ABSTRACT

TITLE:
Road-to-Lab: Validation of the Static Load Test for Predicting On-Road Driving Performance While Using Advanced In-Vehicle Information and Communication Devices

OBJECTIVES:
Validate the Static Load Test (Angell, Young, Hankey, and Dingus, 2002) for predicting on-road variables associated with ease-of-use and safety of in-vehicle information and communications devices.

METHODS:
In the Static Load Test, participants perform telematics tasks on an-vehicle device while parked in a garage, while viewing a videotaped road scene on a monitor. No steering or high fidelity simulator is required. Participants also tap the brake when they see a central or peripheral light, mounted on the vehicle, while performing the tasks. In the On-Road Test, the device, tasks, and lights are the same, except a different group of participants drive a vehicle on a closed road while performing the tasks. In both the static and on-road conditions, ten dependent driver performance variables are measured. Various linear and nonlinear statistical models were applied to predict the on-road variables from the static variables. A separate test data sample (not used to generate the model) independently verified the results.

RESULTS:
A linear model was found to predict on-road variables using static test data with low residual error, high percent variance explained, and few errors in classifying tasks as meeting or exceeding criteria for most dependent variables measured ("strong" variables). Some on-road variables were not well predicted by a linear model ("weak" variables). Potential methods for improving the predictions for the "weak" variables, such as nonlinear transformations, will be discussed.

CONCLUSIONS:
The simple, inexpensive, and low-fidelity Static Load Test has been validated for predicting major on-road performance variables commonly used for evaluating whether telematics products are easy and safe to use while driving.

REFERENCES:
ROAD-TO-LAB: VALIDATION OF THE STATIC LOAD TEST FOR PREDICTING ON-ROAD DRIVING PERFORMANCE WHILE USING ADVANCED IN-VEHICLE INFORMATION AND COMMUNICATION DEVICES

Richard Young, Bijaya Aryal, Marius Muresan, Xuru Ding, Steve Oja, S. Noel Simpson
General Motors Engineering

Acknowledgments

• Virginia Tech Transportation Institute
  – Myra Blanco, Jonathan Hankey, M. Lucas Neurauter, Irena Paschai
• GM
  – Jacqueline Chestnut, Lucy La Hood, Mary Fortier, Linda Angell
Outline

• Background
• Objectives
• Methods
• Results
• Future Work
• Conclusions

Background: Metrics Consortiums

• HASTE –
  – HMI And Safety of Traffic in Europe
  – Multiple countries and laboratories
• CAMP
  – Crash Avoidance Metrics Partnership
  – Multiple OEMs and government
• Alliance of Automotive Manufacturers
  – Most automotive OEMs in North America
Current Study Objectives

• Evaluate the *Static Load Test*
  – How well can it predict on-road driver performance while using in-vehicle navigation and telematics systems?

• Produce a linear model for predicting on-road variables from lab data
  – High % variance explained
  – Low residual error
  – Few errors in task classification

Methods: Static Load Test

• Perform secondary tasks on a device in a stationary vehicle or buck
• View real-life road scene video on TV monitor.
• “Keep hands on wheel and eyes on road as if driving, except when accessing the in-vehicle device - Tap brake when you see a light on the hood or on the left side mirror.”
• “Demo – Practice – Test – Test” while parked, for each task.
Static Load Training Trial Example

On-Road Test

- Vehicles, devices, and tasks same as static test
  - But different participants drove the vehicle on a closed road at VTTI while performing the secondary tasks.
- On-road variables same as lab variables.
- 81 seconds to complete task (0.9 mile at 40 MPH)
- “Demo – Practice” while parked, “Test-Test” while driving, for each task.
On-Road Test Example*

Variables Modeled*

Table 1. Variables measured in static and dynamic tests

<table>
<thead>
<tr>
<th>No.</th>
<th>Variable</th>
<th>Name</th>
<th>Dynamic</th>
<th>Static</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Task Completion Time</td>
<td>tasktime</td>
<td>80th%ile</td>
<td>mean</td>
</tr>
<tr>
<td>2</td>
<td>Number of Steps</td>
<td>numsteps</td>
<td>80th%ile</td>
<td>mean</td>
</tr>
<tr>
<td>3</td>
<td>Eyes-Off-Road Time</td>
<td>eort</td>
<td>85th%ile</td>
<td>mean</td>
</tr>
<tr>
<td>4</td>
<td>Number of Glances to the In-Vehicle System</td>
<td>glances</td>
<td>80th%ile</td>
<td>mean</td>
</tr>
<tr>
<td>5</td>
<td>Subjective Workload</td>
<td>workload</td>
<td>80th%ile</td>
<td>mean</td>
</tr>
<tr>
<td>6</td>
<td>Subjective Situation Awareness</td>
<td>situaware</td>
<td>20th%ile</td>
<td>mean</td>
</tr>
<tr>
<td>7</td>
<td>Percent Successful Task Completion</td>
<td>persucc</td>
<td>percent</td>
<td>percent</td>
</tr>
<tr>
<td>8</td>
<td>Percent of Total Visual Events Missed</td>
<td>alimiss</td>
<td>mean</td>
<td>mean</td>
</tr>
<tr>
<td>9</td>
<td>Mean Single Glance Time to System</td>
<td>glancetime</td>
<td>85th%ile</td>
<td>mean</td>
</tr>
<tr>
<td>10</td>
<td>Time to Respond to Visual Events</td>
<td>evtime</td>
<td>mean</td>
<td>mean</td>
</tr>
</tbody>
</table>

Tasks, Vehicles, and Systems

• Tasks varied in type and degree of difficulty
  – Entertainment, Anchor, and Navigation tasks
  – All manual tasks

• Vehicles and Systems
  – Two non-GM 2003 production vehicles with navigation systems for the modeling data and one part of model validation data.
  – One pre-production GM vehicle with a prototype navigation system for the rest of the model validation test data.

• All were matched for dynamic and static tests.

Participants

• Participants screened for hearing and vision normalcy, licensed drivers

• Multiple task sets:
  – Each had different group of 10-16 participants
  – Each counterbalanced for participant age (25-44 younger, 45-65 older) and gender.
  – Run statically and dynamically
    • different groups of participants
    • about 120 participants total.

• Tasks presented in random order
• Two test trials per participant.
Model Data

- 31 task pairs on 2 devices
  - Static data from MPG Usability Lab
  - Dynamic data from VTTI Smart Road
- Number of “analytical” steps in a task ranged from 1 to 26.
  - Wide range of task difficulties

Validation Data

- 42 task pairs on 3 devices
  - Static data
    - MPG Usability Lab (81 sec limit)
    - VTTI garage
      - 180 sec limit caused 4 outliers in results
  - Dynamic data
    - VTTI Smart Road (81 sec limit)
Models Tested

- **One-to-One Model**
  - Often implicitly assumed

- **Simple Linear Regression (SLR) Model**
  - Single static variable model

- **Multiple Linear Regression (MLR) Model**
  - Multiple static variable model

- **Partial Least Squares (PLS) Model**
  - Controls for co-linearity

Metrics for Model **Accuracy**

Based on *model* data points:

- **Adjusted $R^2$**
  - % total variance explained, adjusted for df

- **$s$**
  - Standard deviation of the regression

**Task Classification Errors**

- Errors in predicting tasks as meeting or not meeting on-road criteria.

**Strong Models**

- High $R^2$, low $s$, few errors

**Weak Models**

- Low $R^2$, high $s$, many errors
Metrics for Model Validation

Based on validation data points:

- **Test R²**
  - Similar to Adjusted R² for model
- **Root Mean Square Error (RMS)**
  - Similar to s metric for model
- **Task Classification Errors**
  - Same as task classification errors for model
- **Strong and Weak Model Validations**
  - Same as for model accuracy
Results *Tasktime*: 1-1 Model

- The 1-1 model (assumes static = dynamic) had 90% variance explained from predicting the on-road 80\textsuperscript{th} \%ile tasktime from the static 80\textsuperscript{th} \%ile tasktime
- The standard deviation of tasktime for the 1-1 model was 5.4 seconds
- The 1-1 model produced a poorer fit to the data than the regression models.
Tasktime: Multiple Regression (MLR) Model

MLR MODEL: Predicted vs Observed On-Road Tasktime 80th%ile

adj. $R^2 = 99\%$, $s = 1.7$

MLR Model:
$R^2 = 99\%$, $s = 1.7$

Tasktime: SLR Model Validation

Validation: SLR Predicted vs Observed On-Road Tasktime 80th%ile

Test $R^2 = 61\%$, $RMS = 5.6$ sec

5 false alarms (1 outlier)
**Tasktime: MLR Model Validation**

Validation: MLR Predicted vs Observed On-Road Tasktime 80th%ile

- **Test**
  - $R^2 = 97\%$
  - RMS = 3.6 sec
  - 2 misses

**ON-ROAD OBSERVED**

**Numsteps: MLR Model Validation**

Validation: MLR Predicted vs Observed On-Road Steps 80th%ile

- **Test**
  - $R^2 = 98\%$
  - RMS = 1.5 steps
  - 2 false alarm (1 outlier)

**ON-ROAD OBSERVED**
Summary: *tasktime, numsteps*

- Model and validation test prediction results for lab to road
  - Multivariate MLR model generally did best

### Overall Task Classification Errors

<table>
<thead>
<tr>
<th>Class</th>
<th>Variable</th>
<th>1-1</th>
<th>SLR</th>
<th>MLR</th>
</tr>
</thead>
<tbody>
<tr>
<td>strong</td>
<td>tasktime</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>numsteps</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>eort</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>glances</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>worklolad</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>sitAware</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td></td>
<td>11</td>
<td>12</td>
<td>6</td>
</tr>
<tr>
<td>weak</td>
<td>perSucc</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>allmiss</td>
<td>9</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>glancedur</td>
<td>12</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>eventtime</td>
<td>9</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td></td>
<td>31</td>
<td>25</td>
<td>17</td>
</tr>
<tr>
<td><strong>GRAND TOTAL</strong></td>
<td></td>
<td>42</td>
<td>37</td>
<td>23</td>
</tr>
</tbody>
</table>

---

6/27/2005 Road-to_Lab 25

---

6/27/2005 Road-to_Lab 26
Strong Models

• The following on-road variables were strongly predicted by all models (except one-to-one):
  1. Tasktime
  2. Numsteps
  3. Eyes-off-road time
  4. Glances
  5. Subjective workload
• The prediction models were strong for both the original model and the validation data.
• The MLR method tended to outperform the SLR method and the one-to-one method.

Weak Models

• The following on-road variables were weakly predicted by all models:
  1. perSucc (percent successful task completion)
  2. glanceDur (glance duration)
  3. Allmiss (percent missed events)
  4. evnttime (brake reaction time)
• All weak models had
  – Low model $R^2$ values
  – High s values
  – Many task classification errors
Task Classification Errors – Conjoint On-Road Criteria

• More than 1 variable is required to describe on-road driver performance, which is multi-dimensional.*
• Set multiple on-road variable criteria in a conjoint manner
  • The number of classification errors can be substantially reduced – in current test, goes to zero.


Future Work

• Improve the predictive capability of the weak models.
• Determine the fundamental neural mechanisms underlying driver event detection and response with secondary task distraction (Young, Hsieh et al., 2005*)
  – Improve the validity of static response time and missed events for predicting on-road event detection driver performance.

Conclusions

1. The Static Load Test is valid for predicting a number of on-road driver performance variables during secondary tasks (using the *strong* models described here).
2. Multivariate methods make more accurate on-road predictions than univariate methods.
3. Conjoint multivariate criteria reduce task classification errors.
4. Event detection methods require further work to make valid on-road predictions.

Thank you for your attention!