Proposed Driver Workload Metrics and Methods Project

Abstract

Ford Motor Company and General Motors Corporation created the Crash Avoidance Metrics Partnership (CAMP) in 1995 to conduct joint pre-competitive projects to accelerate the deployment of future crash avoidance measures. The proposed program utilizes the flexibility of this existing mechanism to bring together Ford Motor Company, General Motors Corporation, Nissan Technical Center North America, Inc. and Toyota Technical Center Inc. USA to propose a Driver Workload Metrics project. This effort will attempt to develop practical, repeatable driver workload metrics and procedures for both visual and cognitive demand that can realistically assess which types of driver interface tasks are appropriate to perform while a vehicle is in motion. In the future, vehicle OEMs will be able to use these workload evaluation procedures to assess what in-vehicle tasks might be accessible to a driver while the vehicle is in motion. The research approach will explore both “ground truth” workload measures taken under test track or on-road driving conditions as well models, simulations or procedures that have been recently developed or proposed. This research will provide a firm foundation for future assessment of feature availability for driver information systems while the vehicle is in motion.

Introduction

The increasing use of in vehicle information systems has made the evaluation of driver workload an area of increasing importance to both government and industry. Attempts to address device evaluation have taken several different directions. Tijerina and his colleagues (Tijerina, Kiger, Rockwell, and Wierwille, 1996) developed a protocol for assessment of heavy truck driver workload. This protocol emphasized empirical data collection using instrumented vehicles, eye glance data analysis, and on-road driving. Such an approach has been criticized as impractical - too expensive, time consuming, and demanding of expertise in study design, data acquisition systems, and data analysis. At the other end of the evaluation spectrum is the SAE Safety and Human Factors Committee's recently drafted SAE J2364, the so-called "15-Second Rule" (Green, 1999a). This rule implies that if a small test participant sample (n =10 or less) can complete a task performed statically (i.e., in a parked vehicle, as a single task rather than done concurrently while driving) within 15 seconds, that task may be accessible while driving. This approach is a bit more taxing than a checklist, but has been hotly debated within the automotive human factors community as perhaps too simplistic and not sufficiently validated. Tijerina, Parmer, and Goodman (in press) conducted a preliminary evaluation using a draft version of the 15-Second Rule and found that it led to a substantial number of mis-classifications of in-vehicle tasks. Other approaches do not involve the collection of any driver performance or behavior data at all. Goals, Operators, Methods, and Selection Rules (GOMS) modeling has been proposed for early-on evaluation of driver interface alternatives (Green, 1999b). This would presumably be of great value for phases in product development where no operational prototype may yet exist. GOMS modeling was originally developed to analytically model single-task, errorless performance using a well-understood task strategy. As such, its relevance to dual-task, error-prone interactions while driving is uncertain. A checklist for the safety assessment of in-vehicle information systems has recently been developed in the U.K. (Stevens, Board, Allen, and Quimby, 1999). The checklist procedure does not require any sophisticated measuring equipment, but it is not clear how well different evaluators will agree among themselves or how well the checklist results will map into actual driver performance and behavior. The Federal Highway Administration (FHWA) has funded Virginia Tech to develop an In-Vehicle Information System DEMAnD (IVIS DEMAnD) model. This computer-based model uses empirical data resident in a database to predict the workload that would be associated with a task completed on a given device. The IVIS DEMAnD model development project was recently completed. This implies that a software package should be available with which to assess IVIS DEMAnD's usability and robustness of prediction. The proposed program offers a means to systematically compare and contrast a wide variety of device evaluation approaches to establish practical guidelines for in-vehicle device accessibility by drivers.

Background

Crash Causation

A crash seldom has a single cause. Therefore, 'Driver Overload,' if it contributes to any given crash, is likely to be **only one of several** contributing factors which co-occurred to cause the crash. Any attempt to predict crash outcomes as a function of one or more measures of driver workload alone will be limited to the extent that it excludes the other contributing factors.

In studying crashes in the past, it has sometimes been common practice to try to relate crashes to driver variables without acknowledging the contributions of other contributing factors in any explicit way. Theories or models which neglect such factors may suffer from what MacGregor & Lewis (1977) termed "gaping middle syndrome." This practice may have the consequence of producing weak theories or models with poor predictive power. Any successful attempt to relate driver performance or workload metrics to crashes should avoid the problem of "gaping middle syndrome."

The gaping middle syndrome represents a substantial hurdle to the development of credible predictive models of crashes. There are many factors that may contribute to a given crash. Some relevant factors may be inadvertently omitted from consideration. Contributing factors may be known but the values of those factors may not be known. Available values for relevant factors may be estimates of uncertain reliability. The interactions among contributing factors may be poorly understood. The probabilities of occurrence and co-occurrence are unknown for many of the factors which contribute to crash causation (e.g., the probability of a co-occurrence of high in-vehicle workload, with high traffic density, bad weather, and sudden braking – or with driver impairment from alcohol or drugs -- or another driver making an unsafe move, or another driver failing to take evasive action – or a failure of a stoplight – etc. --is unknown). It is not possible to account for all relevant variables. For this reason, any attempts at prediction of crash incidence based on driver workload measures will be subject to substantial errors of prediction and will not be undertaken in this project.

Safety Relevance of Driver Performance Measures

The use of an in-vehicle system imposes demands on drivers. If those demands exceed the driver's capacity (e.g., on input modalities, output modalities, or cognitive processing resources), then driver performance on the primary task of driving may be degraded or affected in some way. If the degradation or interference is significant, and if it co-occurs with other contributing factors (such as traffic or unexpected roadway objects), a crash (or near miss) may result. This is the logic which underlies the Driver Workload Metrics and Methods proposal, as shown in Figure 1. The proposed effort focuses on obtaining measures of the demands imposed on drivers by in-vehicle systems - and relating them to measures of driving performance (including any degradation and interference that might be observed). The objective is to develop one or more metrics with which to measure the demands imposed by in-vehicle systems (along with a decision rule for deciding when they may be excessive) – so that new systems which impose excessive demands (that would produce degradation of driving performance) could be re-designed – or locked out from use when the vehicle was in motion. In other words, the second box below is conceptualized as the most opportune place for action that would effectively reduce the incidence of crashes to which driver overload is a contributing factor.

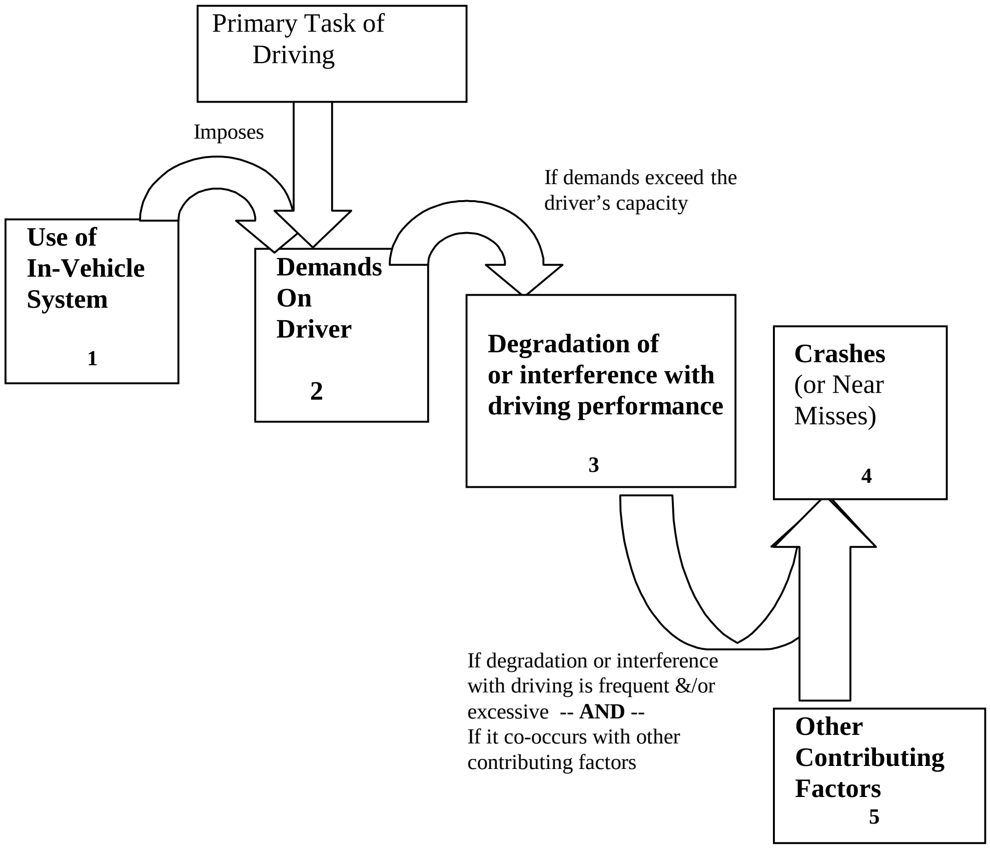

Figure 1. **Logic Underlying the Driver Workload Metrics & Methods Proposal**

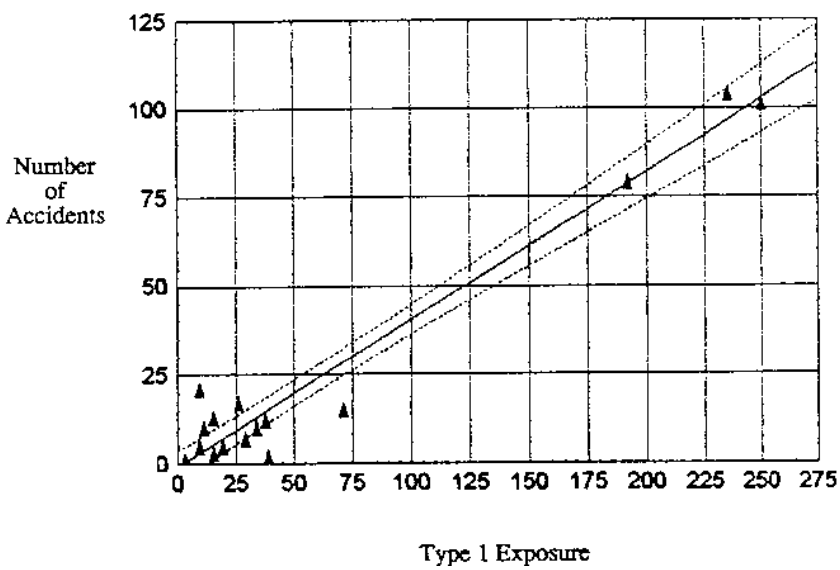

Figure 2. Plot of crash occurrence vs. Type 1 exposure, with regression line and 95% confidence limits shown. (Outlier removed from data set.) From Wierwille & Tijerina (1998).

Although it is not possible to know the general function that relates Box 2 to Box 4 for all in-vehicle systems, what can be done is to look at existing data for already-deployed devices and examine the relationship between Boxes 2 and 4. This is important because it establishes the relevance of the anchor task approach. In the anchor task approach, the relationship of Box 2 to Box 3 for new devices is compared to that for already-deployed devices, such as radios (for which the Box 2 to 3 to 4 link is already approximately known). This kind of evidence for the crash relevance of selected ground-truth workload measures has been described in Wierwille & Tijerina (1998). To illustrate some of their findings, Figure 2 below (from Wierwille & Tijerina, 1998) depicts the relationship between police-reported crashes and eyes-off-road exposure for already-deployed devices. (In this figure, eyes-off-road exposure – or "Type 1 Exposure" – is the product of (mean single-glance time) x (mean number of glances) x (frequency of use)).

Research Approach

Objectives

It is not practical to routinely evaluate "ground truth" workload metrics for all new multimedia functions throughout the product development process. For example, early on, design decisions are being made before any prototype or simulation is available. Later, a simulation may be available for use in a showroom or other 'static' application, yet does not necessarily have the fit and finish suitable to on-road or test track testing. By the time a system is suitable for instrumented vehicle testing, it is important that the probability of any 'showstoppers' has been mitigated by analytical or empirical human factors evaluations carried out at various product development stages along the way.

The proposed research will develop practical, repeatable, and meaningful driver workload metrics and methods. "Practical" means a metric or method is compatible with various phases in the OEM product development process. "Repeatable" implies a metric is consistent in measured results from one test to another. "Meaningful" means a metric is correlated with other safety-relevant, 'ground-truth' measures of driver distraction, such as eyes-off-road time.

Method

In our proposed approach (see Figure 3), a set of in-vehicle tasks that span a wide range of driver demands will be used for evaluation. This will include traditional automotive features (radios, climate controls, mirrors) as baselines or anchors, as well as new multimedia features that have both visual and voice interfaces. For each of these, a set of "ground truth" workload measures such as eyes-off-the-road time, lane exceedances, and driver reaction times to objects and events will be developed. The ground-truth measures of in-vehicle tasks will be obtained while test participants drive on a test track or in traffic.

Next, a set of "surrogate" workload metrics and methods will be developed and evaluated, since it is not practical to routinely evaluate "ground truth" workload metrics for all new multimedia functions throughout the product development process. These "surrogate" workload metrics will be evaluated in terms of their meaningfulness and repeatability when compared with the "ground truth" workload measures, and their practicality. Surrogate metrics and methods will be obtained analytically or by actual testing of the same participants on the same tasks, but under static conditions, i.e., not while driving on a test track or the open road. The meaningfulness of a surrogate will be determined by the degree that it corresponds with one or more ground truth measures. The repeatability of a surrogate measure will be determined by the extent that the results obtained with that measure are consistent from one test to another. That is, repeatability implies that the correlation between a measure from one test and that same measure on a repeat of that test should be relatively high and the distributions should be about the same. A surrogate metric or method will be evaluated as practical to the extent that it meets the practicality criteria developed by the participating OEMs who will use it.

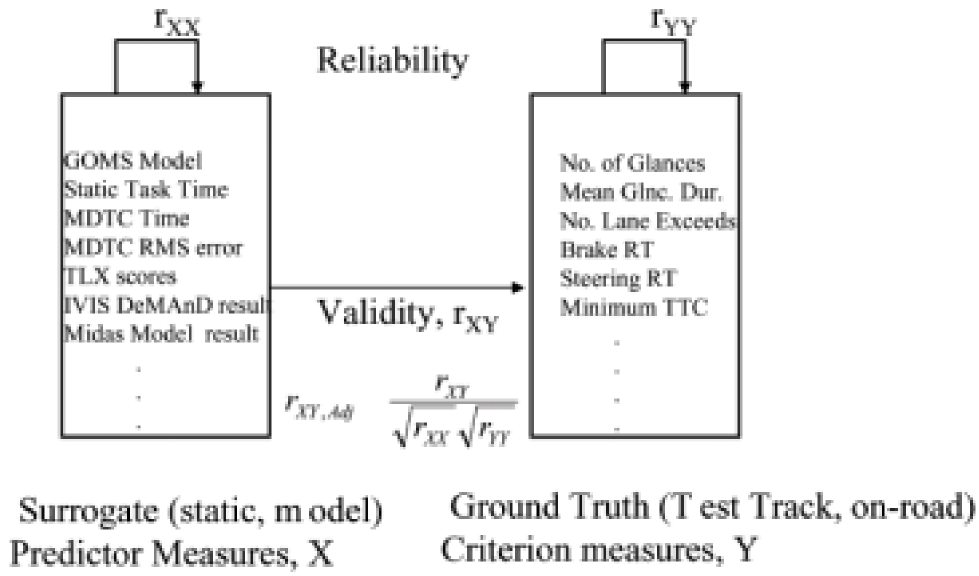

Figure 3. Surrogate workload metrics development approach.

Finally, a "pass/fail" decision threshold will be established for each "surrogate" workload metric. The pass/fail decision threshold will be set by using the results from the baseline tasks. That is, conventional or commonly performed in-vehicle tasks represent the current levels of driver distraction risk encountered by the driving public. A criterion can be identified from the distribution of a given "ground-truth" measure. Setting a threshold at the 85^{th} - 95^{th} percentile values of a distribution is a practice commonly used in traffic engineering. For a "more is worse" ground-truth driver performance measure, e.g., driver reaction time to an object or event, this safety criterion might be obtained by examining the 85th or 95^{th} percentile of the reaction time distribution for conventional tasks (See Figure 4). For a "less is worse" ground-truth measure of driver performance, e.g., eyes-off-road time, this safety criterion might be the, 15^{th} percentile for "less is worse" measures). Actual levels might be determined by consideration of specific common in-vehicle tasks that have been shown in prior research to represent the upper limit of driver distraction (e.g., manually tuning a car radio to a specific frequency). Thus, proposed multimedia functions that exhibit workload metrics that substantially exceed baseline tasks would be considered for redesign or for restricted access while the vehicle was in motion.

A given threshold for a metric can be applied and its classification performance examined. In the example of Figure 5, a True Positive (TP) would legitimately be identified for restricted access or redesign. A False Positive (FP), on the other hand, would be identified for restricted access or redesign with no real justification. A True Negative (TN) would appropriately pass the test, but a False Negative (FN) would pass the test when it should not.

In terms of proposed project tasks, the work is broken down across the following technical tasks:

- Task 1: Select in-vehicle tasks, ground-truth and surrogate measures, and driving scenarios. Select methods to assess how much of the test variability is due to differences between drivers.
- Task 2: Carry out extensive testing as described above to characterize surrogates in terms of repeatability and meaningfulness and to characterize the repeatability of the ground truth measures.
- Task 3: Confirm the repeatability, meaningfulness, and practicality of selected metrics and methods by repeating the evaluation process of Task 2 on a new set of in-vehicle tasks, a new set of test participants, and with a new set of evaluators.
- Task 4: Document the selected surrogate metrics and methods, including evaluation criteria and general project findings.

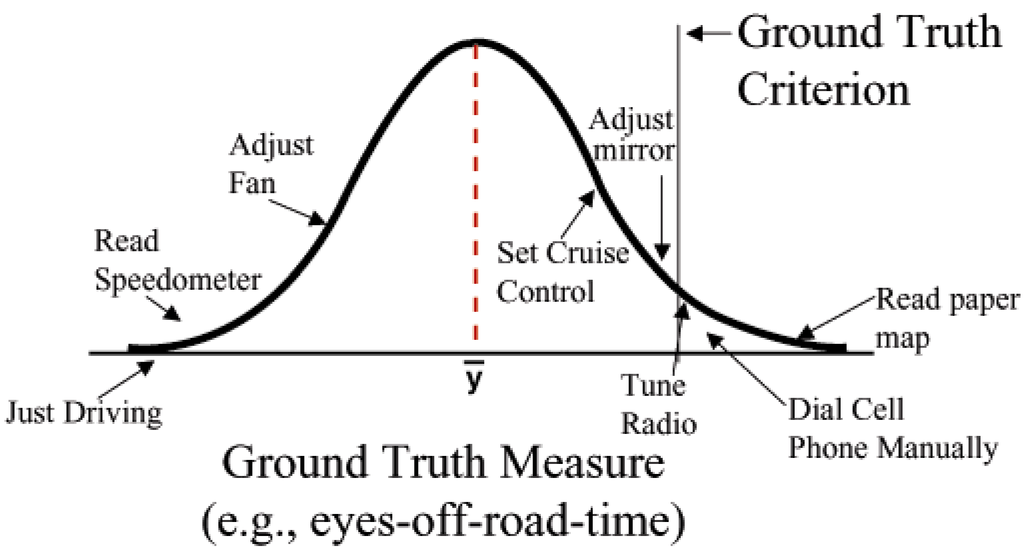

Figure 4. Hypothetical distribution of a ground truth driver performance measure for conventional in-vehicle tasks.

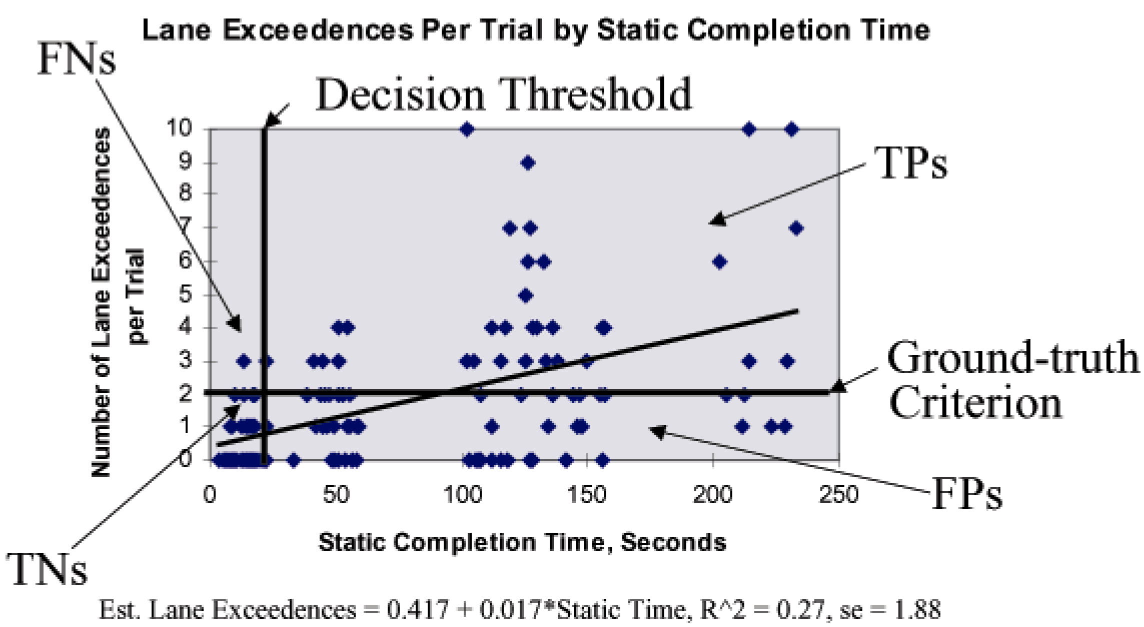

Figure 5. Classification Performance of a surrogate measure of Driver workload.

Another issue relates to how this work will proceed when a surrogate may correlate well with one but not another ground truth measure. In the context of crash scenarios, it is important to note that performance measures are likely to differ in their importance from scenario to scenario. That is, in aggressive conflict driving scenarios (e.g., small headway to lead vehicle with sudden high "G" deceleration of the lead vehicle), driving performance measures of "headway keeping" and "brake reaction time" would likely be the most significant "ground truth" measures. "Eyes-off-road time" may be disproportionately underweighted in such a scenario because it would be difficult to get test participants to take their eyes off the road (glance frequency and duration may be abbreviated in an aggressive conflict driving scenario). Similarly, "lane keeping" measures would also be underweighted, since lanekeeping has little contribution to collisions with the lead vehicle. However, in other driving scenarios (e.g., involving the unexpected appearance of an object on the roadway – or involving narrow lanes and oncoming traffic), both "eyes-off-road time" and "lane-keeping" would emerge as very important. Indeed, it may be the case that composites are not appropriate because weighting implies that various driver workload measures can be traded off against one another. This complex issue will have to be addressed in this project through an assessment of empirical data and the structure within that data.

Deliverables

The proposed project will produce a documented set of surrogate metrics and methods that have been thoroughly assessed in terms of meaningfulness, repeatability, and practicality. The resulting driver workload metrics and criteria will be useful in assessing which system functions or controls may be accessible while driving.

Summary

The goal is to develop practical, reliable, and meaningful (i.e., valid) surrogate metrics with which to estimate or measure the distraction potential associated with a given in-vehicle device or device function.

The proposed approach of using conventional in-vehicle tasks to develop baseline distributions from which a safety criterion will be drawn avoids many problems. First, there is no need to estimate or extrapolate crash probability, crash incidence, or crash severity as a function of one or more driver workload measures. Instead, the proposed approach takes an anchor point for a driver workload measure from the distribution of conventional in-vehicle tasks and operationally defines a safety criterion with respect to this distribution. Second, it acknowledges the fact that common in-vehicle tasks represent a range of socially acceptable driver distraction risks beyond which new systems should not go. It thus avoids the many problems of attempting to make quantitative predictions of crash probability, crash incidence, or crash severity for which many assumptions will be required, few data will exist on contributing factors, and for which the error boundaries are therefore likely to be huge.

The proposed research program is a conservative approach that has great potential to further the development of driver information systems that have been designed to keep distraction within levels associated with current conventional in-vehicle tasks

References

Green, P. A. (1999a). The 15-Second Rule for driver information systems. *Proceedings of the ITS America Ninth Annual Meeting* (CD-ROM).

Green, P. A. (1999b). Estimating compliance with the 15-Second Rule for driver-interface usability and safety. *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting (CD-ROM).*

MacGregor, R. J., & Lewis, F. R., *Neural modeling*. New York: Plenum Press, 1977.

Stevens, A., Board, A., Allen, P., and Quimby, A. (1999, December). *A safety checklist for the assessment of in-vehicle information systems: A user's manual* (Report No. PA 3536/99). Crownthorne, Berkshire, UK: Transport Research Laboratory.

Tijerina, L. (1995). *Executive summary: Heavy vehicle driver workload assessment* (Report No. DOT HS 808 466). Washington, DC: National Highway Traffic Safety Administration.

Tijerina, L., Kiger, S., Rockwell, T. H., & Wierwille, W. W. (October, 1996). *NHTSA heavy vehicle driver workload assessment final report supplement – Task 5: Heavy vehicle driver workload assessment protocol* (DOT HS 808 467). Washington, DC: National Highway Traffic Safety Administration.

Tijerina, L., Parmer, E., and Goodman, M. J. (in press). Preliminary Evaluation of the Proposed SAE J2364 15-Second rule for Accessbility of Route Navigation system Functions while Driving. To be published in *Proceedings of the Human Factors and Ergonomics Society 44th Annual Meeting.*

Wierwille, W. W., and Tijerina, L. (1998). Modeling the relationship between driver in-vehicle demands and accident occurrence. In A. Gale, et al. (Eds.), *Vision in Vehicles VI* (p. 233-243). Amsterdam: Elsevier.