visual Only, Low Cognitive Load

The driver performs a complete visual scan of all gages and indicators within the vehicle for the purpose of checking vehicle operation. Because the driver already has a good understanding of what normal indications are, the cognitive load consists of simple comparisons of present values with memorized values or ranges.

Task 2: Visual Only, High Cognitive Load

The driver reads dispatching instructions from a text display. The instructions include a description of items to be picked up or delivered, the location of the destination, and directions for driving to the destination. Cognitive load consists of interpreting directions and in memorizing items to be picked up or delivered.

Task 3: Visual Primarily, Low Cognitive Load

The driver accesses a specific page from a menu-type display and then performs a visual check. For example, using either hard- or soft-keys, the driver accesses the HVAC page of the display, and then reads the set temperature, the outside temperature, and the fan speed.

Task 4: Visual Primarily, High Cognitive Load

The driver uses an in-vehicle moving-map display to determine the best route to a destination. Present position, destination, and roadway network are shown, and the driver must determine the route. Cognitive load consists of complex spatial reasoning. Manual inputs may be required to select appropriate zoom settings.

Task 5: Visual-Manual, Low Cognitive Load

The driver grasps and inserts a cassette tape into a tape player, and then adjusts the volume and balance controls. Because the cassette has several possible orientations, there is a more-or-less continuous visual-manual load until the tape is correctly inserted. Thereafter the driver must locate and adjust the volume and tone controls. This also requires visual and manual load. Cognitive load is low because of the ease in identifying correct orientation and locations.

Task 6: Visual-Manual, High Cognitive Load

This task is a navigation task using a moving map display. The driver shifts the map laterally to examine the roadway system in an adjacent area. The objective is to determine an alternate roadway to a destination by examination of nearby parallel routes. This task involves complex spatial reasoning as well as manual adjustment of map lateral position as alternate routes are examined.

Task 7: Manual Primarily, Low Cognitive Load

The driver manually tunes a radio to a station based on a specific type of program content, such as country music. Once the tuning control has been located visually and accessed manually, further visual input is usually unnecessary. The driver listens for program content until a match with desired program content is found.

Task 8: Manual Primarily, High Cognitive Load

The driver requests and receives dispatching instructions via two-way radio, using a hand-held microphone with talk switch for verbal output and a loudspeaker for receiving verbal instructions. The instructions include a description of items to be picked up or delivered, the location of the destination, and directions for driving to the destination. Cognitive load consists of interpreting directions and in memorizing items to be picked up or delivered.

Task 9: Manual Only, Low Cognitive Load

The driver deflects the turn signal lever. The driver can usually accomplish this without visual reference and with low cognitive load.

Task 10: Auditory-Vocal

A final task type might be called "Auditory-Vocal". While uncommon at this time, an example might be interaction with a hands-free voice communication system. In such a system, the driver has no visual display since an intercom arrangement is used. Similarly, there is no manual load because voice commands accomplish such activities as dialing and sign-off. Such tasks cannot readily be evaluated with either visual allocation or manual load measurements. The cognitive component of such tasks may be relatively low (e.g., casual conversation) or quite high (e.g., receiving detailed instructions about a route). The level of operator involvement may be quite limited (e.g., simply listening) or quite involved (e.g., providing detailed answers to questions or engaging in highly interactive conversation). The nature of the auditory material may have significantly effects on driver performance. For instance, spatial information such as directions to a waypoint may intrude into the driving task relatively more than strictly verbal information such as a weather report. Finally, such an interaction, unlike those of other in-cab devices, may be measured in minutes or portions of an hour rather than in seconds. Because significantly less is currently known about this class of in-cab tasks than the tasks listed above, it has not been integrated into Table 3-2 to be presented next. However, this is considered an important class of in-cab device tasks which deserve serious consideration in protocol development.

Development of a Sensitivity Matrix

Over 50 driver/vehicle/roadway system measures have been defined that may be sensitive to in-vehicle workload. The next issue then is what one might wish to use in the workload assessment protocol for this project. This subsection describes the extent of likely sensitivity of specific types of measures to specified types of loading. Nine benchmark tasks having varying degrees of visual, manual, and cognitive load are described above. The set of measures and the set of tasks can now be presented in a matrix format that describes likely sensitivity; it is called a sensitivity matrix and is shown in Table 3-2.

The key to the matrix is as follows:

- i. A ✓✓+ symbol designates a measure that is likely to be sensitive to the specified benchmark task, and the measure has prima face safety relevance.
- ii. A ✓✓ symbol designates a measure that is likely to be sensitive to a specified benchmark task, but the measure does not have prima face safety relevance.
- iii. A ✓+ symbol designates a measure that may be sensitive in some situations but not others, or the measure may be sensitive for some drivers but not others. The measure has prima face safety relevance.
- iv. A ✓ symbol designates a measure that may be sensitive in some situations but not others, or the measure may be sensitive for some drivers but not others. The measure does not have prima face safety relevance.

- v. A blank space is used to designate a measure that is unlikely to be sensitive to the specified benchmark task.

The key designates sensitivity by checks and designates safety relevance by pluses. If a measure is unlikely to be sensitive, its safety relevance is of no consequence. Thus, safety relevance is only included for measures that are likely to possess a degree of sensitivity.

The key can be summarized as shown below:

		Safety Relevance	
		YES	NO
SENSITIVITY	LIKELY	✓✓+	✓✓
	POSSIBLE	✓+	✓
	UNLIKELY	(blank)	

This matrix was derived by one of the authors (WWW) based on extensive experience and familiarity with such measures. The matrix reflects an expert judgement of what measures are likely to correlate well with in-cab device loads. However, the literature review represented in Appendix 3A lends support to the predictions of sensitivity indicated in the matrix. Furthermore, the safety relevance of many measures is often apparent from inspection. Thus, this sensitivity matrix provides useful guidance on an important aspect of protocol development and will be validated during subsequent tasks of this project.

TABLE 3-2. LIKELY SENSITIVITY OF DRIVER-VEHICLE PERFORMANCE MEASURES TO PROTOTYPIC IN-CAB TASKS

TASKS: MEASURES	TASK 1: V.O., LO COG SCAN OF GAGES	TASK 2: V.O., HI COG READ TEXT FROM DISPATCH DISPLAY	TASK 3: V.P., LO COG ACCESS MENU DISPLAY & CHECK	TASK 4: V.P., HI COG PLAN ROUTE WITH NAV. DISPLAY	TASK 5: V.M., LO COG INSERT CASSETTE AND ADJUST	TASK 6: V.M., HI COG PLAN ALT. PARALLEL ROUTE WITH NAV. DISPLAY	TASK 7: M.P., LO COG TUNE RADIO BASED ON PROGRAM CONTENT	TASK 8: M.P., HI COG DISPATCHING AND ROUTING VIA 2-WAY RADIO	TASK 9: M.O., LO COG SET TURN SIGNAL
I. LANE RELATED MEASURES									
MEAN LANE POSITION	✓+	✓✓+	✓+	✓✓+	✓+	✓✓+	✓+		
LANE STANDARD DEVIATION	✓+	✓✓+	✓+	✓✓+	✓+	✓✓+	✓+		
LANE RMS DEVIATION	✓+	✓✓+	✓+	✓✓+	✓+	✓✓+	✓+		
PEAK LANE DEVIATION	✓+	✓✓+	✓+	✓✓+	✓+	✓✓+	✓+		
LANE EXCEEDENCES		✓+		✓+	✓+	✓+			
MEAN LANE EXCEED. DUR.				✓+	✓+	✓+			
II. HEADING/ACCEL. REL. MEASURES									
YAW STANDARD DEVIATION	✓+	✓+	✓+	✓+	✓+	✓+	✓+		
YAW RATE STD. DEVIATION	✓+	✓+	✓+	✓+	✓+	✓+	✓+		
LAT. ACC. STD. DEVIATION	✓+	✓✓+	✓+	✓✓+	✓+	✓✓+	✓+		
PEAK LATERAL ACCEL.	✓+	✓✓+	✓+	✓✓+	✓+	✓✓+	✓+		
III. STEERING-RELATED MEASURES									
STEERING STD. DEVIATION									
PEAK STEER. DEFLECTION	✓+	✓+	✓+	✓+	✓+	✓+	✓+		
STEER. VEL. STD. DEVIATION		✓	✓	✓	✓	✓	✓		
STEERING REVERSALS									
LARGE STEER. REVERSALS	✓+	✓+	✓+	✓+	✓+	✓+	✓+	✓	
SMALL STEER. REVERSALS		✓		✓	✓	✓	✓	✓	
STEERING HOLDS	✓	✓	✓	✓	✓	✓			
STEERING ZERO CROSSINGS									
*STEER. RESP. TIME	✓+	✓✓+	✓+	✓✓+	✓+	✓✓+			

* Stimulus is assumed to be presented visually.

TABLE 3-2. LIKELY SENSITIVITY OF DRIVER-VEHICLE PERFORMANCE MEASURES TO PROTOTYPIC IN-CAB TASKS (Continued)

TASKS: MEASURES	TASK 1: V.O., LO COG SCAN OF GAGES	TASK 2: V.O., HI COG READ TEXT FROM DISPATCH DISPLAY	TASK 3: V.P., LO COG ACCESS MENU DISPLAY & CHECK	TASK 4: V.P., HI COG PLAN ROUTE WITH NAV. DISPLAY	TASK 5: V.M., LO COG INSERT CASSETTE AND ADJUST	TASK 6: V.M., HI COG PLAN ALT. PARALLEL ROUTE WITH NAV. DISPLAY	TASK 7: M.P., LO COG TUNE RADIO BASED ON PROGRAM CONTENT	TASK 8: M.P., HI COG DISPATCHING AND ROUTING VIA 2-WAY RADIO	TASK 9: M.O., LO COG SET TURN SIGNAL
IV. ACCELERATOR REL. MEASURES									
ACCEL. STD. DEVIATION									
ACCELERATOR REVERSALS		J	J	J	J	J	J		
LARGE ACCEL. REVS.		J	J	J	J	J	J		
SMALL ACCEL. REVS.		J	J	J	J	J	J		
ACCELERATOR HOLDS	J	J	J	J	J	J	J		
MEAN ACCEL. HOLD. DUR.	J	J	J	J	J	J	J		
TOTAL ACCEL. HOLD TIME	J	J	J	J	J	J	J		
ACCELERATOR RELEASES		J	J	J	J	J	J		
MEAN ACCEL. RELEASE DUR.									
TOTAL ACCEL. RELEASE TIME		J	J	J	J	J	J		
V. BRAKE/DECEL. REL. MEASURES									
VI. SPEED RELATED MEASURES									
MEAN SPEED ERROR									
SPEED STD. DEVIATION									
SPEED RMS DEVIATION									
VII. VISION-RELATED MEASURES									
MEAN RDWAY GLANCE DUR.	J	J	J	J	J	J	J	J	
NUM. OF RDWAY GLANCES	J	J	J	J	J	J	J	J	
TOTAL RDWAY GLANCE TIME	J	J	J	J	J	J	J	J	

* Stimulus is assumed to be presented visually.

TABLE 3-2. LIKELY SENSITIVITY OF DRIVER-VEHICLE PERFORMANCE MEASURES TO PROTOTYPIC IN-CAB TASKS (Continued)

TASKS: MEASURES	TASK 1: V.O., LO COG SCAN OF GAGES	TASK 2: V.O., HI COG READ TEXT FROM DISPATCH DISPLAY	TASK 3: V.P., LO COG ACCESS MENU DISPLAY & CHECK	TASK 4: V.P., HI COG PLAN ROUTE WITH NAV. DISPLAY	TASK 5: V.M., LO COG INSERT CASSETTE AND ADJUST	TASK 6: V.M., HI COG PLAN ALT. PARALLEL ROUTE WITH NAV. DISPLAY	TASK 7: M.P., LO COG TUNE RADIO BASED ON PROGRAM CONTENT	TASK 8: M.P., HI COG DISPATCHING AND ROUTING VIA 2-WAY RADIO	TASK 9: M.O., LO COG SET TURN SIGNAL
VIII. MANUAL-RELATED MEASURES									
HANDS-ON-WHEEL OCCURR.	✓	✓	✓	✓	✓+	✓+	✓	✓	✓
MEAN HANDS-ON-WHEEL DUR.	✓	✓	✓	✓	✓+	✓+	✓	✓	✓
TOT. HANDS-ON-WHEEL TIME	✓	✓	✓	✓	✓+	✓+	✓	✓	✓
IX. VEHICLE-FOLLOWING-REL. MEAS.									
FOLLOWING DIST. MEAN ERROR	✓+	✓+	✓+	✓+	✓+	✓+	✓	✓	✓
FOLLOWING DIST. STD. DEV.	✓+	✓+	✓+	✓+	✓+	✓+	✓	✓	✓
FOLLOWING DIST. RMS ERROR	✓+	✓+	✓+	✓+	✓+	✓+	✓	✓	✓
MIN. FOLL. DISTANCE	✓+	✓+	✓+	✓+	✓+	✓+	✓	✓	✓
FOLL. DIST. PK. CLOS. VEL.	✓+	✓+	✓+	✓+	✓+	✓+	✓	✓	✓
X. OBJECT/OBST. REL. MEAS.									
MINIMUM MISS DISTANCE	✓+	✓+	✓+	✓+	✓+	✓+	✓	✓	✓
PEAK CLOSING VELOCITY	✓+	✓+	✓+	✓+	✓+	✓+	✓	✓	✓
XI. STIMULUS RESPONSE MEAS.									
*RESPONSE TIME TYPE I	✓+	✓+	✓+	✓+	✓+	✓+	✓	✓	✓
*RESPONSE TIME TYPE II	✓+	✓+	✓+	✓+	✓+	✓+	✓	✓	✓
*RESPONSE DISTANCE	✓+	✓+	✓+	✓+	✓+	✓+	✓	✓	✓
*ERRORS OF OMISSION	✓+	✓+	✓+	✓+	✓+	✓+	✓	✓	✓
*ERRORS OF COMMISSION	✓+	✓+	✓+	✓+	✓+	✓+	✓	✓	✓

* Stimulus is assumed to be presented visually.

References

- Allen, R. W., Jex, H. R., McRuer, D. T. and DiMarco, R. J. (1975). Alcohol effects on driving behavior and performance in a car simulator. IEEE Transactions on Systems, Man, and Cybernetics, SMC-5, 498-505.
- Blaauw, G. J., Godthelp, J. & Moraal, J. (1977). Driver's lateral control strategy as affected by task demands and driving experience. Society of Automotive Engineers, Paper no. 770876.
- Brouwer, W. H., Waterink, W., Van Wolffelaar, P. C., & Rothengatter, T. (1991). Divided attention in experienced young and older drivers: lane tracking and visual analysis in a dynamic driving simulator. Human Factors, 33(5), 573-582.
- Brown, I. D. (1962). Measuring the 'spare mental capacity' of car drivers by a subsidiary auditory task. Ergonomics, 5, 247-250.
- Brown, I. D. (1965). A comparison of two subsidiary tasks used to measure fatigue in car drivers. Ergonomics, 8, 467-473.
- Brown, I. D. (1966). Subjective and objective comparisons of successful and unsuccessful trainee drivers. Ergonomics, 9, 49-56.
- Brown, I. D., & Poulton, E. C. (1961). Measuring the spare 'mental capacity' of car drivers by a subsidiary task. Ergonomics, 4, 35-40.
- Ceder, A. (1977). Drivers' eye movements as related to attention in simulated traffic flow conditions. Human Factors, 19, 571-581.
- Chiles, W. D., & Alluisi, E. A. (1979). On the specification of operator or occupational workload with performance-measurement methods. Human Factors, 21, 525-528.
- Chiles, W. D., & Jennings, A. E. (1970). Effects of alcohol on complex performance. Human Factors, 12(6), 605-612.
- Colbourn, C. J., Brown, I. D., & Copeman, A. K. (1978). Driver's judgements of safe distances in vehicle following. Human Factors, 20(1), 1-12.
- Dewar, R. E., Ells, J. G., & Mundy, G. (1976). Reaction time as an index of traffic sign perception. Human Factors, 18(4), 381-392.
- Dingus, T. A., Antin, J. F., Hulse, M. C., & Wierwille, W. W. (1989). Attentional demand requirements of an automobile moving-map navigation system. Transportation Research A, 23(A), 301-315.
- Dingus, T. A., Hardee, L., & Wierwille, W. W. (1987). Development of models for on-board detection of driver impairment. Accident Analysis and Prevention, 19(4), 271-283.

- Drory, A. (1985). Effects of rest and secondary task on simulated truck-driving task performance. Human Factors, 27(2), 201-207.
- Farber, E. & Gallagher, V. (1972). Attentional demand as a measure of the influence of visibility conditions on driving task difficulty. Highway Research Record, 414, 1-5.
- Godthelp, H. (1986). Vehicle control during curve driving. Human Factors, 28(2), 211-221.
- Godthelp, H., & Kappler, W-D. (1988). Effects of vehicle handling characteristics on driving strategy. Human Factors, 30(2), 219-229.
- Godthelp, H., Milgram, P., & Blaauw, G. J. (1984). The development of a time-related measure to describe driving strategy. Human Factors, 26(3), 257-268.
- Hancock, P. A. & Chignell, M. H. (1988). Mental workload dynamics in adaptive interface design. IEEE Transactions on Systems, Man, and Cybernetics, 18, 647-657.
- Helander, M. G. (1975). Physiological reactions of drivers as indicators of road traffic demand. In Driver performance studies: Transportation Research Record 530 (Report No. TRB/TRR-530) (pp 1-17). Washington, DC: U. S. Transportation Research Board.
- Hicks, T. G., & Wierwille, W. W. (1979). Comparison of five mental workload assessment procedures in a moving-base driving simulator. Human Factors, 21, 129-143.
- Hulse, M. C., Dingus, T. A., Fischer, T., & Wierwille, W. W. (1989). The influence of roadway parameters on driver perception of attentional demand. In A. Mital (Ed.), Advances in Industrial Ergonomics and Safety I. (pp. 451-456). Taylor and Francis.
- Imbeau, D., Wierwille, W. W., Wolf, L. D., & Chun, G. (1989). Effects of instrument panel luminance and chromaticity on reading performance and preference in simulated driving. Human Factors, 31(2), 147-160.
- Johansson, G., & Rumar, K. (1971). Drivers' brake reaction times. Human Factors, 13(1), 23-27.
- Korteling, J. E. (1990). Perception-response speed and driving capabilities of brain-damaged and older drivers. Human Factors, 32(1), 95-108.
- Kurokawa, K. & Wierwille, W. W. (1991). Effects of instrument panel clutter and control labelling on visual demand and task performance. Plaza del Rey, CA: Society for Information Display, SID International Symposium Digest of Papers, Vol. XXII, pp. 99-102.
- MacDonald, W. A., & Hoffmann, E. R. (1980). Review of relationships between steering wheel reversal rate and driving task demand. Human Factors, 22(6), 733-739.
- Mathews, M. L. (1986). The influence of visual workload history on visual performance. Human Factors, 28, 623-632.

- McDonald, L. B. & Ellis, N. C. (1975). Driver work load for various turn radii and speeds. In Driver performance studies: Transportation Research Record 530 (Report No. TRR 530) (pp. 18-30). Washington, D.C.: U. S. Transportation Research Board.
- McLean, J. R., Hoffmann, E. R. (1975). Steering reversals as a measure of driver performance and steering task difficulty. Human Factors, 17, 248-256.
- Mourant, R. R., Herman, M., & Moussa-Hamouda, E. (1980). Direct looks and control location in automobiles. Human Factors, 22(4), 417-425.
- Mourant, R. R., & Rockwell, T. H. (1970). Mapping eye-movement patterns to the visual scene in driving: An exploratory study. Human Factors, 12, 81-87.
- Muto, W. H., & Wierwille, W. W. (1982). The effect of repeated emergency response trials on performance during extended-duration simulated driving. Human Factors, 24(6), 693-698.
- Noy, Y. I. (1987). Theoretical review of the secondary task methodology for evaluating intelligent automobile displays. Proceedings of the Human Factors Society 31st Annual Meeting-1987, pp. 205-209.
- Noy, Y. I. (received Sept. 1991). Selective attention with auxiliary automobile displays. (in preparation)
- Noy, Y. I. (Feb. 1990). Attention and performance while driving with auxiliary in-vehicle displays. Transport Canada Report No. TP 10727(E).
- Repa, B. S., & Wierwille, W. W. (1976). Driver performance in controlling a driving simulator with varying vehicle response characteristics. Society of Automotive Engineers, Paper no. 760779.
- Rockwell, T. H. (1987). Spare visual capacity in driving revisited: New empirical results for an old idea. In A. G. Gale, et al. (Eds.) Vision in Vehicles II - Proceedings of the Second International Conference on Vision in Vehicles. Nottingham. UK. 14-17 Sept. 1987 (pp. 317-324). Amsterdam, Netherlands: Elsevier.
- Senders, J. W., Kristofferson, A. B., Levison, W. H., Dietrich, C. W., and Ward, J. L. (1967). The attentional demand of automobile driving. Highway Research Record, 195, 15-33.
- Simmons, R. R. (1979). Methodological considerations of visual workloads of helicopter pilots. Human Factors, 21, 353-367.
- Sivak, M., Post, D., Olson, P. L., & Donohue, R. J. (1981). Driver responses to high-mounted brake lights in actual traffic. Human Factors, 23(2), 231-235.
- Summala, H. (1981). Driver/vehicle steering response latencies. Human Factors, 23(6), 683-692.
- Sweatman, P., & Joubert, P. N. (1974). Detection of changes in automobile steering sensitivity. Human Factors, 16(1), 29-36.

- Venvey, W. B. (1990). Adaptable driver-car interacting and mental workload: A review of literature. Institute for Perception RVO-TNO, Netherlands. Report No. IZF-1990-B-3; TDCK-90-0016.
- Verwey, W. B. (1990). Adaptable driver-car interaction in the GIDS system; Guidelines and a preliminary design from a human factors point of view. Institute for Perception RVO-TNO, Netherlands. Report No. IZF-1990-B-12.
- Verwey, W. B. (1991). Towards guidelines for in-car information management: Driver workload in specific driving situations. Institute for Perception RVO-TNO, Netherlands. Report No. IZF-1991-C-13.
- Vidulich, M. A., & Wickens, C. D. (1986). Causes of dissociation between subjective workload measures and performance. Applied Ergonomics, 17, 4, 291-296.
- Wierwille, W. W., Antin, J. F., Dingus, T. A., & Hulse, M. C. (1988). Visual attention demand of an in-car navigation display system. In A. G. Gale, et al. (Eds.) Vision in Vehicles II - Proceedings of the Second International Conference on Vision in Vehicles, Nottingham, UK 14-17 Sept. 1987. Amsterdam, Netherlands: Elsevier.
- Wierwille, W. W., Casali, J. G., & Repa, B. S. (1983). Driver steering reaction time to abrupt-onset crosswinds, as measured in a moving-base driving simulator. Human Factors, 25(1), 103-116.
- Wierwille, W. W., & Connor, S. A. (1983). Evaluation of 20 workload measures using a psychomotor task in a moving-base simulator. Human Factors, 25, 1-16.
- Wierwille, W. W., & Gutmann, J. C. (1978). Comparison of primary and secondary task measures as a function of simulated vehicle dynamics and driving conditions. Human Factors, 20, 233-244.
- Wierwille, W. W., Guttman, J. C., Hicks, T. G., & Muto, W. H. (1977). Secondary task measurement of workload as a function of simulated vehicle dynamics and driving conditions. Human Factors, 19, 557-565.
- Wierwille, W. W., & Peacock, B. (1988). Human factors technology. Automotive Engineering, 96-101.
- Wierwille, W. W., Hulse, M. C., Fischer, T. J., & Dingus, T. A. (1988). Strategic use of visual resources by the driver while navigating with an in-car navigation display system. XXII FISITA Congress Technical Papers: Automotive Systems Technology: The Future, Vol. II, SAE-P-211. Warrendale, PA: Society of Automotive Engineers, pp. 2.661-2.675. (Paper No. 885180.)
- Zeier, H. (1979). Concurrent physiological activity of driver and passenger when driving with and without automatic transmission in heavy city traffic. Ergonomics, 22, 799-810.

Appendix 3A

Selective Review of Driver-Vehicle Performance Measures

Appendix A References

- Allen, R. W., Jex, H. R., McRuer, D. T. and DiMarco, R. J. (1975). Alcohol effects on driving behavior and performance in a car simulator. IEEE Transactions on Systems, Man, and Cybernetics, SMC-5, 498-505.
- Blaauw, G. J., Godthelp, J. & Moraal, J. (1977). Driver's lateral control strategy as affected by task demands and driving experience. Society of Automotive Engineers, Paper no. 770876.
- Brouwer, W. H., Waterink, W., Van Wolffelaar, P. C., & Rothengatter, T. (1991). Divided attention in experienced young and older drivers: lane tracking and visual analysis in a dynamic driving simulator. Human Factors, 33(5), 573-582.
- Brown, I. D. (1962). Measuring the 'spare mental capacity*' of car drivers by a subsidiary auditory task. Ergonomics, 5, 247-250.
- Brown, I. D. (1965). A comparison of two subsidiary tasks used to measure fatigue in car drivers. Ergonomics, 8, 467-473.
- Brown, I. D. (1966). Subjective and objective comparisons of successful and unsuccessful trainee drivers. Ergonomics, 9, 49-56.
- Brown, I. D., & Poulton, E. C. (1961). Measuring the spare 'mental capacity' of car drivers by a subsidiary task. Ergonomics, 4, 35-40.
- Ceder, A. (1977). Drivers' eye movements as related to attention in simulated traffic flow conditions. Human Factors, 19, 571-581.
- Chiles, W. D., & Alluisi, E. A. (1979). On the specification of operator or occupational workload with performance-measurement methods. Human Factors, 21, 525-528.
- Chiles, W. D., & Jennings, A. E. (1970). Effects of alcohol on complex performance. Human Factors, 12(6), 605-612.
- Colbourn, C. J., Brown, I. D., & Copeman, A. K. (1978). Driver's judgements of safe distances in vehicle following. Human Factors, 20(1), 1-12.
- Dewar, R. E., Ells, J. G., & Mundy, G. (1976). Reaction time as an index of traffic sign perception. Human Factors, 18(4), 381-392.
- Dingus, T. A., Antin, J. F., Hulse, M. C., & Wierwille, W. W. (1989). Attentional demand requirements of an automobile moving-map navigation system. Transportation Research A, 23(A), 301-315.
- Dingus, T. A., Hardee, L., & Wierwille, W. W. (1987). Development of models for on-board detection of driver impairment. Accident Analysis and Prevention, 19(4), 271-283.

- Drory, A. (1985). Effects of rest and secondary task on simulated truck-driving task performance. Human Factors, 27(2), 201-207.
- Farber, E. & Gallagher, V. (1972). Attentional demand as a measure of the influence of visibility conditions on driving task difficulty. Highway Research Record, 414, 1-5.
- Godthelp, H. (1986). Vehicle control during curve driving. Human Factors, 28(2), 211-221.
- Godthelp, H., & Kappler, W-D. (1988). Effects of vehicle handling characteristics on driving strategy. Human Factors, 30(2), 219-229.
- Godthelp, H., Milgram, P., & Blaauw, G. J. (1984). The development of a time-related measure to describe driving strategy. Human Factors, 26(3), 257-268.
- Hancock, P. A. & Chignell, M. H. (1988). Mental workload dynamics in adaptive interface design. IEEE Transactions on Systems, Man, and Cybernetics, 18, 647-657.
- Helander, M. G. (1975). Physiological reactions of drivers as indicators of road traffic demand. In Driver performance studies: Transportation Research Record 530 Washington, D.C.: U. S. Transportation Research Board, Technical Report TRB/TRR-530: 1-17.
- Hicks, T. G., & Wierwille, W. W. (1979). Comparison of five mental workload assessment procedures in a moving-base driving simulator. Human Factors, 21, 129-143.
- Hulse, M. C., Dingus, T. A., Fischer, T., & Wienville, W. W. (1989). The influence of roadway parameters on driver perception of attentional demand. In A. Mital (Ed.), Advances in Industrial Ergonomics and Safety I. Taylor and Francis, pp. 451-456.
- Imbeau, D., Wierwille, W. W., Wolf, L. D., & Chun, G. (1989). Effects of instrument panel luminance and chromaticity on reading performance and preference in simulated driving. Human Factors, 31(2), 147-160.
- Johansson, G., & Rumar, K. (1971). Drivers' brake reaction times. Human Factors, 13(1), 23-27.
- Korteling, J. E. (1990). Perception-response speed and driving capabilities of braindamaged and older drivers. Human Factors, 32(1), 95-108.
- Kurokawa, K. & Wierwille, W. W. (1991). Effects of instrument panel clutter and control labeling on visual demand and task performance. Plaza del Rey, CA: Society for Information Display, SID International Symposium Digest of Papers, Vol. XXII, pp. 99-102.
- MacDonald, W. A., & Hoffmann, E. R. (1980). Review of relationships between steering wheel reversal rate and driving task demand. Human Factors, 22(6), 733-739.
- Mathews, M. L. (1986). The influence of visual workload history on visual performance. Human Factors, 28, 623-632.

- McDonald, L. B. & Ellis, N. C. (1975). Driver work load for various turn radii and speeds. In Driver performance studies: Transportation Research Record 530 Washington, D.C.: U. S. Transportation Research Board, Technical Report TRR 530, 18-30.
- McLean, J. R., Hoffmann, E. R. (1975). Steering reversals as a measure of driver performance and steering task difficulty. Human Factors, 17,248-256.
- Mourant, R. R., Herman, M., & Moussa-Hamouda, E. (1980). Direct looks and control location in automobiles. Human Factors, 22(4), 417-425.
- Mourant, R. R., & Rockwell, T. H. (1970). Mapping eye-movement patterns to the visual scene in driving: An exploratory study. Human Factors, 12, 81-87.
- Muto, W. H., & Wierwille, W. W. (1982). The effect of repeated emergency response trials on performance during extended-duration simulated driving. Human Factors, 24(6), 693-698.
- Noy, Y. I. (1987). Theoretical review of the secondary task methodology for evaluating intelligent automobile displays. Proceedings of the Human Factors Society-31st Annual Meeting-1987, pp. 205-209.
- Noy, Y. I. (received Sept. 1991). Selective attention with auxiliary automobile displays, (in preparation)
- Noy, Y. I. (Feb. 1990). Attention and performance while driving with auxiliary in-vehicle displays. Transport Canada Report No. TP 10727(E).
- Repa, B. S., & Wierwille, W. W. (1976). Driver performance in controlling a driving simulator with varying vehicle response characteristics. Society of Automotive Engineers, Paper no. 760779.
- Rockwell, T. H. (1988). Spare visual capacity in driving revisited. New empirical results for an old idea. Gale, A. G., et al. (Eds.) Vision in Vehicles II - Proceedings of the Second International Conference on Vision in Vehicles. Nottingham. UK. 14-17 Sept. 1987. Amsterdam, Netherlands: Elsevier. pp. 317-324.
- Senders, J. W., Kristofferson, A. B., Levison, W. H., Dietrich, C. W., and Ward, J. L. (1967). The attentional demand of automobile driving. Highway Research Record, 195, 15-33.
- Simmons, R. R. (1979). Methodological considerations of visual workloads of helicopter pilots. Human Factors, 21, 353-367.
- Sivak, M., Post, D. B., Olson, P. L., & Donohue, R. J. (1981). Driver responses to high-mounted brake lights in actual traffic. Human Factors, 23(2), 231-235.
- Summala, H. (1981). Driver/vehicle steering response latencies. Human Factors, 23(6), 683-692.
- Sweatman, P., & Joubert, P. N. (1974). Detection of changes in automobile steering sensitivity. Human Factors, 16(1), 29-36.

- Venvey, W. B. (1990). Adaptable driver-car interacting and mental workload: A review of literature. Institute for Perception RVO-TNO, Netherlands. Report No. IZF-1990-B-3; TDCK-904016.
- Venvey, W. B. (1990). Adaptable driver-car interaction in the GIDS system; Guidelines and a preliminary design from a human factors point of view. Institute for Perception RVO-TNO, Netherlands. Report No. IZF-1990-B-12.
- Verwey, W. B. (1991). Towards guidelines for in-car information management: Driver workload in specific driving situations. Institute for Perception RVG-TNO, Netherlands. Report No. IZF-1991-C-13.
- Vidulich, M. A., & Wickens, C. D. (1986). Causes of dissociation between subjective workload measures and performance. Applied Ergonomics, 17,4, 291-296.
- Wierwille, W. W., Antin, J. F., Dingus, T. A., & Hulse, M. C. (1988). Visual attention demand of an in-car navigation display system. In Gale, A. G. et al. Vision in Vehicles II, North Holland: Elsevier Science Publishers B. V.
- Wierwille, W. W., Casell, J. G., & Repa, B. S. (1983). Driver steering reaction time to abrupt-onset crosswinds, as measured in a moving-base driving simulator. Human Factors, 25(1), 103-116.
- Wierwille, W. W., & Connor, S. A. (1983). Evaluation of 20 workload measures using a psychomotor task in a moving-base simulator. Human Factors, 25,1-16.
- Wierwille, W. W., & Gutmann, J. C. (1978). Comparison of primary and secondary task measures as a function of simulated vehicle dynamics and driving conditions. Human Factors, 233-244.
- Wierwille, W. W., Guttman, J. C., Hicks, T. G., & Muto, W. H. (1977). Secondary task measurement of workload as a function of simulated vehicle dynamics and driving conditions. Human Factors, 19,557-565.
- Wierwille, W. W., & Peacock, B. (1988). Human factors technology. Automotive Engineering, 96-101.
- Wierwille, W. W., Hulse, M. C., Fischer, T. J., & Dingus, T. A. (1988). Strategic use of visual resources by the driver while navigating with an in-car navigation display system. XXII FISITA Congress Technical Papers: Automotive Systems Technology: The Future, Vol. II, SAE-P-211. Warrendale, PA: Society of Automotive Engineers, pp. 2.661-2.675. (Paper No. 885180.)
- Zeier, H. (1979). Concurrent physiological activity of driver and passenger when driving with and without automatic transmission in heavy city traffic. Ergonomics, 22, 799-810.

4.0 RISK TAKING AND DRIVER ADAPTATION IN THE USE OF HIGH-TECHNOLOGY IN-CAB DEVICES

Introduction

Over the past three decades, the role of driver risk taking in highway safety has been the subject of much study and debate in the human factors community. During this period, researchers have attempted to relate risk perception as a primary determinant of driver behavior leading to the occurrence of traffic crashes. These efforts have been explored with the intent of identifying the mechanisms for crash causation and methods for reducing the frequency and severity of crashes.

Few people would dispute that driving on today's highways involves some element of risk. In a highway driving context, one can define risk as the likelihood of being involved in a traffic crash with its associated losses that include property damage, injury or death.

Before proceeding with a discussion of risk taking in driving, it is important to distinguish between two types of risk. Objective risk is actuarial in nature. It refers to the probability of being involved in a crash under given driving conditions. Objective risk can be quantified by an examination of traffic crash data for a selected site, a specified set of conditions (i.e., highway, weather, lighting, traffic, etc.) and over some defined unit of exposure.

Subjective risk is psychological in nature. It refers to an individual driver's estimate of the likelihood of being involved in a crash as the result of some action taken by the driver. Subjective risk is the perceived risk, usually consciously assessed by a driver. It is influenced by several external factors which include the vehicle state, highway state, environmental conditions, and traffic conditions in effect at the time the assessment is made. Internal factors, such as fatigue, alcohol, motivations, attitudes, and past experiences, can also influence perceived risk. Because of the combined effects of these factors, perceived risk may or may not equal the objective risk for a situation encountered by the driver. Further, two drivers may also not perceive the same risk when faced with identical situations.

Subjective **risk also** implies that there is uncertainty about the outcome of the action taken by the driver. That is, when making a decision involving the assessment of risk in a given situation, the driver does not know whether a favorable outcome will result from a decision. Risk takers may perceive the outcome as less uncertain than non-risk takers, the payoff as sufficiently high to make the risk worthwhile, or both.

It is this latter type of risk (i.e., perceived or subjective risk) that is of primary interest in the study of heavy vehicle driver workload. With the introduction of high technology electronic devices into the cabs of heavy vehicles, there is a concern on the part of NHTSA that use of these devices could potentially lead to increased driver risk taking. Such risk-taking behavior might be exhibited by an increased willingness to engage in risky driving maneuvers in response to the improved safety benefits offered by the device or to attend to in-cab device at an inopportune time. In turn, increased risk-taking may result in a net reduction in highway safety despite the projected benefits of a device.

Before considering the subject of risk-taking it is worth noting that risk and risk adaptation are complex topics. There are no universally agreed upon definitions, theories of risk-taking, or methods

for assessing the effects of risk-taking on driver behavior within the research community. With this in mind, the following sections present a selective review of the literature on driver risk taking and its implications for the heavy vehicle driver workload project.

Driver Risk Taking Research

Concerns over driver adaptation to safety improvements in vehicles is not a new concept. In the late 1930s, researchers (Gibson and Crooks, 1938) suggested that improved braking systems on cars would not lead to increased safety because drivers would soon learn to modify their braking behavior in response to the shorter stopping distances afforded by the improved brakes.

In recent years, a number of theories have been proposed to describe the relationship between risk taking, driver behavior and highway safety. These theories are known, generally, as risk compensation or risk adaptation models. They include, for example, Peltzman (1975), Naatanen and Summala (1976), Fuller (1984), and Adams (1983). While they espouse some variations, they are common in the view that drivers consciously assess perceived risk while driving and alter their behavior in response to changes in perceived risk. Usually, drivers execute behaviors to reduce risk. For example, they slow down in fog.

When a new safety innovation is introduced into the vehicle-highway system, the driver, through direct and observable feedback from the innovation, may perceive a reduction in the level of driving risk. The theory proposes that the drivers may subsequently alter their behavior relative to their current level of desired risk, offsetting some of the benefits of the innovation in the process. For example, drivers may select higher cruising speeds when protected by air bags. In these theories, the level of risk the driver is willing to accept changes over time in response to factors both internal to the driver (e.g., motivation, attitude, stress) and external (traffic, weather, etc.). The driver is said to compensate for the reduced perception of risk by changing driving behavior.

The risk homeostasis theory, proposed by Wilde (1982), is a special case of the compensatory models. This theory postulates that an individual will maintain a more or less constant level of risk over time. When the individual's perceived level of risk differs from their desired level, the individual will alter behavior to realign the perceived risk with the desired risk. Under this model, the safety benefits of an innovation introduced into the vehicle-highway system will, over time, be completely negated by increased risk taking on the part of the driver.

Numerous attempts to validate risk theories have been undertaken through the years, producing inconclusive results. Taylor (1964), Rumar et al. (1976) and Peltzman (1975), for example, report data that the authors concluded supported the theory of risk compensation. Others, including McKenna (1988), O' Neill et al. (1985), and Evans (1985) cite data which lead to the opposite conclusion. Evans (1984), in a summary report, identifies data that are neutral, supportive, and contrary to the theory. Thus, after 20 some years of study and vehement debate, the issue of risk compensation and its role in determining driver behavior still has not been resolved. There is no agreement on which performance measures are indicants of risk taking. Lund and O' Neill (1986), for example, contemplate whether drivers consciously assess risk in every maneuver they execute while driving.

The major conclusion illustrated by this discussion is that risk taking is, at this time, seen as a concept whose use must be thoughtfully applied when assessing the effects of novel in-cab devices on highway safety. Several reasons for this exist. Risk-adaptation theory is applicable only to the introduction of safety related devices, as only this class of device has the potential to produce benefits that could lead to changes in a driver's perceived level of driving risk and, subsequently, lead to changes in driving behavior. If the new high technology in-cab device is not a safety device (e.g., a satellite tracking system or driver's electronic log book), then increased risk-taking due to the device should not be expected on the basis of risk-adaptation theory. This prediction results because the non-safety related device could not produce benefits which would influence the driver's perceived level of risk. Therefore, no change in behavior due to the risk compensatory mechanisms of the driver would result. This is not to say that the driver's behavior would remain unchanged, but only that observed behavior change could not be caused by increased risk taking.

Moreover, the concept is of limited use because risk-taking is one of many intervening variables that effect a driver's decision making process. Whether the behavioral changes introduced by a device were caused by risk taking, or one of the other factors (e.g., driver motivation, distraction, fatigue), is really a moot point. The pertinent issues to be considered here are:

- How the introduction of new devices affect driver behavior.
- Whether the behavioral changes have adverse implications for traffic safety.

For these reasons, the remaining discussion will focus on potential driver behavioral changes brought about by the introduction of new in-cab devices.

Risk Adaptation Effects with the Use of New In-Cab Devices

Without knowing the specific characteristics of a new in-cab device (i.e., the display design, the nature of warnings, the control activations required or the cognitive loads imposed on the user), it is difficult to discuss the range and extent of the behavioral changes one might expect from heavy vehicle drivers using a device. As a consequence, discussions included here are presented in general terms. Examples of how behavior might change in drivers are provided to illustrate points raised.

If one considers the introduction of a device into the cab, such as an IVHS headway detection and warning system, it is reasonable to expect some driver adaptation, i.e., behavior change on the part of the driver. With the safety benefits provided by this device, one might expect to see changes in the primary driving measures at the "maneuver" level of driving. The primary measures include the vehicle's speed, lane position, headway to the vehicle ahead, and control inputs made by the driver. In the current example, one might hypothesize that the new device will lead drivers to visually sample the road scene less often now that an automated "sentry" is available. The driver may follow at shorter headways or overtake at higher relative velocities if the driver perceives that there is increased margin for error available with the headway monitoring system. Shorter headways and increased closing velocity are not necessarily unsafe driving practices. But if the time headways fell to intervals shorter than, say, 2 seconds, one could conclude that the device introduced a behavioral change with adverse consequences on safety.

The introduction of a near object detection and warning system into heavy trucks serves to illustrate another type of behavioral change that might occur. Before the introduction of the system, drivers

could exhibit some baseline level of lane keeping ability (i.e., average lateral position and variability over time). After the introduction of this warning system, one might speculate that the drivers would reduce their lane keeping precision, relying instead on the warning system's ability to monitor and advise on lane excursions in the presence of traffic in the adjacent lanes. Again, such changes do not guarantee accidents but they presumably increase the likelihood of accidents, especially if the automation malfunctions.

One might also hypothesize that mirror sampling prior to a lane change would decrease. If, after the introduction of the device, drivers were observed to allocate fewer glances or less total time to the left and right mirrors than before, then one's initial reaction might be that the device introduced an undesirable change in visual search behavior. But if the glances in question were reallocated to the road scene ahead, the change might not represent unsafe driving, especially if the detection system were functioning properly. Another concern regarding the above example might be whether the change in the distribution of mirror glances represents a bad habit that would not be shed by the driver if the near object detection system become inoperative, or if the driver moved from a suitably equipped vehicle to an unequipped vehicle.

A second type of behavioral change can also be postulated with the use of a new in-cab device. Here, the driver could become more willing to interact inappropriately with the in-cab device while driving. Inappropriate interaction refers to using the device at an inopportune time (when attention should be focused on the road) or attending to the device for longer time intervals than are prudent for the driving conditions. In essence, the driver might mis-allocate driving resources (i.e., visual, manual, auditory and cognitive resources) to the in-cab device to the detriment of the driving task. The effect of the mis-allocation could ultimately lead to the increased incidence of traffic crashes. However, given previous discussions it does not appear beneficial to consider such misallocation to be risk taking per se, and so is not considered here.

A Proposed Methodology to Examine the Safety Criticality of In-Cab Devices

There are numerous hypotheses that can be proposed about driver adaptation to in-cab devices considering (a) the range of driving condition observed, (b) the driving maneuvers undertaken in truck driving and (c) the host of new devices projected for the future. It is clear that any methodology used to assess the impact of new devices on driving behavior must necessarily be broad-based and able to assess performance along many dimensions. The methodology would include a measurement system that would ideally record data about many of the primary driving behaviors including:

1. Vehicle speed.
2. Headway to the vehicle ahead or to the rear, as appropriate for the driving maneuver.
3. Relative velocity between vehicles (front and rear).
4. Lateral lane position.
5. Braking time-locations in response to traffic control devices.
6. Control activations by the driver (i.e., steering wheel, clutch, brake pedal and accelerator).

Knowledge of such measures would enable average value performance to be established as well as the variability of performance.

Measure	Comments	Reference
Mean-Lane Position	<p>Authors found this measure to be sensitive to the instructions given to drivers (i.e. either concentrate on keeping a constant lateral position or to do this and report items seen on side of the road).</p> <p>Authors found that lane position did not significantly differ while driving a curve or straightaway either with or without a secondary visual task.</p> <p>In this study, mean lateral position was not sensitive to variations in imposed speed or the introduction of an occlusion task.</p> <p>Mean lateral position did not vary as a function of vehicle configuration (under- or over-steering) or vehicle speed (20, 40, 60, 80, and 100 km/h) in a straight-road lane keeping task.</p>	<p>Blaauw, G. J., Godthelp, J., & Moraal, J. (1977)</p> <p>McDonald, L. B., & Ellis, N. C. (1975)</p> <p>Godthelp, H., Milgram, P., & Blaauw, G. J. (1984)</p> <p>Godthelp, H., & Kappler, W-D. (1988)</p>
Lane Standard Deviation	<p>Using lateral deviation, authors discovered this technique to be less sensitive than yaw STD or steering reversals and subject ratings to workload manipulation, but more sensitive than a secondary task (reading digits), occlusion technique, and cardiac arrhythmia.</p> <p>Authors discovered that lateral STD was sensitive to workload manipulations such as two levels of wind-curvature disturbance and two levels of steering ratio, but not to two levels of vehicle response.</p> <p>Using driving simulator, author found that lane STD was significantly degraded by adding auxiliary task of visual search of a CRT display or by manipulating driving difficulty by reducing curve radii.</p> <p>In at least three studies Zwahlen and colleagues have used lateral STD as a driver performance measure for the auxiliary tasks of driving a) while vision is occasionally occluded, b) while reading text, and c) while operating a CRT touch-screen. No significance tests available to assess sensitivity of measure.</p> <p>Authors found this measure to be sensitive to the instructions given to drivers to either concentrate on keeping a constant lateral position or to also report items seen on side of the road.</p> <p>In this study, STD lateral position was not sensitive to imposed speed variations, but did differ with introduction of occlusion task.</p> <p>Lateral position STD's were larger when driving with partially occluded vision (LCD visor configuration) and/or at higher speeds during a straight-road lane keeping task.</p>	<p>Hicks, T. G., & Wierwille, W. W. (1979)</p> <p>Wierwille, W. W., & Gutmann, J. C. (1978)</p> <p>Noy, Y. I. (Feb. 1990)</p> <p>Zwahlen, H. T., Adams, C. C. Jr., & DeBald, D. P. (1988)</p> <p>Blaauw, G. J., Godthelp, J., & Moraal, J. (1977)</p> <p>Godthelp, H., Milgram, P., & Blaauw, G. J. (1984)</p> <p>Godthelp, H., & Kappler, W-D. (1988)</p>

Time to Lane Crossing (not included in Wierwille list)	Using driving simulator, author found that TLC was significantly degraded by adding an auxiliary task of visually searching a CRT display or by increasing driving difficulty by decreasing curve radii.	Noy, Y. I. (Feb. 1990)
	Authors evaluated TLC in an occlusion task with varying speed levels. Found it to correspond to driver's self-chosen occlusion times.	Godthelp, H., Milgram, P., & Blaauw, G. J. (1984)
	Authors discovered that in a straight-lane keeping task drivers produced shorter TLCs for over-steering but not under-steering vehicles relative to normals and that, with a fitted occlusion device, drivers did not look often enough to compensate for this shorter TLC.	Godthelp, H., & Kappler, W-D. (1988)
Lane RMS Deviation	Authors used this measure to investigate effects of various vehicle handling characteristics on driving performance during random and step wind gust disturbance tasks. Measure showed some sensitivity to various conditions.	Repa, B. S., & Wierwille, W. W. (1976)
Peak Lane Deviation	Peak lane position overshoot plus a variety of integral performance measures were used to assess effects of step wind gusts on driver performance while various vehicle handling characteristics were varied. Peak lane position overshoot was found to be a sensitive measure.	Repa, B. S., & Wierwille, W. W. (1976)
	Author examined lateral displacement as a function of time to respond to a car-door opening on the side of the road. In this and other experiments by same author he was able to determine the mean reaction time to respond to unforeseen driving events.	Summala, H. (1981)
Lane Exceedences (Lane Busts)	Number and duration of lane exceedences were found to be too few to allow differentiation of operator or road variables in a driving study, however, it provided a rank ordering of different tasks required of the driver. The results of this ordering were not consistent with other dependent variables.	Dingus, T. A., Antin, J. F., Hulse, M. C., & Wierwille, W. W. (1989)
Mean Lane Exceedence Duration (or duration ratio)	In simulator study, author found lane exceedence ratio to be sensitive to different levels of an auxiliary task (e.g. visual Sternberg task).	Noy, Y. I. (Feb. 1990)
	Authors used related measure of time-on-target defined as the time during which the whole car was in the right lane. Found that it discriminated between young and old drivers and the introduction of a secondary task.	Brouwer, W. H., Waterink, W., Van Wolfelaar, P. C., & Rothengatter, T. (1991)
Yaw Standard Deviation	Using yaw standard deviation, authors discovered this to be a sensitive measure of workload variation comparable to steering reversals and more sensitive than several other techniques including lateral deviation, subjective ratings, a secondary technique, and cardiac arrhythmia.	Hicks, T. G., & Wierwille, W. W. (1979)
	Authors discovered yaw STD to be sensitive to workload manipulations such as two levels of wind-curvature disturbance, two levels of steering ratio, two levels of vehicle response, and the intrusion of a digit-reading secondary task.	Wierwille, W. W., & Gutmann, J. C. (1978)

	For yaw angle RMS, authors used this measure to investigate effects of various vehicle handling characteristics on driving performance during random and step wind gust disturbance tasks. Measure showed some sensitivity to various conditions.	Repa, B. S., & Wierwille, W. W. (1976)
	Yaw deviation mean and yaw standard deviation were found to be sensitive to drowsiness manipulations.	Dingus, T. A., Hardee, L., & Wierwille, W. W. (1987)
Yaw-Rate Standard Deviation		
Standard Deviation Lateral Speed*	Authors found this measure to be sensitive to imposed speed variations and to inclusion of an occlusion task.	Godthelp, H., Milgram, P., & Blaauw, G. J. (1984)
	Authors found that a vehicle configured for over-steering produces relatively large lateral speed changes at higher speeds in a straight-road lane keeping task.	Godthelp, H., & Kappler, W-D. (1988)
Lateral Acceleration Standard Deviation		
Peak Lateral Acceleration	Authors found no effect of nighttime visibility simulated by neutral-density filtered goggles on this measure for driving a slalom course at an automatically controlled speed (30 or 45 mph).	Farber E., & Gallagher, V. (1972)
Steering Standard Deviation	Authors discovered steering wheel STD was sensitive to workload manipulations such as two levels of wind-curvature disturbance, two levels of steering ratio, two levels of vehicle response, and the intrusion of a digit-reading secondary task.	Wierwille, W. W., & Gutmann, J. C. (1978)
	Authors found this measure to be sensitive to the instructions given to drivers (i.e. either concentrate on keeping a constant lateral position or to do this and report items seen on side of the road). This measure also showed an interaction between longitudinal control instructions and driver experience.	Blaauw, G. J., Godthelp, J., & Moraal, J. (1977)
	STD of steering angle was found to be sensitive to imposed speed variations only when self-chosen occlusion task was added.	Godthelp, H., Milgram, P., & Blaauw, G. J. (1984)
	For the RMS of steering wheel angle, authors used this measure to investigate effects of various vehicle handling characteristics on driving performance during random and step wind gust disturbance tasks. Measure showed some sensitivity to various conditions.	Repa, B. S., & Wierwille, W. W. (1976)
	At higher speeds, a vehicle configured for over-steering resulted in a relatively larger steering wheel angle STD in a straight-road lane keeping task.	Godthelp, H., & Kappler, W-D. (1988)
Peak Steering Deflection		
Frequency Analysis of Steering Movements*	Authors found this measure to be sensitive to the instructions given to drivers (i.e. either concentrate on keeping a constant lateral position or to do this and report items seen on side of the road). This measure also was sensitive to longitudinal control instructions and driver experience.	Blaauw, G. J., Godthelp, J., & Moraal, J. (1977)

	<p>Authors compared several steering related measures derived from steering reversals including primary dominant frequency and, from spectral analysis, high frequency area above 0.4 Hz. In two expts. various factors were manipulated including lane width, speed, and criterion for a steering reversal (gap width). A gap of 0.5 to 0.7 deg was best. Authors claim that frequency measures are better viewed as reflecting control effort rather than absolute tracking performance. Conclude that steering reversal rates can be a valid measure if used with caution.</p>	McLean, J. R., & Hoffman, E. R. (1975)
	<p>Spectral analysis of steering wheel movements demonstrated that a vehicle configured for over-steering produces rapid and large steering wheel movements to maintain control in a straight-road lane keeping task.</p>	Goehelp, H., & Kappler, W-D. (1988)
Steering Velocity Mean*	<p>This measure was sensitive to manipulations of blood-alcohol level, drowsiness, and their combined effects.</p>	Dingus, T. A., Hardee, L., & Wierwille, W. W. (1987)
Steering Velocity Standard Deviation	<p>Authors found this measure to be sensitive to drowsiness manipulations and combined effects of drowsiness and alcohol.</p> <p>Authors found that most sensitive measure to changes in road types, traffic conditions, operator variables, and task manipulations was the variance of the time that the steering velocity was zero (arguing that attention demanding tasks induced the subject to hold the steering wheel constant).</p>	Dingus, T. A., Antin, J. F., Hulse, M. C., & Wierwille, W. W. (1989)
Steering Reversals	<p>Using a 2-deg steering reversal criterion, authors discovered this measure to be sensitive to workload manipulations such as two levels of wind-curvature disturbance, two levels of steering ratio, two levels of vehicle response, and the intrusion of a digit-reading secondary task.</p> <p>Using a 2-deg steering reversal criterion, authors discovered this to be a more sensitive measure of workload variation than several other techniques including lateral deviation, yaw deviation, subjective ratings, a secondary technique, and cardiac arrhythmia.</p> <p>Authors measured how far subjects would allow vehicle to drift before correcting and found no significant differences when a digit-reading secondary task was added.</p>	Wierwille, W. W., & Gutmann, J. C. (1978)
		Hicks, T. G., & Wierwille, W. W. (1979)
		McDonald, L. B., & Ellis, N. C. (1975)
	<p>Authors compared several steering related measures derived from steering reversals including primary dominant frequency and, from spectral analysis, high frequency area above 0.4 Hz. In two expts. various factors were manipulated including lane width, speed, and criterion for a steering reversal (gap width). A gap of 0.5 to 0.7 deg was best. Authors claim that frequency measures are better viewed as reflecting control effort rather than absolute tracking performance. Conclude that steering reversal rates can be a valid measure if used with caution.f</p>	McLean, J. R., & Hoffman, R. (1975)
	<p>Authors review use of steering wheel reversal rate as a driving performance measure and conclude that it is sensitive to high driving task demand subject to a few qualifications.</p>	MacDonald, W. A., & Hoffmann, E. R. (1980)

	Study demonstrated steering wheel reversal measure was sensitive to manipulations in a 7-hour simulated driving task of reducing fatigue and additions of various secondary tasks.	Drory, A. (1985)
Small/Large Steering Reversals	Author found that number of times the steering wheel was reversed sufficiently to affect the wheels was not a good predictor of success at a bus driving school.	Brown, I. D. (1966)
Steering Holds		
Steering Zero-Crossings		
Steering Response Time	Authors used this dependent variable to investigate several effects of wind gusts and vehicle characteristics. Found drivers respond more quickly when physical as well as visual cues are present to wind gusts and when locus of gusts is toward front of vehicle. However, vehicle yaw rate, steering sensitivity, and vehicle understeer do not effect reaction time to side gusts. In addition, time on task (30, 60, 150 min) and previous exposure to gusts had no appreciable effect.	Wierwille, W. W., Casali, J. G., & Repa, B. S. (1983)
Accelerator Standard Deviation	Authors found that accelerator position was not a sensitive measure to variations in instructed tasks performed in a moving vehicle.	Dingus, T. A., Antin, J. F., Hulise, M. C., & Wierwille, W. W. (1989)
Accelerator Reversals		
Accelerator Velocity*	Authors did not find that accelerator velocity was sensitive to differences in road types, traffic conditions, operator variables, or specific tasks asked of drivers.	Dingus, T. A., Antin, J. F., Hulise, M. C., & Wierwille, W. W. (1989)
Small/Large Accelerator Reversals		
Accelerator Holds		
Mean Accelerator Hold Duration		
Total Accelerator Hold Time		
Accelerator Releases		
Mean Accelerator Release Duration		
Total Accelerator Release Time		
Brake Applications	Author found that number and force of brakes was not a good predictor of success at a bus driving school.	Brown, I. D. (1966)
	Study did not find that number of brake applications to appearance of brake lights ahead in 7-hour simulated driving situation was sensitive to manipulations of rest length or various secondary tasks.	Drory, A. (1985)

	Authors found that adding a single-high-mounted brake light or dual-high-mounted brake lights increased chances of braking to a lead car's brake-signal presentations. No difference between these two systems however.	Sivak, M., Post, D., Olson, P. L., & Donohue, R. J. (1981)
	Authors did not find this measure to be sensitive to differences in road types, traffic conditions, operator variables, or specific tasks asked of drivers.	Dingus, T. A., Antin, J. F., Huise, M. C., & Wierwille, W. W. (1989)
Mean Brake Application Duration		
Total Brake Time		
Brake Response Time	Study did find that average brake reaction time to appearance of brake lights ahead in 7-hour simulated driving situation was sensitive to manipulations of rest length and to various secondary tasks.	Drory, A. (1985)
	Authors found that mean brake reaction time to a loud klaxon/buzzer was .66 sec when somewhat anticipated and .73 sec when unanticipated.	Johansson, G., & Rumar, K. (1971)
	Author investigated reaction time to brake when lead car lightly tapped brakes in normals, older drivers, and brain-damaged. Found increased RTs for brain-damaged relative to other groups. Brake reaction time was also sensitive to task load manipulation (straight vs. curved road) for all three groups.	Korteling, J. E. (1990)
	Authors discovered that brake reaction time to lead car brake lights increased from a pre-extended driving baseline to a post-extended baseline, however, it did not vary depending upon how long the extended driving situation lasted (i.e. 30, 60, or 150 min). Simulator study.	Muto, W. H., & Wierwille, W. W. (1982)
	Authors discovered no difference in average brake reaction time to a lead car braking with either regular two-back brake lights, the addition of a single-high-mounted light, or the addition of dual-high-mounted lights.	Sivak, M., Post, D., Olson, P. L., & Donohue, R. J. (1981)
Peak Longitudinal Deceleration		
Maximum Achievable Safe Speed*	Authors found that drivers could drive faster when the visor shutting off their visual field was down less.	Senders, J. W., Kristofferson, A. B., Levison, W. H., Dietrich, C. W., & Ward, J. L. (1967)
Mean Speed Error	Author found no substantial differences in average speed when drivers were fatigued or when auditory secondary task was introduced.	Brown, I. D. (1982)
	Authors found lower average speed when driving through a 'shopping area' as opposed to a 'residential area', but adding an auditory secondary task did not affect this measure.	Brown, I. D., & Poulton, E. C. (1961)
Speed Standard Deviation	In driving simulator study, author found that speed STD was significantly degraded by adding either an auxiliary task of visually searching a CRT display or a modified Sternberg task. This measure was also sensitive to driving difficulty manipulations (e.g. decreasing curve radii) but this factor interacted with the auxiliary task manipulations.	Noy, Y. I. (Feb. 1990)

	Author found no differences in the distribution of average speeds between automatic and manual transmission cars on the same routes.	Zeier, H. (1979)
Speed Correlation between Vehicles*	Author asked normals, older drivers, and brain-damaged patients to follow a vehicle and match its speed. Three different correlation measures were obtained. Found no difference between normal and older, worse performance for brain-damaged.	Korteling, J. E. (1990)
Speed RMS Deviation	For RMS of velocity, authors used this measure to investigate effects of various vehicle handling characteristics on driving performance during random and step wind gust disturbance tasks. Measure showed some sensitivity to various conditions.	Repa, B. S., & Wierwille, W. W. (1976)
Mean Roadway Glance Duration	Author found that as driving "stress" increases (for increasing flow conditions and for driving between signalized intersections) eye-movement distances became shorter while fixation durations became longer.	Cedar, A. (1977)
Number of Roadway Glances	Authors found more glances to roadway via an occluding visor during higher driving speeds.	Senders, J. W., Kristofferson, A. B., Levison, W. H., Dietrich, C. W., & Ward, J. L. (1967)
	Authors found that the mean number of roadway fixations decreased in a car following situation versus open driving, however, number of fixations pertaining to checking lane position doubled.	Mourant, R. R., & Rockwell, T. H. (1970)
	Authors found that as the macro- and micro-clutter of instrument panels increased (i.e. number of distinct panels and complexity of each panel increased) number of glances toward panel increased and, for micro-clutter, the average length of these glances increased. Longer button labels also increased the number of glances toward the display.	Kurokawa, K., & Wierwille, W. W. (1991)
Total Roadway Glance Time	Authors investigated proportion of total time that driver's eyes were on the road or were on an in-car navigation display system. Conclusions couched in driver adaptability noting that as "attentional demand" went up so did "eyes on road" and concurrently, "eyes on display" went down. "Attentional demand" defined by objective variables (sight distance, curvature, lane restriction, road width) and subjective estimates.	Wierwille, W. W., Hulise, M. C., Fischer, T. J., & Dingus, T. A. (1988)
Position of Roadway Glances*	Authors found that as route familiarity increased visual area sampled became more compact and mean center of glances moved down and slightly to the left.	Mourant, R. R., & Rockwell, T. H. (1970)
Hands-On-Wheel Occurrences		
Mean Hands-On-Wheel Duration		
Total Hands-On-Wheel Time	Mean hands-off-wheel time was found to be lengthened when a simulated instrument panel configuration contained microclutter (i.e. more buttons within the same amount of space.) The length of the display/button label (e.g. abbreviated or not) had no effect of this variable.	Kurokawa, K., & Wierwille, W. W. (1991)

Mean Headway Distance *	Authors discovered no difference in mean headway as a function of experience or probability that lead car would break, however did find that traveling at faster speeds did induce larger mean headway distances.	Colbourn, C. J., Brown, I. D., & Copeman, A. K. (1978)
Following Distance Mean Error	In simulated driving, author found that by adding auxiliary task of visually scanning a CRT, performance on instructed headway distance degraded. It was also sensitive to different levels of auxiliary task difficulty. However, not sensitive to auxiliary visual Sternberg task.	Noy, Y. I. (Feb. 1990)
Following Distance Standard Deviation	Authors found that at faster speeds larger STDs resulted. However no main effects were found for driver experience or probability that lead driver would break (simulating country versus city driving).	Colbourn, C. J., Brown, I. D., & Copeman, A. K. (1978)
Following Distance RMS Error		
Minimum Following Distance (Closest Longitudinal Approach)	Authors calculated the percentage of headway lost when lead car braked suddenly. Found that there were no significant differences on this measure due to the speed of the two cars or the experience of the drivers.	Colbourn, C. J., Brown, I. D., & Copeman, A. K. (1978)
Following Distance Peak Closing Velocity		
Minimum Miss Distance (Closest Approach)	Authors found no effect of nighttime visibility simulated by neutral-density filtered goggles on number of cones knocked down for driving a slalom course at an automatically controlled speed (30 or 45 mph).	Farber, E., & Gallagher, V. (1972)
Peak Closing Velocity		
Response Time (Type I)	Study found that response time to embedded control task (switch off light that lit up in one of four positions) was sensitive to amount of rest and to various secondary tasks in 7-hr simulated driving situation.	Drory, A. (1985)
	Study investigated car's lateral displacement to the opening of a car-door alongside the road. Found response begins approx. 1 - 1.5 sec and maximum displacement to occur 4.0 sec after opening.	Summala, H. (1981)
Response Time (Type II)	Task completion time to press button on a display increased as the number of different displays increased, the complexity of each display increased, and the labels of each display/button increased (e.g. abbreviations versus full spelling).	Kurokawa, K., & Wierwille, W. W. (1991)
Response Distance		
Errors of Omission		
Errors of Commission		

Total Number of Control Movements*	Author found that introducing an auditory secondary task and manipulating factor of fatigue level did not produce differences in total number of control movements (brake, steering, etc..)	Brown, I. D. (1962)
	Authors found number of control movements to increase in a 'shopping area' as opposed to driving through a 'residential area', but did not differ when an auditory secondary task was added.	Brown, I. D., & Poulton, E. C. (1961)

*Starred items were not included in Table 3-1 definitions because the item was redundant with an included item or posed significant problems with data collection, reduction or analysis.

Further, the methodology must also incorporate measures of driver visual behavior, the manual activities involved with driving and the auditory and cognitive loads placed on the driver. It should be noted that although the allocations of manual, cognitive and auditory resources are important considerations, the visual system will likely be the resource of most importance, given the characteristics of the devices anticipated in the immediate future. What is required is the driver's allocation of vision to the right and left side mirrors, to the instrument panel, to any devices overhead in the header area and to the road scene.

The measurement methodology needed to achieve these goals could be an adaptation of the data recording system developed for Task 3 of this project and reported in the Task 3 interim report. This system, using a two-camera video recording system, provided data on the duration and frequency of glances to locations in the road scene, to the mirrors and to locations in the cab. From data reduced from the video the visual allocation strategy used by drivers (as measured by the percent of time spent in looking at specified locations) was identified and compared across selected driving conditions.

The assessment of behavioral changes would entail driver-vehicle performance and behavioral measures taken in four phases:

- Before introduction of the in-cab device;
- During an initial period of use with the device;
- After some period of use with the device; and
- After the device is removed or otherwise made unavailable to the driver.

The first phase establishes a baseline level of driver performance through on-the-road data collection across selected highway, lighting, traffic, and weather conditions in a standard truck not equipped with one of the new in-cab devices. In this data collection effort, subject drivers would be asked to execute a series of selected driving maneuvers (e.g., open road driving, lane changes, close car following, turns at intersections, etc.) under specified conditions and various driver-vehicle measures are taken.

In the second phase, novelty and learning effects would be assessed. The same procedures used in the first phase would be applied here too. Special attention, however, would be directed to assessing the driver's understanding of the device, its use, strengths and weaknesses. Care would be taken to assess novelty effects by measures, say, additional interactions with the device even during periods when such interaction is not called for.

In the third phase, after introduction of and some period of use with a new device in the cab, the data collection procedure would be replicated with subjects performing the same maneuvers given above, under similar conditions. The differences in driver performance before and after the introduction would be identified to assess the effect of the device on the driver. At the same time, subjective measures of relative risk could be made by the drivers to ascertain if their perceptions of risk, given the device, are correlated with any compensatory behavioral strategy (as indicated by the objective measures of resource allocation mentioned above).

In the fourth phase, the device would be removed some time after the driver had grown accustomed to it. This manipulation is intended to mimic conditions where the device is unavailable or unreliable because of road hazards (e.g., precipitation) or because a driver switches to a vehicle which is not

equipped with the device. This manipulation will allow for an assessment of carry-over effects from working with or relying upon the device.

References

- Adams, J.U.G. (1983). Public safety legislation and the risk compensation hypothesis: the example of motorcycle helmet legislation. Environments and Planning C: Government and Policy, 1, 193-203.
- Evans, L. (1984). Human Behavior Feedback and Traffic Safety. (Report No. GMR-4766). Warren, MI: General Motors Research Laboratories.
- Evans, L. (1985). Risk Homeostasis and Traffic Accident Data. (Report No. GMR-4910). Warren, MI: General Motors Research Laboratories.
- Fuller, R. (1984). A conceptualization of driving behavior as threat avoidance. Ergonomics, 27(11), 1139-1155.
- Gibson, J.J & Crooks, L.E. (1938). A theoretical field-analysis of automobile driving. The American Journal of Psychology, 51,453-471.
- Lund, A.K. & O' Neill, B. (1986). Perceived Risks and Driving Behavior. Accident Analysis and Prevention, 18(5), 367-370.
- McKenna, F. (1988). What role should the concept of risk play in theories of accident causation? Ergonomics, 31(4), 469-484.
- Naatanen, R. & Summala, H. (1976). Road User Behavior and Traffic Accidents. Amsterdam: North Holland.
- O'Neill, B., Lund, A.K. & Zador, P. (1985). Mandatory Belt Use and Driver Risk Taking: An Empirical Evaluation of the Risk-Compensation Hypothesis. In L. Evans & R. C. Schwing (Eds.), Human Factors and Traffic Safety (pp. 93-107) New York: Plenum Press.
- Peltzman, S. (1975). The effects of automobile safety regulation. Journal of Political Economy, 83(4), 677-725.
- Rumar, K., Berggmnd, U., Jemberg, P. & Ytterbom, U. (1976). Driver reaction to a technical safety measure - studded tires. Human Factors, 18(5), 443-454.
- Taylor, D.H. (1964). Driver's galvanic skin response and the risk of accident. Ergonomics, 7, 439-451.
- Wilde, J.G.S. (1982). The theory of risk homeostasis: Implications for safety and health, Risk Analysis, 2(4), 249-258.

5.0 DISCUSSION

This report provides a selective review of literature on workload and driver performance to support establishing a link between workload measures which might be included in a heavy vehicle driver workload assessment protocol and highway safety. This link has been discussed by means of referring to two kinds of approaches: actuarial and inferential. The actuarial approach seeks to analyze crash records which involve attention to in-cab tasks, characterize these records according to the putative workload involved with each of several classes of in-cab tasks, then correlate these workload measures to crash incidence. Advantages and disadvantages associated with this approach are discussed.

The inferential approach takes up the remainder of the report. This approach involves the following:

- 1) Measure safety relevant driver-vehicle performance in “baseline” conditions (i.e., without the to-be-evaluated in-cab device). Note that the baseline can also involve use of low-technology analogues to the in-cab devices (paper map use vs. electronic map use).
- 2) Measure safety relevant driver-vehicle performance while the driver is using the in-cab device to be evaluated.
- 3) Measure driver loads imposed by the device itself.
- 4) Relate the results of 2) and 3). A stable relationship which shows driver-vehicle performance degradation during in-cab device use would be used to infer safety impacts.
- 5) Test for significant differences between 2) and 1). This will allow for an assessment, for example, that an electronic map is or is not more workload intensive than working **with** a paper map.

Linking workload to safety is a difficult task for several reasons which have not yet been fully resolved. A number of important issues which bear upon the link between safety and workload measurement which are discussed below.

The Importance of Approximate Science. The theoretical framework which has been presented in this report is a very approximate model: of driver interaction with in-cab devices. Approximate models are common in human factors work. For example, human working memory can be approximately modeled as consisting of seven slots, each capable of holding one item (or chunk), having a decay rate with a half-life of 7 seconds, information being lost due to interference from new information that takes up available slots. Approximate models like these are often clearly wrong in all their details and open to criticism at the theoretical level, yet can be very useful for applied work (Norman, 1988). The models presented in this report need to be verified, to be sure, and changed or enriched with additional theoretical detail as needed. However, it is expected that the framework presented in this report will provide a sound and practical foundation for further work in protocol development.

Necessity for Realistic Tasks vs. "Laboratory" Tasks. At least some of the results reported in the driving literature make use of highly artificial stimulus materials which are intended to maximize the size of an experimental effect. For example, the hands free telephoning research reviewed in Section 2.0 of this report makes use of mental arithmetic and grammatical reasoning problems as test messages. Perhaps not surprisingly, these types of stimuli show an effect of "workload" on driver-vehicle performance. However, it is important to extend these results into the real world with message length, content, pace, and level of driver interaction which more closely resembles that common to relevant in-cab devices. To use the voice communications example again, it might be important to develop an approach where typical messages between dispatchers and drivers are recorded and used to develop the stimulus materials which will be integrated into workload assessment protocol scenarios.

First Order vs. Second Order Issues. Norman (1988) has also noted the importance of prioritizing portions of a system that will be worked on in the face of limited time and budget. In the present case, the most important driver loads are likely to be visual, manual, and cognitive in nature. Other aspects of driver interaction with an in-cab device, such as biomechanical interference (Tijerina, et al., 1991), are assumed to have lesser importance.

The Use of Baseline Measurements with Conventional Tasks. In Task 3 of this project (Kiger, et al., 1992), commanded tasks were used to collect baseline data on the visual loads (glance duration, glance frequency) which are a part of driving in a conventional cab without new high technology in-cab devices. It seems appropriate to consider extending this notion of commanded tasks to include, when feasible, baseline "control" conditions as part of the protocol development effort. This would involve non-automated alternatives to the functions provided by an in-cab device, e.g.:

- use of a paper map instead of an electronic map;
- use a pad and pencil to make notes instead of a video text display;
- use of the conventional radio instead of a special communications system for weather updates.

Collecting data on driver-vehicle performance both with and without high technology in-cab devices might provide useful assessment information, e.g., that the in-cab device takes no longer than a conventional means to accomplish the same function. This type of baseline might complement more general baseline measures on the use of visual or manual resources on a variety of tasks not directly related to in-cab device functionality.

There is a potential problem associated with using conventional in-cab tasks as a standard of acceptable or "safe" device loads. For example, if use of a car radio or cassette player is potentially dangerous, the actuarial analysis, if successful in correlating in-cab loads to crashes, will likely demonstrate this. This actuarial approach might support using conventional in-cab driver loads as benchmarks for safety when evaluating a new in-cab device. Alternatively, the most one might conclude from comparison of new in-cab device use with baseline data as described here is that the in-cab device is or is not imposing any driver demands beyond those which are already present in the fleet.

Criticality of Primary Task Performance Covariation with In-cab Device Use. In this report, emphasis has been placed on the notion that in-cab devices can be judged to negatively impact on safety if primary task performance, i.e., driving, is degraded beyond to-be-determined levels. Primary performance might not be sensitive to changes in workload yet sensitive measures are traditionally sought for workload assessment. In traditional laboratory approaches to workload, sensitivity of measures to workload manipulations led to the development of alternatives to primary task performance (e.g., unembedded secondary tasks). While sensitive measures are highly desirable in a practical HV driver workload assessment protocol, if primary driving performance is unaffected under a reasonably broad range of driving conditions, then it would be very useful to be able to conclude that the device assessed had no deleterious effects on highway safety. Thus, we seek, primarily, safety relevant driver-vehicle performance measures; unlike laboratory studies, sensitivity must be subservient to safety relevance.

Problems with Best-case/Worst-case Analysis It is possible to develop a test scenario which any device will look bad in (the worst case) or any device will look good in (the best case). Between these extremes is a range of driving conditions which will vary by driver, season, route taken, etc. There does not appear to be a clear means to determine what might be a "fair" set of scenarios to test a device in other than representative ones. However, emphasis should be placed on determining under what conditions a system might intrude on the driver and compromise safety. Then the likelihood of these conditions occurring in the real world might be assessed and some estimate made of how critical workload measurements taken under such conditions are. Further, one must also consider actual and perceived demand characteristics of a device. Devices that require immediate attention such as warnings may need to be treated differently in a protocol than devices which allow the driver to decide when it is safe to respond.

The Effects of Individual Difference. It is a truism that workload is an interaction between a task (or set of tasks) and a driver. In interacting with an in-cab device, there will most likely be considerable individual differences among drivers and even within the same driver over a time course of learning, fatigue, etc. In military applications, individual differences in design are often handled by presenting design criteria or evaluation results in terms of percentiles. Thus, a workload assessment might be reported in terms of the 5th percentile or 95th percentile values for workload measures obtained.

Tradeoffs Among Device-induced Workload and More General Safety Benefits, Much of what has been discussed has been in the area of device-specific loads. Yet, it is also relevant to consider the broader picture. For example, an In-Vehicle Safety Advisory and Warning System (IVSAWS) might impose a relatively high driver resource loading (which might compromise highway safety). On the other hand, IVSAWS provides the driver with critical information about hazards along the route (which might enhance highway safety). At this point, some type of cost-benefit tradeoff would be the most likely means to bring these two discrepant findings together and arrive at a consolidated device assessment.

Multiple Transactions and In-cab Device Workload Assessments. Most sophisticated in-cab devices will perform more than one function and so more than one driver transaction is possible. It may well be that, for the same device, one transaction (e.g., menu search) can be accomplished quickly while another transaction (say orienting a moving map display) is cumbersome (but also highly informative once completed). How should these be weighted if one wishes a composite score

for the whole system? At the simplest level, the workload assessment protocol might be used to find and evaluate the simple and more demanding transactions and uncover opportunities for improvement.

Problem of Interactions Between Driving Conditions and In-cab Tasks. It is important to acknowledge that a safety-relevant driver workload measure will be influenced by the degree of demand imposed by driving conditions. For example, while sitting in a parked car, visual glance duration grows as long as the time it takes to complete the task. Glance frequency is likely to reduce to one continuous glance until the task is finished. At the other extreme, moving the eyes from the road scene for even an instant in a race down a winding mountain road would likely be unacceptable. Thus, the sensitivity of device-specific workload measures will depend on driving condition demands as well as on the device itself.

Novelty Effects. One additional dimension of in-cab workload assessment has to do with novelty effects. When the driver is working with the device initially, there may be some period of learning, as indicated previously. Understanding driver-vehicle performance during this initial period of use therefore has safety relevance. More subtle novelty effects might involve driver interaction with a device when none is called for. For example, a device with a visually compelling display might prompt drivers to look at even though it has no part in their goal-oriented behavior (other than perhaps entertainment value). These types of effects also deserve consideration in a comprehensive device assessment.

Refining the Measures of Merit. One problem associated the figure of merit equations presented in Section 2.0 of this report is that there is no criterion per se with which to fit weights for the various driver resources. By changing the weights, alternative designs can be made to look more or less different from each other. Furthermore, the value of including cross product (interaction) terms in the model is unclear at this time. Ultimately, the value of a particular figure of merit is determined by the uses to which that figure may be put. Further research will determine what, if any, improvements are available to the approach presented in this report.

The to-be-developed assessment protocol should provide valuable guidance for in-cab device design by indicating the types of loads which a design imposes on the driver. This, in turn, indicates design attributes to redesign. Eventually, the assessment protocol could provide more diagnostic information about the source of an undesirable driver load. The issues listed above clearly indicate the complexities involved in developing a powerful evaluation tool to aid in device design and selection. These issues cannot be resolved here. The development of the workload assessment protocol will proceed on a better course, however, with the awareness that they exist.

References

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.

Norman, D. A. (1988). The psychology of everyday things, New York: Basic Books.

Tijerina, L., Kiger, S., Wierwille, W., Rockwell, T., Kantowitz, B., Bittner, A., Nygren, T., Myers, L., Tolbert, C., & McCallum, M. (January, 1992). NHTSA heavy vehicle driver workload assessment Task 1 Interim Report: Task analysis data and protocols review (Contract No. DTHN22-91-C-07003). Columbus, OH: Battelle.