Resurrecting driver workload metrics: A multivariate approach

Jack L. Auflick*

Engineering Systems Inc., 1174 Oak Valley Drive, Ann Arbor, MI 48103 USA

Abstract

This paper presents new, multivariate analyses of data collected during the Driver Workload Metrics (DWM) project. In a cooperative effort with the National Highway Transportation Safety Administration, the DWM project had several goals including the development of performance metrics and test procedures to assess visual, manual, and cognitive aspects of driver workload. Workload was defined as the competition in driver resources (perceptual, cognitive, or physical) between the driving task and a concurrent secondary task, occurring over that task’s duration. It was hypothesized that, depending on the type of secondary task performed while driving, measured workload and the correlated quality of driving should either remain the same or decline, but would manifest in degraded measures of lane keeping, longitudinal control, or eye glance behavior. However, the original DWM project had an unrealized goal, i.e. to apply Exploratory Factor Analysis (EFA) methods, in an attempt to uncover the underlying unobserved structure within the project’s relatively large set of variables. It is this hidden multi-dimensional structure that must be examined to empirically comprehend the concept of driver workload. DWM kinematic vehicle data, driving performance, and eye glance data were reanalyzed using Maximum Likelihood Factor Analysis (MLFA). These analyses found that task-induced workload affected driving performance and was multi-dimensional in nature. Visual-manual tasks exhibited fundamentally different performance profiles than auditory-vocal tasks or just driving. Furthermore, when secondary statistical analyses of the normalized factor scores were done using Multivariate Analysis of Variance (MANOVA) the results found highly statistically significant workload differences in age groups, task type, and at times, gender.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: Driver workload; Driver distraction; Exploratory Factor Analysis

* Corresponding author. Tel.: +1-734-794-8103; fax: +1-734-794-8115.
E-mail address: jlauflick@esi-mi.com
1. Introduction

In recent years, technological advances in the automotive world have brought a number of new systems and devices into play often suggesting that they will provide enhanced safety for drivers. On one hand, advanced telematics systems, like navigation and route guidance, collision avoidance systems, or intelligent cruise control tout improved safety. On the other hand, these systems and other handheld devices like cell phones or tablets have generated fears from researchers and regulatory agencies that the use of these new devices within the automotive cockpit, will at times, overload and distract the driver. In attempts to understand the dimensions of driver distraction, agencies such as the National Highway Transportation Safety Administration (NHTSA), universities, international agencies and private corporations have funded large research efforts, like the Human Machine Interface and the Safety of Traffic in Europe (HASTE) [1], the Naturalistic Driving Program at Virginia Tech [2], the European Advanced Driver Attention Metrics Program [3], the Crash Avoidance Driver Metrics, and the Driver Workload Metrics (DWM) [4].

In a cooperative effort with NHTSA, the DWM project¹ investigated and established correlates between the demands placed on the driver from secondary discretionary tasks that had the potential to interfere with the primary driving task. Workload was defined as the competition in driver resources (perceptual, cognitive, or physical) between the driving task and a concurrent secondary task, occurring over that task’s duration. It was hypothesized that, depending on the type of secondary task performed while driving, measured workload and the correlated quality of driving would manifest in degraded measures of lane keeping, longitudinal control, object-and-event detection, or eye glance behavior.

Based on an extensive review of the driver workload literature and Multiple Resource Theory (MRT) [5], DWM created a set of fourteen conventional experimental tasks commonly performed in vehicles today. DWM classified these tasks by the input and output modalities needed to perform the task: either visual input and manual output, or auditory input and vocal output [6].

The tasks were comprised of seven visual-manual tasks and seven auditory-vocal tasks. Examples of the visual-manual tasks included the CD7 task that directed the subjects to select a CD from the visor, place it in the CD player, and tune to track seven. The seven auditory-vocal tasks included a biographic task where the experimenter asked the subjects a series of questions intended to elicit a verbal response. Questions such as, “Where do you live?” and “How many children do you have?” were included in this section. There was also a baseline, “Just Drive” task and a combination task where subjects picked up a cell phone, pressed a preset button, and then interacted with an automated voice recognition flight schedule service. This project accomplished several goals including:

1. Development of performance metrics and test procedures that reliably assessed how driving performance may be negatively affected due to auditory, visual, manual, and cognitive aspects of driver workload associated with using in-vehicle systems, and
2. Creation of a metrics toolkit that could be used by automotive engineers during all stages of the design process to assess the implications of driver workload while using future systems.

1.1 Multi-dimensional interrelationships

However workload as the construct of interest (a transient mental state), had no direct measure that was available at the time the DWM study was initiated. As a result, the existence of driver workload was inferred from an extensive string of bivariate correlation and regression analyses, observed across multiple measures of performance. These individual analyses provided brief glimpses into the complex, hidden inter-relationships within the collected data, but had an unrealized goal, i.e. to apply EFA methods in an attempt to uncover the latent, unobserved structure

¹The Driver Workload Metrics project, a co-operative agreement between the NHTSA, Ford, GM, Nissan, and Toyota, was conducted under the Crash Avoidance Metrics Partnership (CAMP), a partnership established by Ford and GM to undertake joint precompetitive work in advanced collision avoidance systems.
of correlated measures of workload. It is this hidden structure that must be examined to empirically comprehend the multi-dimensional concept of driver workload.

As explained in the DWM final report, driver performance data was collected from 108 test subjects using a canonical repeated measures experimental design where each test subject repeatedly performed the secondary tasks during real driving on an interstate highway. The sample of participants was approximately balanced by gender and age. Task sequences and presentations were randomized for each subject. Based on analysis of distraction-related crash data, the driving condition selected for testing was a highway speed, car-following scenario on a straight level road under clear, dry, daytime conditions.

During testing, subjects drove an instrumented car that was the center car of a three-vehicle platoon. This platoon operated as a single testing unit and provided a realistic car-following driving experience. During each task, an extensive array of sensors, cameras, and on-board instrumentation recorded kinematic data for longitudinal and lateral vehicle control, including task duration, speed, range, range rate, time headway, and time to contact. For each of the kinematic variables, data was captured for mean, median, standard deviation, minimum and maximum distances, and time durations (at the minimum or maximum). In addition, steering wheel behavior was measured for three variables, for example, how much time during a task the steering wheel was held beyond a zero-degree location, and how long it was held at either a 15 or 20 degree offset from zero degrees, including the number of lane exceedances and durations, plus the number of center stripe touches and their durations.

Finally, driver eyeglance patterns were also recorded for glances to the road, mirror, situation awareness (outside left or right), task related (glances to in-cockpit locations during visual manual tasks), not road, and to not assigned (other) locations. These eye glance metrics included median glance duration, standard deviations, and the percent of a task’s duration attributed to that specific location. Eye glances were manually scored through review of on-board video taken during each task. However, due to timing and funding constraints, only 42 of the 108 test subjects had complete eye data at the time the final DWM analyses were completed.

2. Factor Analysis

This new analysis used a subset of 42 vehicle kinematic variables (like speed, range, range rate, longitudinal and later lane positions, etc.), time variables (like time for the vehicle at minimum or maximum lane position right and left, plus minimum and maximum times for the kinematic variables), and the steering wheel position and durations. The analysis also included the subset of test subjects with complete eye glance data, using eye glance variables (such as median, standard deviation, and percent task duration, etc. from on-road, mirror, situation awareness, task-related, and not-road or not-assigned locations). The purpose of this analysis was to extract the underlying latent, highly inter-correlated structure within these data, believed to be workload, caused by the variation in driving performance during multitasking. Specifically, this analysis used the maximum likelihood factor analysis (MLFA) method [7, 8] one of several commonly used EFA techniques that in general, seeks to: (1) reduce the number of variables and (2) to detect unobserved structure (latent variables) in the relationships between variables. MLFA approaches this with general methods that estimate the loadings and communalities in a data set and then maximizes the probability of the observed correlation matrix occurring. In MLFA, maximizing the likelihood function determines the parameters that are most likely to produce the observed data.

2.1. Hidden Structure

The first step of MFLA identified how many factors should be retained in the analysis. In factor analysis, a factor is a latent (unmeasured) variable that expresses itself through its relationship with other measured, observable variables. In this analysis, MFLA was used in an iterative fashion starting with all 42 independent variables, meaning that there could have been up to 42 orthogonal or independent factors. Then, using both the Kaiser criterion [9] (i.e. retain only Eigenvalues greater than 1.0) and Cattell’s scree plots [10], initial results showed that with this set of variables, there were only seven new orthogonal factors with Eigenvalues greater than 1.0. These seven factors described the hidden structure within the data. Given this information, MFLA was run a second time using a
maximum of seven factors, maximizing the likelihood for the observed correlation matrix. The Eigenvalues from these factors measured the variance from all the variables that were accounted for by a given factor. As seen in Table 1 below, the seven factors explain approximately 66% of the original variance in the data.

The initial solution left the factors and factor scores (i.e. estimates of actual values of individual observations for each factor) in an un-rotated seven-dimensional space making the given solution difficult to interpret. Factor loadings are correlations of an observed variable with the underlying factor and they are conceptually similar to Pearson's \( r \), the common correlation coefficient. When the loadings are squared, the resulting values describe the percentage of variance within a specific independent variable explained by a given factor. Based on their individual directionality and magnitude, factor loadings provide important insights when interpreting the MFLA solutions. As an aid for interpretation, factor scores were rotated and normalized using the varimax orthogonal transformation [11, 12]. Varimax rotations were so named because the process maximizes the sum of the variances (e.g. Vari-Max) of the squared loadings (i.e. the squared correlations between variables and factors). A varimax rotation makes the output more understandable, by maintaining the orthogonality of the factor axes but maximizing the variance of the squared loadings in the factor matrix (i.e. correlation coefficients between the cases (rows) and factors (columns)). These rotated factors identify the simple structure or hidden relationships within the unobserved configuration in the data, if such structure exists.

Because the rotated factors scores and loadings indicate how each hidden factor is associated with the observable variables, the loadings must be interpreted through a highly subjective activity where “names” are applied to each factor based on the magnitude and direction of the rotated loadings. While there are several approaches on how to do this, a generic rule of thumb has been developed suggesting that interpretations should be done only on those loadings exceeding \(|0.7|\). The rational for this is that when a loading of 0.7 or higher is squared, about half of the variance in that variable is being explained by the factor. Table 1, shown above, contains the names applied to the seven factors that were derived through analysis of loadings on each of the seven factors.

During the MLFA process explained above, individual scores for each test subject on each variable and task were normalized, and then rotated using the varimax approach. This resulted in a matrix of factor scores representing numerical values defining a person's relative spacing or standing on each of the seven latent factors. Test subjects’ factor scores on each factor were averaged by gender, age group, and task type, and were then plotted on radar plots. These plots present a multi-dimensional driving workload profile for different groupings of subjects based on gender, age group, or type of task. As can be seen in the following figures, there were significant, discernable differences in workload in relation to type of task, gender, and age. Figure 1 below presents a profile graphically comparing males and females. Note that Factor 1, the Eyes up-Aud. Vocal Work Load, is in the 12 o’clock position while Factor 2 through 7 rotate in a clockwise direction.

Figure 2 below presents a similar profile showing differences in driving workload due to the effect of age group. Depending on the factor, there are discernable and significant differences in workload due to the age group of the subject. One can see that on some factors, older and younger drivers differ completely. For example, while all three age groups appear similar on Factor 1, on Factor 2, Mirror and Situation Awareness glances, younger drivers have a higher positive average score while the older age group is at the opposite pole in the negative range. This suggests that older drivers are making significantly fewer eye glances to mirrors and situation awareness locations.
Figure 3 graphically compares workload differences due to the effect of task type. As above, depending on the factor, there are discernable differences in workload. For example, on Factor 1, the Eyes up-Auditory Vocal Workload, the average factor scores for auditory-vocal tasks and Just Drive are relatively higher positive values while the average factor score for visual-manual tasks is in the negative range. Graphically, this suggests that when test subjects were performing the seven visual manual tasks, they had significantly fewer glances to the forward road scene when compared to the same drivers performing auditory-vocal tasks.

2.2. Multivariate Analysis of Variance (MANOVA)

MANOVA is a generalized form of analysis of variance where there are two or more dependent variables. However, instead of testing univariate means, MANOVA tests the statistical significance of the variance-covariance differences between groups. Because the radar plots from the averaged factor scores graphically demonstrated
significantly workload differences due to age, gender, and task type, a MANOVA was done on the original data set to further explain “workload”, i.e., the meaning from the latent factors. The MANOVA tested hypotheses that there were statistically significant differences between gender, age groups (Younger 19-39, Middle 40-59, Older 60-79), and task type (Visual-Manual, Auditory-Vocal, Just Drive, Combination). The 42 kinematic and timing variables were the independent variables with gender, age, and task type serving as the dependent, categorical variables. Table 2 below presents the MANOVA results. Values in italics indicate statistical significance. As can be seen, there were highly statistically significant main effects for gender, age group, and task type. In addition, there were highly significant two-way interactions between task type and age group, and gender and age group. As shown in the figures below, results found that visual-manual tasks exhibited fundamentally different performance profiles than auditory-vocal tasks or just driving.

These results from the MANOVA affirm the qualitative differences as shown in the proceeding figures. Given the original data, there are notable differences in driving workload as defined by the dependent grouping variables. This overall MANOVA analysis is still undergoing interpretation through examination of the univariate statistics for each of the 42 independent variables. Figure 4 below presents a preliminary look at one statistically significant comparison from the GenderName by AgeGroup interaction for Median Task-related Duration (medTRDur). The medTRDur was one of several independent variables that contributed to the naming of Factor 1. Post-Hoc significance tests used the Tukey Unequal N [13] for “medTRDur,” in the GenderName by AgeGroup Interaction found that older females differ significantly from young males, young females, and middle females with respect to the length of task-related glances while performing visual-manual tasks.

Table 2. MANOVA results.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Test</th>
<th>Value</th>
<th>F</th>
<th>Effect(df)</th>
<th>Error(df)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>Wilks</td>
<td>0.000005</td>
<td>2095474</td>
<td>42</td>
<td>417.000</td>
<td>0.000000</td>
</tr>
<tr>
<td>TaskType</td>
<td>Wilks</td>
<td>0.020743</td>
<td>26</td>
<td>126</td>
<td>1250.347</td>
<td>0.000000</td>
</tr>
<tr>
<td>GenderName</td>
<td>Wilks</td>
<td>0.854425</td>
<td>2</td>
<td>42</td>
<td>417.000</td>
<td>0.005794</td>
</tr>
<tr>
<td>AgeGroup</td>
<td>Wilks</td>
<td>0.634314</td>
<td>3</td>
<td>84</td>
<td>834.000</td>
<td>0.000000</td>
</tr>
<tr>
<td>TaskType*GenderName</td>
<td>Wilks</td>
<td>0.721723</td>
<td>1</td>
<td>126</td>
<td>1250.347</td>
<td>0.146480</td>
</tr>
<tr>
<td>TaskType*AgeGroup</td>
<td>Wilks</td>
<td>0.481782</td>
<td>1</td>
<td>252</td>
<td>2489.048</td>
<td>0.002205</td>
</tr>
<tr>
<td>GenderName*AgeGroup</td>
<td>Wilks</td>
<td>0.684612</td>
<td>2</td>
<td>84</td>
<td>834.000</td>
<td>0.000000</td>
</tr>
<tr>
<td>TaskType<em>GenderName</em>AgeGroup</td>
<td>Wilks</td>
<td>0.547387</td>
<td>1</td>
<td>252</td>
<td>2489.048</td>
<td>0.279897</td>
</tr>
</tbody>
</table>
Figure 5 below presents one of several two-way interactions between TaskType and AgeGroup for the Median Task-related Duration variable. Post-Hoc comparisons were done again using the Tukey Unequal N test. Results found that there are significant differences in TaskType between the Combo Task, Visual Manual Tasks, and Auditory Vocal-Just Drive. There were no significant differences between Auditory-Vocal and Just Drive comparisons. Age groups are similar within each task type with the exception of Visual Manual Tasks. In this comparison, younger subjects differed from middle age subjects and both groups differed significantly for older subjects. Older subjects spent much longer glance durations to Task-Related locations while doing Visual-Manual Tasks.

3. Discussion - conclusions

The purpose of the MFLA was to examine the large DWM data set trying to identify hidden structure that was indicative of driver workload. MLFA found that these DWM data contained extensive multi-colinearity in the data set, hidden relationships, i.e. the latent structure that could be described by a minimum of seven factors while still being able to explain ~66% of the original variation in the data. The seven underlying factors began to reveal insights into key effects of multitasking on driving performance. Driver workload caused distraction or interference.
while driving was shown to be multidimensional in nature, meaning that it was represented in the data by simultaneous effects on multiple variables. Driver workload also was reflective of allocations of driver resources across input modalities, output modalities, working memory, and central attention, as well as being affected by task type, gender, or age group. The MANOVA confirmed that within these data there are highly significant main effects due to gender, age group, and task type, as well as significant two-way interactions between task type and gender plus task type and age group.

In applying this methodology, however, there was an explicit recognition that it was exploratory in nature, and that the underlying dimensions it identified would need to be attributed with meaning and interpreted through subjective analysis. It is worthwhile to reiterate that this is exploratory work and that the nature of the dimensions could change if the input to the analysis were different. Similarly, the interpretations of the underlying dimensions may be refined as a deeper understanding of the data set is acquired over time. Finally, the full DWM data contains a second data set where 69 drivers, driving on a test track performed twenty-two in-vehicle tasks, plus the task of just driving, using the same experimental design. These data should be analyzed using the same MFLA approach and then compared back to the results from the current analyses. Assuming both data provide similar results, one could then begin the laborious process of using Confirmatory Factor Analysis (CFA), a special form of factor analysis, used to test whether measures of a construct are consistent with a researcher's understanding of the nature of that factor. CFA tests whether the data fit a hypothesized measurement model where the hypothesized model is based on theory and/or previous analytic research.

Acknowledgements

Much praise goes to the Driver Workload Metrics project team: L. Angell, P.A. Austria, D. Kochhar, L. Tijerina, W. Biever, T. Diptiman, J. Hogsett, and S. Kiger. They endured much but made a significant contribution to the understanding of driver workload.

References