ABSTRACT

This paper provides a project overview and summarizes the general lessons learned from a major research project on driver information systems. Emphasized are lessons that transcend experiments and have programmatic impact for project managers and sponsors.

The goals of the project were to develop (1) human factors guidelines, (2) methods for examining the safety and ease of use of driver interfaces, and (3) a driver performance model. Five systems (navigation, traffic information, road hazard warning, vehicle monitoring, and car phones) were examined in 20 experiments. Experiments included surveys at driver licensing offices, response time tasks, driving simulator studies, part-task simulations, and on-the-road evaluations.

A major group of lessons concerned how realistic, inexpensive, and rapidly-produced interface prototypes can be, and how to achieve a high level of fidelity. (Use SuperCard and HyperCard to develop them. Have prototypes operated surreptitiously by experimenters in response to real world events (the Wizard of Oz method).)

Of the methods explored, there were lessons concerning focus groups (ineffective for products beyond participant experience), response time tasks (use them to evaluate display readability) and usefulness of the subjects-in-tandem method (in which pairs of subject collaborate in using a product).

The research provided several lessons concerning the inadequacies of low-fidelity driving simulators (unsatisfactory estimates of driving performance, sign legibility inadequate for route guidance assessments, numerous problems with videotape-based simulations). Lessons from the on-the-road evaluations related to test vehicle shake down (extensive time is needed), and determining workload (for comparable test conditions) and data reduction. (New methods are needed in both cases.)

During these evaluations, design guidelines emerged from interface design decisions (the preferred approach), not from a summary of the literature. General guidelines and principles (especially consistency), proved very useful in design.

PROJECT OVERVIEW

This paper summarizes a four-year project on driver interfaces for motor vehicles of the future sponsored by the U.S. Department of Transportation (DOT). This topic includes controls and displays with which the driver interacts, as well as the information presentation logic and sequencing.

The project had three objectives:
• Provide human factors guidelines for the design of in-vehicle information systems.
• Provide methods for testing the safety and ease of use of those systems.
• Develop a model that predicts driver performance while using those systems.

Each of the three project objectives pushed the research focus in a different direction. To support guideline development, research was needed on specific (and well designed) interfaces for many systems, so detailed recommendations could be developed. To support recommendations for methods, varied methods need to be explored; so often there was a concentration on a few systems, an emphasis different from that for guideline development. For modeling work, the need was to explore a wide variety of systems. To make the model robust, both good and bad designs were considered. Thus, the research for this project often involved significant compromises, with each experiment addressing only some of the project objectives.

Although only passenger cars were considered in the project, the results apply to light trucks, minivans, and vans, because the driver population and likely use of these vehicles is similar to cars. It should also be noted that only able-bodied drivers were considered. Disabled and impaired drivers are likely to be the focus of future DOT research. Further, this project did not consider driver performance monitoring or collision avoidance, topics covered by other DOT contracts.

An overview of the initial project plans appears in Green, Williams, Serafin, and Paelke (1991). The research program conducted is summarized in figure 1, table 1, and Green (1993). The project began with a literature review (Green, 1992), and with several focus groups (Brand, 1990; Green and Brand, 1992) to understand the domain of the problem. Subsequently, the extent to which various driver information systems might reduce accidents, improve traffic operations, and satisfy driver needs and wants was analyzed (Green, Serafin, Williams, and Paelke, 1991; Serafin, Williams, Paelke, and Green, 1991). That analysis led to the selection of two systems for detailed examination (traffic information and car phones) while contractual requirements stipulated three others (navigation, road hazard warning, and vehicle monitoring). Each of the five systems selected was examined separately in a sequence of experiments. Typically, patrons at a local driver licensing office were shown mockups of interfaces, and driver understanding of the interfaces and preferences were investigated. Interface alternatives were then compared in laboratory experiments involving response time, performance on driving simulators, and part-task simulations. The results for each system are described in separate reports (Green, Hoekstra, Williams, Wen, and George, 1993; Green and Williams, 1992; Hoekstra, Williams, and Green, 1993; Paelke, 1992; Paelke and Green, 1993; Serafin, Wen, Paelke, and Green, 1993; Williams and Green, 1992; Williams, Hoekstra, and Green, 1993). Development of precompetitive interfaces as part of this project identified issues the design guidelines should address. This approach also provided a basis for further experimental work. For example, to compare visual versus auditory route guidance interfaces, reasonable implementations were needed. Otherwise, the modality question is confounded with the quality of the implementation.

Following this were three on-the-road experiments, the first two of which considered navigation, traffic information, hazard warning, and vehicle monitoring systems (Green, Williams, Hoekstra, George and Wen, 1993). The final experiment considered the navigation system and car phones (Green, Hoekstra, and Williams, 1993).

Development of the design guidelines (Green, Levison, Paelke, and Serafin, 1993), a human performance model (Levison, 1993; Levison and Cramer, 1993), and assessment procedures (Green, 1993a; Green, 1993b) occurred in conjunction with the experimental work.
This paper describes what was learned regarding how to design and evaluate driver interfaces. Lessons pertaining to the model of driving, interface design alternatives, and results from specific experiments will appear elsewhere, along with the lessons on presenting results.

![Project overview diagram](image-url)

Figure 1. Project overview.
Table 1. Experiments conducted.

<table>
<thead>
<tr>
<th>#</th>
<th>UMTRI report</th>
<th>System</th>
<th># Subjects</th>
<th>Method</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>90-22</td>
<td>All systems</td>
<td>46</td>
<td>Focus group</td>
<td>Public desire for new information systems and specific features</td>
</tr>
<tr>
<td>2</td>
<td>93-16</td>
<td>IVSAWS</td>
<td>10</td>
<td>Survey</td>
<td>Warnings for hazards</td>
</tr>
<tr>
<td>3</td>
<td>93-16</td>
<td>IVSAWS</td>
<td>75</td>
<td>Survey</td>
<td>Best warning text and graphics</td>
</tr>
<tr>
<td>4</td>
<td>93-16</td>
<td>IVSAWS</td>
<td>20</td>
<td>In-car interview</td>
<td>Best location cues for warnings</td>
</tr>
<tr>
<td>5</td>
<td>93-17</td>
<td>Car phone</td>
<td>19</td>
<td>Interview</td>
<td>Suggestions for labels for simulated telephone</td>
</tr>
<tr>
<td>6</td>
<td>93-17</td>
<td>Car phone</td>
<td>12</td>
<td>Interview</td>
<td>Preferred abbreviation method</td>
</tr>
<tr>
<td>7</td>
<td>93-17</td>
<td>Car phone</td>
<td>12</td>
<td>Driving simulator</td>
<td>Dialing times and driving performance for HUD and IP location, voice and manual dialing</td>
</tr>
<tr>
<td>8</td>
<td>93-22</td>
<td>Vehicle monitor</td>
<td>27</td>
<td>Interview</td>
<td>General knowledge of vehicles and malfunctions</td>
</tr>
<tr>
<td>9</td>
<td>93-22</td>
<td>Vehicle monitor</td>
<td>60</td>
<td>Survey</td>
<td>Wording of vehicle monitor warnings</td>
</tr>
<tr>
<td>10</td>
<td>93-22</td>
<td>Vehicle monitor</td>
<td>20</td>
<td>Survey</td>
<td>Understanding of various warnings</td>
</tr>
<tr>
<td>11</td>
<td>93-20</td>
<td>Traffic information</td>
<td>&lt;20</td>
<td>Interview</td>
<td>Understanding of graphics, gesture stereotypes</td>
</tr>
<tr>
<td>12</td>
<td>93-20</td>
<td>Traffic information</td>
<td>20</td>
<td>Survey</td>
<td>Color coding, retrieval strategies</td>
</tr>
<tr>
<td>13</td>
<td>93-20</td>
<td>Traffic information</td>
<td>16</td>
<td>Driving simulator</td>
<td>Retrieval times, eye fixations for various screen designs, general understandability</td>
</tr>
<tr>
<td>14</td>
<td>92-21</td>
<td>Route guidance</td>
<td>&lt;20</td>
<td>Interview</td>
<td>Display format</td>
</tr>
<tr>
<td>15</td>
<td>92-21</td>
<td>Route guidance</td>
<td>60</td>
<td>Survey</td>
<td>Plan versus aerial versus perspective formats</td>
</tr>
<tr>
<td>16</td>
<td>92-21</td>
<td>Route guidance</td>
<td>12</td>
<td>Response time</td>
<td>Display format, location, graphics for road</td>
</tr>
<tr>
<td>17</td>
<td>93-31</td>
<td>Route guidance, traffic info.</td>
<td>48</td>
<td>Respond to videotape</td>
<td>Role of landmarks and presentation modality (auditory, visual) on interface usability</td>
</tr>
<tr>
<td>18</td>
<td>93-32</td>
<td>All systems except phone</td>
<td>12</td>
<td>On road, subjects-in-tandem</td>
<td>Looked for severe problems with interface design</td>
</tr>
<tr>
<td>19</td>
<td>93-32</td>
<td>All systems except phone</td>
<td>43</td>
<td>On road (singly)</td>
<td>Lane, speed, throttle, steering wheel variance, glance frequency, safety &amp; usability ratings</td>
</tr>
<tr>
<td>20</td>
<td>93-35</td>
<td>Route guidance and phone</td>
<td>8</td>
<td>On road (singly)</td>
<td>Lane, speed, throttle, steering wheel variance, glance frequency, safety &amp; usability ratings</td>
</tr>
</tbody>
</table>
LESSONS LEARNED ABOUT SIMULATING IN-VEHICLE INFORMATION SYSTEMS

Lesson 1: SuperCard and HyperCard Are Preferred Applications for Simulating In-Vehicle User Interfaces.

During the early phases of the project, several platforms and applications were considered for simulating user interfaces. Requirements included low cost, support of standard television monitors and various input devices, power and size compatibility with in-car use, the ability to easily present colors, and the ability to accurately time user inputs and obtain rapid system responses to them (Green, Boreczky, and Kim, 1990).

Operating systems considered included UNIX and Silicon Graphics, DOS and Windows, and Macintosh. The Silicon Graphics application of interest was VAPS, quite popular within the aviation industry. However, hardware, software, and support costs, as well as concerns for in-car use, led to eliminating Silicon Graphics and UNIX platforms from further consideration.

Toolbook, a Windows application, was also considered. However, it did not provide the desired level of support for color usage or straightforward support for a variety of input and output devices in the same manner as Macintosh applications.

At the University of Michigan, because of ease of use, support of graphics, and widespread availability, Macintosh computers are the preferred platform. Of the Macintosh applications considered, SuperCard was preferred because of its support for color and graphic objects. SuperCard has been used by GM for developing the TravTek interface, by Motorola for the ADVANCE project, and by NASA for the space station. It is widely accepted by the human factors community for interface prototyping, and recommended for developing future automotive projects. Other authoring software considered included HyperCard (used for simulating phones in this project) and MacroMind Director, a moderately expensive application.

Lesson 2: Rapid Prototyping, While Fast, Is Not Instantaneous for Real Interfaces.

Scaling from experience prior to this project, the impression was that real interfaces could be put together in a few weeks. The navigation and other interfaces actually required two months or more, mostly for debugging (and in the case of the navigation system, comparison with real roads). This time is still considerably less than the time to develop a real system, or modify a real system for experimental work. Modifications of working interfaces were quite rapid.

Lesson 3: HUD's Can Be Simulated Using Mirrors.

One of the navigation presentation schemes examined in this project used a HUD (Head-Up Display). The color graphic display device had to be readily interfaced to a Macintosh. Developing a real HUD system for this purpose or customizing an off-the-shelf HUD (most likely from an aircraft), would have cost at least $100,000 and taken a year to complete. In this project, a 7-inch color video monitor was mounted facing forward at just below shoulder height between the driver and passenger. A small mirror mounted on a suction cup was attached to the windshield, and adjusted by each driver so they could see the mirror image of the monitor's reverse scanned image. To make the HUD partially transparent, the width of the mirror was less than the driver's interocular spacing, enabling drivers to "look around it." The image was quite bright and readily seen in sunlight. While this approach may not be acceptable to customers for products, it worked very well for exploring HUD concepts. Given the interest in utilizing HUds for a variety of systems (e.g., collision avoidance, vision enhancement, etc.), use of this technique is encouraged.
Lesson 4: For Systems That Require Considerable In-Vehicle Hardware for Sensing or Processing, or Outside-the-Vehicle Infrastructure, Use the Wizard of Oz Method to Simulate the System for Interface Evaluations.

In several experiments, driver interfaces in the test vehicle were surreptitiously controlled by a back seat experimenter using a numeric keypad. The experimenter behaved much like the wizard in the Wizard of Oz story.

In on-the-road tests (Green, Hoekstra, and Williams, 1993; Green, Williams, Hoekstra, George and Wen, 1993), navigation information was presented by either a small LCD display mounted on top of the instrumented panel, a HUD, or by computer-controlled voice messages. A Macintosh computer running SuperCard controlled the navigation system, presenting a series of screens and voice messages. When subjects completed a turn, the experimenter in the back seat pressed a button on a keypad to advance to the next screen in the sequence. Screens showing the current location were updated utilizing the vehicle speed provided by a 486 computer (in conjunction with route data on the Macintosh). When a navigation error occurred, the experimenter pressed a button for the "off route" screen, and verbally guided the driver back to the test route.

Another example application of the Wizard of Oz method was a simulation of the In-Vehicle Safety Advisory Warning System (IVSAWS) (Hoekstra, Williams, and Green, 1993). In a real system, radio transmitters on objects (vehicles, bridges, trains, etc.) would send warnings to drivers during hazardous situations (police cars while on high speed runs, school buses while discharging children, etc.). In on-the-road evaluations of IVSAWS (Green, Williams, Hoekstra, George, and Wen, 1993), the experimenter pressed a key when a hazard was observed, causing the appropriate warning to appear on an LCD in front of the driver.

In both cases, drivers did not realize the experimenter was controlling the display, and behaved as if the system was real. Implementing and operating a real system would have added years to the project schedule and at least hundreds of thousands of dollars to the project cost. Use of the Wizard of Oz method is encouraged for relatively low cost exploration of the usability and usefulness of new technologies prior to costly engineering development.

LESSONS LEARNED ABOUT DRIVING SIMULATORS


The driving simulator used during portions of this research was very basic, showing a series of road edge markers (figure 2). In the cellular phone (Serafin, Wen, Paelke, and Green, 1993a,b) and traffic information experiments (Paelke, 1992; Paelke and Green, 1993), when participants were distracted by in-vehicle tasks, they seemed to make larger lane departures and depart more often then they might on the road. Consequently, summary measures of performance (lateral variance, speed variance) may be inaccurate estimates of absolute performance on the road. On-the-road confirmations have not been conducted, though from other experiments in progress, higher fidelity simulation (showing road edges as solid lines, a detailed background, some roadside texture, etc.) has elevated the priority of the steering task, and is suggested for future evaluations. Interestingly, the completion times for the in-vehicle tasks in the low-technology simulator were close to those times measured on the road.

As an aside, there also have been difficulties in getting engineering managers to accept low fidelity simulations as providing a "valid" context. Acceptance has not been a problem with human factors colleagues.
Lesson 6: Most Driving Simulators Are of Limited Utility for Examining Driver Performance with Route Guidance Interfaces.

Most driving simulators do not have the resolution to present signs that are legible at the same visual angle they are legible on the road. (The simulator at FHWA circumvents the limitations of computer graphics by presenting signs on slides.) As a consequence, signs need to be larger than normal so they can be read in driving simulators. This creates significant problems for studies of navigation, where signs are used in conjunction with the navigation interface to make decisions. Therefore, until significant technological advances occur, on-the-road tests should be used for driver performance evaluations of route following.

Lesson 7: Videotape-Based Simulations Are of Limited Use for Examining Driver Performance with In-Vehicle Information Systems.

The lack of low-cost, moderate-fidelity driving simulators has been a major roadblock to conducting research on ITS driver interfaces. (See Sweet and Green, 1993 for a recent example.) As an alternative, a videotape-based simulation tool was developed in this project to assess navigation interfaces. Participants were shown a videotape of a view of a road scene, and coordinated with it, the appropriate information from a navigation system (Green, Williams, Hoekstra, George and Wen, 1993). The road scene was recorded by a camera near the driver's eye point. This method had the potential advantage of displaying road signs and landmarks that were moderately legible.

During the pilot phases of the experiment, the videotape was projected onto a reflective wall in front of the participant. In a few cases, motion discomfort was both immediate and significant after the first turn. The level of discomfort was accentuated by the recording method (a fixed camera on the vehicle), the magnitude of turns (90 degrees), the enclosed vehicle cockpit, the overall lighting level (dark, to facilitate viewing the projected image), and the wide field of view. Displaying the image on
a video monitor on the hood of the mockup, and increasing the ambient illumination reduced the level of motion discomfort. Doing such, however, compromised realism.

In the videotape-based method, the driver's task was to press a button when they could see the turn point. The farther from a turn point that a decision could be planned, the safer the maneuver should be. To maintain attention to the road scene, drivers stepped on the brake whenever a lead vehicle braked. For some subjects, this did not occur, as they nodded off while watching the videotape. The presentation of navigation displays and the clocking of driver responses was coordinated with the manual start of the videotape. However, problems of tape stretch and variations in playback speed made it difficult to determine exactly when navigation events occurred relative to events on the videotape.

An improvement would be to use Society of Motion Picture and Television Engineers (SMPTE) time code to trigger the navigation system. Computer playback of QuickTime movies would also provide more exact timing, though current resolution limitations and significant storage requirements could present difficulties. Playback of Motion Picture Experts Group (MPEG) compressed files, laser-disks, or CD-ROM recordings may be options. All of these options are costly and would require significant development time.

Thus, use of the videotape for driver performance experiments is not recommended. However, where providing participants with a product experience is important (such as for focus groups), this method may be an option.

LESSONS LEARNED ABOUT ON-THE-ROAD EVALUATIONS

Lesson 8: Comparisons of ITS Interfaces Require the Use of Similar Roads for the Various Conditions. Finding Comparable Road Sections for Tests Can Be a Problem.

Comparisons of interfaces and tasks, conducted either on the road or in a driving simulator should be made within subjects and use similar (if not identical) roads, because differences due to drivers and roads can be large. Baseline data are required (performance without ITS devices). For interface studies, especially with the test route, familiarity is a critical factor. Furthermore, because of day-to-day variations in weather and traffic, it is highly desirable that interface differences be examined at almost the same time.

In this project, road sections of baseline driving and driving use of a navigation system (Green, Hoekstra, and Williams, 1993) were alternated. Project evaluations focused on straight sections for a few roads, using the posted speed to categorize segments.

The effects of driving workload on use of such systems have not been quantified. For that matter, how such factors as speed, the radius of curvature of a road, lane width, sight distance, traffic volume, number of lanes, etc. affect driving workload is not well understood. The lack of an expression for quantifying workload, allowing the comparison of simulations and on-the-road evaluations, is a major impediment to the advancement of ITS. Since the effects of road geometry and traffic on workload have not been quantified, the interim test route for the standardized interface assessment protocol requires use of the test route in Michigan from this project (Green, 1993). This is the only route for which the desired baseline data exists, a significant constraint.

Lesson 9: At a Minimum, Several Months Should Be Set Aside to Shake Down a Test Vehicle Before It Is Used for On-the-Road Evaluations.

Three major on-the-road experiments were conducted as part of this project (Green, Hoekstra, and Williams, 1993; Green, Williams, Hoekstra, George, and Wen, 1993). While the test vehicle had been used for previous research (Hoekstra, Williams, Green, and Paelke, 1992, Sweet and Green,
1993), use of the vehicle for this project involved new hardware (lane tracker) and other enhancements for which minimal shake down time was scheduled. There were problems with dropouts of the speed signal (suggesting the impossible, a 30 mph decrease in a 33 ms interval) and failure of the lane tracker to lock on to the road markings. The time allocated for shake down should be equal to the time allocated for testing drivers. Instrumented cars are not turnkey systems, even in the best situations.

**Lesson 10: Duplicate (or Triplicate) Hardware Should Be Provided for Development and Testing.**

There were several times during this project in which development was attempted in the office while testing occurred in the instrumented car or in the simulator. Project funding constraints did not allow for purchase of duplicate computers or displays so work could proceed in parallel (or all three at the same time). In addition, when there was a test hardware failure, or suspected failure, duplicate hardware could not be swapped to minimize delays. Duplicate hardware is also necessary for accurately representing interfaces during in-office demonstrations.

In this project, the main display used for navigation failed in the middle of subject testing. That display was not sold in the U.S. and shipment from Japan would have been time consuming. By coincidence, the SAE Annual Congress was being held at about that time nearby in Detroit and a swap of a demonstration model was arranged.

**Lesson 11: Reduction of Vehicle Performance Data Is a Significant Problem in On-the-Road Experiments, and Can Be a Problem in Driving Simulator Research.**

The data collected in on-the-road experiments (and sometimes in driving simulation experiments) can be voluminous. As an example, in Green, Hoekstra, and Williams, 1993, steering wheel angle, throttle position, speed, lateral position, and eye fixation location, were all sampled at 30 Hz, along with other measures. As a consequence, every 30 minutes 4 megabytes of data were collected, only some of which was of interest. However, prior to analysis, the data needed to be filtered. Matters needing attention included occasional signal losses of the speed pulse, loss of lock of the lane tracking signal, etc. While some of the filtering could be automated, the data of each subject had to be inspected by an experimenter to identify anomalies. This time can exceed a day for each subject, a considerable time in view of the duration of the test segment of interest.

One implication of this lesson concerns when driver performance data should be collected. It may be that during the initial phases of on-the-road testing a reduced level of data analysis may be desired, with the focus being on navigation errors and subject comments. During product development, rapid turn around of results is required so design can be iterative. That is not possible if months are required to analyze results. Reliance upon such limited analysis have proven effective in design of human-computer interfaces by many companies (Wiklund, 1994).

Particularly useful in this regard are tools that automate the data reduction process. This includes software to identify and replace outliers, and to generate summary statistics for trip segments. Those carrying out in-vehicle studies are encouraged to develop software for this purpose prior to collecting data.

**Lesson 12: Affordable Eye Fixation Recording Systems Do Not Provide the Desired Characteristics. Even for the Best Systems, Expect to Perform Some Manual Analysis of the Data.**

Eye fixation recording systems are uncomfortable to wear, do not provide both day and night capability, often have problems with glasses and contacts, and limit the field of view. In this project, a NAC model V eye fixation recording system was used to identify the driver's direction of gaze.
To provide a greater field of view, the left pupil sensor was removed. For the model V, the wear limit was 30 minutes.

Reduction of data required considerable effort. While features providing for digitized output were helpful, manual reduction of data (in conjunction with a videotape of the forward scene) was often required because of a lack of software to identify fixated objects. This problem was exaggerated by slippage of the headpiece during testing.

Development of a low-cost, off-head recording system for in-vehicle use should be a high development priority, especially for use with mature drivers. As an alternative to specialized eye fixation recording products, in many situations two cameras are installed in the test vehicle, one aimed at the driver and a second camera aimed at the forward scene. Data is reduced manually by reviewing a split screen image of the road scene and driver's face. Data reduction time can be significantly abbreviated if the number of fixations to various locations (but not their durations) are used, with further reductions occurring when software to support counting is provided (Green, Williams, Hoekstra, George and Wen, 1993). If durations are required, then determining them for a few cases should be considered. For product evaluations, use of this more rapid method is recommended.

Lesson 13: Specifying Absolute Levels of Safe and Unsafe Driving Performance is Generally Not Possible at This Time.

The project contract called for specifying safe and unsafe levels of driving performance (e.g., lane variance, speed variance), an impossible task. How safe something is falls along a continuum, not in two discrete levels. Further, uniformly minimal performance on all measures is unacceptable. The solution was therefore to select three levels (best expected, desired/planned, worst case) and to further require that if only worst case performance was achieved for one variable (e.g., lateral variance), then best expected performance should be achieved for another (e.g., speed variance).

In addition to theoretical concerns, there is an absence of data on which to base a judgment. To specify what is unsafe (or even abnormal), data are required on what constitutes normal driving. Needed are typical values for lane variance, eyes-off-the-road time, etc. and how those measures vary with driver age, time of day, road type, traffic, and other important factors. Those data do not exist and they are urgently needed.

LESSONS LEARNED ABOUT OTHER EVALUATION METHODS

Lesson 14: Focus Group Participants Cannot Provide Useful Input Concerning Driver Interfaces That Are Beyond Their Experience.

During the early phases of this project, as required by DOT, focus groups were conducted in Los Angeles and New York. The purpose was to obtain a sense of driver reactions to available advanced driver information systems, and suggest what drivers might want in future products (Brand, 1990; Green and Brand, 1992). While the results provided some useful insights, there was some disappointment with them. At that time, few drivers knew anything about electronic navigation systems and other advanced products, and none had any experience using them. Consequently, drivers had considerable difficulty in making assessments about the ease of use, usefulness, their willingness to pay, and other marketing related qualities of navigation systems, or for that matter, any products that were beyond their experience. Ideally, manufacturers would like to have an impression of a product market before engaging in research or development. It is uncertain if the necessary customer experience can be provided using static physical mockups, a videotaped demonstration, part-task simulations, driving simulator experience, a test vehicle drive, or long term exposure in a borrowed vehicle. What is necessary may vary with the class of system and with focus group moderator creativity. Identifying experience necessary to make good new product
marketing decisions (and provide useful background data for engineering enhancements) is a topic worthy of further study.

**Lesson 15: Response Time Experiments Are a Cost-Effective Means for Collecting Data on the Readability of In-Vehicle Displays.**

Significant work was completed using this method to compare the readability of several types of visual displays (Green and Williams, 1992; Williams and Green, 1992). In one experiment, drivers were seated in a vehicle mockup. Simultaneously they were shown a slide of an intersection (projected on a wall), and a navigation display (inside the mockup). Their task was to determine if the same or different intersections were being presented. Using this method, differences due to display location (HUD versus instrument panel mounting) and graphics implementation (6 designs) were readily examined. The experimental protocol was sensitive to 40 ms differences in how roads were presented (where the mean times were approximately 1575 ms), and the data for response times, errors, and preference ratings were correlated with each other. A task involving driving simulation or on-the-road evaluation would have been several times more costly and taken significantly longer to complete. This method is recommended for readability evaluations of visual displays.

**Lesson 16: Classification of Keying Errors Is Not Straightforward or Easily Automated.**

In response time experiments, the responses are generally categorized as either correct or errors. Other than classifying errors into general time categories (fast guesses, responses over some time limit), little is done to examine the different types of errors. For more complex interfaces, such as phones, destination entry systems, and hierarchical menu systems, the pattern of errors needs attention. For example, shown below are some of the types of errors than might occur for dialing a phone number.

<table>
<thead>
<tr>
<th>Type of Error</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correctly dialed</td>
<td>764 6589&lt;send&gt;</td>
</tr>
<tr>
<td>Selection error</td>
<td>764 6588&lt;send&gt;</td>
</tr>
<tr>
<td>Transposition</td>
<td>674 6589&lt;send&gt;</td>
</tr>
<tr>
<td>Missing digit</td>
<td>74 6589&lt;send&gt;</td>
</tr>
<tr>
<td>Repeated digit</td>
<td>7644 6589&lt;send&gt;</td>
</tr>
<tr>
<td>Row shift</td>
<td>431 3256&lt;send&gt;</td>
</tr>
<tr>
<td>Missing terminator</td>
<td>764 6589</td>
</tr>
</tbody>
</table>

Error classification is difficult because most of the errors can occur any digit, leading to a large number of possibilities.

**Lesson 17: Use the Subjects-in-Tandem Method during the Early Phases of Evaluation to Identify Problems with Products.**

In this method, pairs of people are videotaped while they collaboratively solve a problem to provide insight into the solution process. In the computer industry, the most common problem examined is setup of a personal computer ("the out-of-the-box experience") (Green, 1986; Wiklund, 1994). To work together, participants verbalize what they are planning to do ("First we should stick this blue thing in the hole.") and problems they are experiencing. ("That does not make sense. See, here in the getting started manual it refers to the red one and in the technical manual it says something else.") They also describe solutions. ("This would be a lot simpler if there was one custom set of assembly instructions that said...")

The first on-the-road experiment in this project used the subjects-in-tandem approach (Green, Williams, Hoekstra, George, and Wen, 1993). Without instructions on how to use various in-vehicle interfaces, subjects navigated to a destination. The primary questions were (1) were the
interfaces safe enough to be used by individual drivers in subsequent experiments, and (2) to which information elements in the interface were drivers paying attention? The method was easy to execute and provided useful results. Accordingly, the author recommends that every initial evaluation of a new in-vehicle information system include a brief subjects-in-tandem phase.

LESSONS LEARNED ABOUT GUIDELINES

Lesson 18: Guidelines Should Emerge from Design, Not the Literature.

The original proposal for this project called for developing a set of guidelines by summarizing sets in the literature, such as those of Smith and Mosier (1986), and Military Standard 1472D (U.S. Department of Defense, 1989). Instead, the approach followed was to record all major design decisions as they were made, along with the research data, principle, or other basis for the decisions. This led to a set of guidelines (Green, Levison, Paelke, and Serafin, 1993) that was focused on the subject matter. Had a document been produced by summarizing the literature, it would have been longer, missed some of the automotive-specific literature, contained irrelevant material, and not been very usable.


In many instances, there was no specific literature or computation method to compare two alternative designs of interest. However, often design decisions could be made based on principles. At the outset of this project, it was thought such generalities would not prove to be useful for engineering applications. The opposite was found to be true.


A significant number of design decisions were based on the principle of consistency. Examples included having the sequence of entries on electronic maps (exit number, road name, destination) match the sequence on exit signs, using the same sequence of key entries for different functions, using the same abbreviations for a function in various places, etc. Consistency minimizes the number of mental operations required of a user to understand an interface, enhancing ease of use.

CONCLUSIONS

This project identified several useful ideas concerning evaluation methods and guidelines. With regard to methods, videotape-based simulations should not be used for empirical work and focus groups should occur later in the design cycle than was the case for this project. The use of subjects-in-tandem and Wizard of Oz methods should be expanded. Without use of the Oz method, in conjunction with rapid prototypes, this project would have been impossible to complete. Needing additional attention are methods for collecting data (especially eye fixations), and tools for reducing and analyzing data. Data collected should include baseline (normal) driving data, data necessary to identify the effects of ITS products.

The guidelines development work identified the need for a variety of guideline types (principles, general guidelines, and specific guidelines) and highlighted the usefulness of principles in guiding design. In future guidelines efforts, specific guidelines should be developed as design needs dictate, not because particular issues are covered in the research literature.

It is hoped that following the lessons provided here will speed developments related to ITS technology by eliminating potential false steps in ITS evaluations, encourage the use of methods that are both quick and inexpensive, and accelerate the delivery of research results to engineers.
developing new products. Application of these lessons will cause human factors to be viewed as a positive force in the development of new products and services, rather than an anchor.

REFERENCES


