

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/241582192>

Detecting short periods of elevated workload. A comparison of nine workload assessment techniques

Article in *Journal of Experimental Psychology Applied* · September 1996

DOI: 10.1037//1076-898X.2.3.270

CITATIONS

10

READS

277

2 authors, including:



[Willem B Verwey](#)

Universiteit Twente; A&M University

152 PUBLICATIONS 4,291 CITATIONS

SEE PROFILE

Detecting Short Periods of Elevated Workload: A Comparison of Nine Workload Assessment Techniques

Willem B. Verwey and Hans A. Veltman
TNO Human Factors Research Institute

The present experiment tested the merits of 9 common workload assessment techniques with relatively short periods of workload in a car-driving task. Twelve participants drove an instrumented car and performed a visually loading task and a mentally loading task for 10, 30, and 60 s. The results show that 10-s periods of visual and mental workload can be measured successfully with subjective ratings and secondary task performance. With respect to longer loading periods (30 and 60 s), steering frequency was found to be sensitive to visual workload, and skin conductance response (SCR) was sensitive to mental workload. The results lead to preliminary guidelines that will help applied researchers to determine which techniques are best suited for assessing visual and mental workload.

In recent years, major research programs in Europe, the United States, and Japan have boosted the development of information systems in road traffic. In Europe, these systems are referred to as *transport telematics*, but in the United States the term *intelligent transport systems* (ITS) is more common. It is generally expected that transport telematic applications will change the nature of car driving considerably in the near future. A vast array of telematics systems is currently being considered. Some well-known examples are route guidance systems, anticollision systems, systems for monitoring driver status, and parking advice systems. However, installing even a subset of these telematic systems, many of which are aimed at supporting the driver and increasing traffic safety, may lead to distraction and overload of the driver and will result in increased accident rates. Therefore, care should be taken to design telematics

systems according to a driver's needs and capabilities (e.g., Michon, 1993). This implies that the human factors specialist should have detailed knowledge of the advantages and disadvantages of the existing techniques used for assessing driver workload.

In a review of workload assessment methodologies, O'Donnell and Eggemeier (1986) proposed five criteria for selecting workload assessment techniques. Two major criteria were sensitivity (whether the technique discriminates between levels of workload) and diagnosticity (whether the technique distinguishes between types of workload). The purpose of the present article is to compare a series of nine common workload assessment techniques to relatively short periods of elevated workload in driving. Special emphasis is given to the sensitivity and diagnosticity of these techniques.

Several studies have compared workload assessment techniques. These studies have yielded different results with respect to sensitivity of these techniques. For example, Casali and Wierwille (1984) compared workload assessment techniques in a flight simulator study. Hicks and Wierwille (1979) did the same in a driving simulator study. In both studies, subjective rating scales and primary task performance were able to quantify workload of the primary task. However, secondary task performance was a poor workload indicator in the

Willem B. Verwey and Hans A. Veltman, Human Factors Research Institute by Toegepast Natuurwetenschappelijk Onderzoek (TNO), Soesterberg, The Netherlands.

This work was funded in part under Contract No. DRIVE V2065 (GEM) of the European Commission.

Correspondence concerning this article should be addressed to Willem B. Verwey, TNO Human Factors Research Institute, P.O. Box 23, NL 3769 ZG Soesterberg, The Netherlands. Electronic mail may be sent via Internet to verwey@tm.tno.nl.

second study. Neither study showed an effect of workload on heart rate (HR) and HR variability, but still other studies did show that HR measures were sensitive indicators for workload caused by traffic jams and bad weather (Vivoli, Bergomi, Rovesti, Carrozzi, & Vezzosi, 1993) and workload caused by telephoning while driving (Brookhuis, de Vries, & de Waard, 1991). The outcomes of these studies illustrate that there are marked sensitivity differences among techniques in different tasks and in different driving environments.

A cause for the differences between studies may be that the techniques differ with respect to their sensitivity to rapid fluctuations in workload level. For example, if workload is assessed in a task involving short workload peaks and a technique is used that is highly sensitive to peaks, the study will show a generally high level of workload. When a technique is used that is not sensitive to peaks, the conclusion will be that workload is low. However, in tasks involving only gradual changes in workload, both techniques may indicate similar workload levels.

The sensitivity of workload assessment techniques to workload peaks has received some attention (e.g., Knowles, 1963; Wierwille, 1981), but empirical studies are scarce. Antin and Wierwille (1984) compared the sensitivity of six workload assessment techniques with respect to the workload induced by two laboratory tasks. Level of workload changed rapidly in these tasks. Primary task performance and subjective estimates yielded the best results, but the duration of workload elevations was not systematically manipulated. It remains unclear how long workload increments must last for them to be detected by the various techniques. Moreover, it is unclear to what extent these results can be extrapolated to the driving environment, which is the task environment of primary interest in this article.

In some models, there are various types of workload. For example, Wickens (1984b) distinguished workload on perceptual and response modalities, on cognitive coding mechanisms, and on processing stages. In line with this model, Verwey (1993) demonstrated that during driving visual and mental workload may diverge. Visual workload was relatively high in situations involving changes in course, such as at turns and on curves. Mental workload was affected much less in these situations. This shows that a workload assessment

technique cannot be selected without taking the question at hand into account. For example, when workload is used to indicate error probability, another technique is required than when exploring the possibilities to add a task to a task environment. In the latter case, a technique is required to allow the researcher to diagnose types of elevated workload. Estimating error probability requires a technique sensitive to all types of workload. However, little is known about the diagnosticity of the various workload assessment techniques, that is, to what extent the various techniques are sensitive to one or another type of workload.

The aim of the present study was to compare the sensitivity and diagnosticity of a series of workload assessment techniques for short periods of elevated visual and mental workload. Experienced drivers operated an instrumented car on a straight four-lane freeway. This task involved the basic elements of driving, that is, lane keeping and controlling speed. To increase workload in a controlled manner, we added a visually loading task and a mentally loading task to the driving task. These are called *loading tasks*. Both loading tasks lasted for 10, 30, or 60 s. Even though peaks shorter than 10 s can occur in driving, workload elevations of 10 s are still considered *peaks*. Periods of 30 and 60 s of elevated workload were introduced to obtain an indication of the sensitivity and diagnosticity of the workload assessment techniques for longer periods of high workload.

Workload was assessed with nine different techniques: two primary task measures, one secondary task measure, four physiological measures, and two subjective measures. Primary task measures were driving speed and the intervals between successive steering movements. Speed has been found to decrease with increasing workload (Harms, 1986; Verwey & Janssen, 1989). Steering frequency increased under such conditions (Antin, Dingus, Hulse, & Wierwille, 1990; MacDonald & Hoffman, 1980; Verwey, 1991).

A fairly common technique for measuring workload is secondary task performance. The rationale is that performance on a low-priority task reflects the workload induced by a concurrent high-priority task (e.g., Noy, 1987). To ensure that participants acknowledge this priority and continue doing so, the researcher should repeatedly tell them to perform the secondary task only if the primary task performance does not deteriorate

(Wickens, 1984a). The diagnosticity of the secondary task technique depends on the characteristics of the secondary task being used (O'Donnell & Eggemeier, 1986; Wickens, 1984b). A secondary task generating substantial visual workload is diagnostic for the visual workload of the primary task. A mentally loading secondary task is diagnostic for mental workload.

Various physiological measures for workload were used: two measures derived from the electrocardiogram, including interbeat interval (IBI) and HR variability, the interval between successive eyeblinks, and the interval between skin conductance responses (SCRs). An increase in workload would result in a decrease of IBI and HR variability because operators increase their level of arousal or mental effort in order to accomplish their goals (Gaillard & Wientjes, 1994; Hockey, 1986).

Eyeblink intervals were measured because people are known to suppress the eyeblink reflex when they have to process visual information (Stern, Boyer, & Schroeder, 1994; Veltnan & Gaillard, 1996; Wilson & Fisher, 1990). The extent to which mental workload affects eyeblinks is not entirely clear. There are indications that eyeblinks do not respond to mental task demands (Casali & Wierwille, 1984; Wierwille & Eggemeier, 1993). This suggests that eyeblinks are diagnostic with respect to visual workload.

A measure that is traditionally associated with workload, and especially with arousal states accompanied by effort and emotions, is the spontaneous electrodermal activity or SCR (e.g., Edelberg, 1972). In a classic study, Taylor (1964) found a relationship between the number of SCRs and accident rate, number of turns per distance unit, and driving speed. Zeier (1979) observed that driving in heavy city traffic with a manual transmission yielded more SCRs than driving with an automatic transmission, which would suggest that SCR indexes activation of the sympathetic nervous system.

Because physiological measures of workload are often associated with arousal and emotions, and consequently effort, one might expect that SCR, IBI, HR variability, and eyeblinks are closely related and have limited correlations with performance-based measures of workload (Casali & Wierwille, 1984; Hart & Wickens, 1990). The reason is that effort and task performance need not coincide. It is important to find whether these

physiological measures will be sensitive to short workload periods in driving, as many previous studies showed either limited or no sensitivity of physiological measures (Wierwille, 1979).

Finally, workload was assessed by two subjective rating scales: the Subjective Workload Assessment Technique (SWAT) and the Rating Scale Mental Effort (RSME). The SWAT (Reid & Nygren, 1988) is based on the premise that workload is a multidimensional concept that can be assessed by asking participants for their workload in terms of time pressure, mental effort, and psychological stress. Assessing workload on these three dimensions would increase the sensitivity and the diagnosticity of this technique. In addition, the SWAT takes individual differences into account. Participants are requested to perform a card-sorting procedure that identifies relative sensitivity among these dimensions. The RSME (Zijlstra, 1993) is a univariate rating scale. This scale ranges from 0 to 150 and has nine descriptive indicators along its axis (e.g., *not effortful* and *awfully effortful*). Validation of this technique ensures that the meaning of the verbal labels are the same for different people.

A multidimensional and a univariate scale were included because there has been a dispute about the usefulness for rating scales to have more than one dimension. Comparative evaluations of subjective scales conducted in a wide range of tasks showed little advantage for the multidimensional scales (see Wierwille & Eggemeier, 1993, for a review). In addition, there is mounting evidence that taking individual differences into account by way of the sorting task adds little or no sensitivity. However, because this has been found with rather homogeneous participant groups such as pilots, the question of whether individual differences are important in more heterogeneous groups, such as car drivers remains.

In summary, the present study compared nine workload assessment techniques in the driving environment with respect to their sensitivity and their diagnosticity to 10-, 30-, and 60-s periods of elevated workload. These techniques encompassed two driving parameters (speed and steering intervals), secondary task performance (visual detection), two rating scales (SWAT and RSME), and four physiological measures (IBI, HR variability, eyeblink intervals, and SCR). Detailed insight into the merits of these workload assessment techniques for the driving task will aid in the

design of future studies of driver workload and be of theoretical interest for understanding workload.

Method

Driving and Loading Tasks

The experimental route consisted of 40-km, four-lane freeway connecting two Dutch cities, Amersfoort and Apeldoorn. This road is characterized by relatively stable traffic with little or no congestion. It took 25 to 30 min to drive the experimental route in a single direction. Participants were instructed to drive 80 to 90 kph in the right lane and to avoid overtaking.

To impose workload for predefined periods of time on top of the workload caused by lane and speed keeping, participants carried out loading tasks for 10, 30, or 60 s. Peaks below 10 s seemed too short to be registered by the present techniques. The 30- and 60-s periods were included to assess sensitivity and diagnosticity to longer periods of increased workload as well. One loading task used a set of four high luminance light-emitting diodes (LEDs) placed in a rectangular form. Only three of the LEDs were switched on at the time. The resulting shape changed each 1 to 2 s. There was a probability of .25 that the LEDs formed an L shape. Participants were to detect and count the number of times the L shape was shown. This task was called the *L-counting task*.

The second loading task was the *continuous memory task* (CMT). The CMT requires participants to count each of a number of target letters in a stream of spoken letters (Boer & Jorna, 1987). In the present study, the targets were *A*, *B*, and *C*. Targets had a higher presentation probability than nontargets; for each nontarget presentation probability was .03; for each target this probability was .14. In effect, there was a probability of .42 that a target was presented. The interval between the onset of successive letters was 2 s.

Participants were instructed to pay as much attention to each of these two loading tasks as possible and not to miss any of the targets. The start of each loading task was indicated by a vocal warning from the loudspeaker. The high priority of the CMT and the L-counting task was emphasized by performance feedback immediately upon responding at the end of each loading task period.

Workload Assessment Techniques

Nine techniques were used to assess the workload of the driver. Two involved primary task measures, that is, indicators of driving performance that are known to be affected by driver workload. The first primary task measure was driving speed. Second, a measure of steering frequency called *steering reversal rate* (SRR) was assessed by analyzing the average time between successive wheel movements. A *movement* was defined as a change from a negative (clockwise movement) to a positive (counterclockwise movement) rotational velocity, given that the positive rotational velocity exceeded 1° per second.

The secondary task technique for assessing workload (not to be confused with the loading tasks) was a visual detection task. It required participants to say the Dutch equivalent of "yes" (i.e., "ja") upon detecting a target stimulus (two random punctuation marks) on a dashboard-mounted display. To prevent peripheral detection of stimuli, we presented a neutral stimulus (called "gg") between the target stimuli. The intervals between onset of succeeding messages varied randomly between 2 and 4 s. Presentation time was 750 ms. Participants were urged to give this task low priority relative to all other tasks, including the loading task. Care was taken to make sure that there was no synchronization between stimulus presentation in the secondary task and in the loading tasks. The secondary task continued irrespective of whether a loading task was carried out.

A third group of workload assessment techniques involved the SWAT and RSME rating scales. The SWAT is based on participant ratings from 1 to 3 on time pressure, effort, and stress. Individual differences are taken into account by integrating the results of a card-sorting procedure that identifies relative weight of each of these dimensions. The RSME is a one-dimensional scale with ratings between 0 and 150. The scale has nine descriptive indicators along its axis (e.g., 2 corresponds to *not effortful*, 58 to *rather effortful*, and 113 to *awfully effortful*). It is designed so as to minimize individual differences.

Finally, there were four different physiological measures for workload: IBI, HR variability, intervals between successive eyeblinks, and intervals between successive SCRs. IBI and HR variability

were derived from the electrocardiogram (ECG) and eyeblinks from the electrooculogram (EOG).

Apparatus and Data Collection

The experiment was carried out in an instrumented Volvo 240 station wagon with dual controls. This car contained an IBM 486 personal computer and various auxiliary apparatus for measuring driving behavior and generating stimuli. In this study, speed, steering wheel position, and stimuli presented were registered at a sample rate of 10 Hz.

The readings for ECG, EOG, and SCR were digitally recorded at 100 Hz with Coda software (DATAQ Instruments, Akron, OH) by means of silver/silver chloride (Ag/AgCl) electrodes. We used electrodes placed on drivers' chests for the ECGs. The HR variability coefficient was computed by dividing the standard deviation of the IBIs by the average IBI (van Dellen, Aasman, Mulder, & Mulder, 1983). The EOG was measured with electrodes above and below the right eye and recorded with an AC-coupled amplifier ($\tau = 3$ s). The derivative of the EOG signal was used for the detection of eyeblinks. A *blink* was defined as a valley followed by a peak in the differentiated signal within less than 0.2 s and the amplitude of the peaks and valleys had to be at least 50% of the standard deviation of the entire signal. The SCR was measured with a Wheatstone bridge and amplified by an AC-coupled amplifier ($\tau = 10$ s). The electrodes were attached to the palm of the participant's left hand about 4 cm apart. Fluctuations in SCR were detected by Coda software and checked visually.

The CMT stimuli were uttered by a digitized female voice and were clearly audible under all driving conditions. L-counting stimuli were presented by LEDs mounted in a box at the left side of the steering wheel. The LEDs were protected from glare by a cardboard cover. The visual angle between the normal fixation point on the road ahead and the LEDs was about 27.0° horizontally and about 20.0° vertically. The distance between adjacent LEDs was about 0.9°.

Secondary task stimuli were presented on a plasma display mounted high on the center part of the dashboard (i.e., right of the steering wheel). The visual angle between the normal fixation point on the road ahead and the screen was also about

27.0° horizontally and 20.0° vertically. Again, a black cardboard shield was placed around this display. The height of the two digit stimuli on the screen was about 0.9° and the width was 0.7° per digit. The L-counting LEDs and the secondary task plasma display could not be seen at a single glance.

Participants

Twelve men participated in the present study. They were all between the ages of 23 and 50, had their licenses for more than 5 years, and had driven more than 10,000 km per year in the last 5 years.

Procedure

Participants were familiarized with the aim of the study, the instrumented car, the SWAT and RSME, the secondary task, and the two loading tasks. They were instructed to rate workload on the SWAT and RSME only with respect to the preceding period of loading task exposure. It was explicitly stated before and during the experiment that the secondary task (the visual detection task) was relatively unimportant, whereas the loading tasks (the L-counting task and the CMT) were extremely important and that target stimuli should not be missed. Furthermore, participants were repeatedly instructed to prevent the driving task from deteriorating when performing other tasks. Then, the SWAT card-sorting task was carried out, and electrodes for ECG, EOG, and SCR registration were attached.

First, physiological baseline conditions were registered for 15 min in the car while the participants were reading a newspaper. While driving toward the experimental route, the participants practiced the loading tasks and the secondary task for about 10 min. Order of the seven experimental conditions was counterbalanced across participants. For safety reasons, a licensed driver-training instructor who was able to take over with dual controls acted as experimenter.

Participants were warned by an auditory signal when L-counting or the CMT started. The 10-, 30-, and 60-s loading task blocks were presented in a random order. Response accuracy in the loading tasks was recorded at the end of each block by the experimenter after accuracy was verified by the technician. When a response was incorrect, the participant was informed immediately. To avoid

mishearing the vocal responses, participants wore a lightweight head-mounted microphone that was connected to a headphone used by the technician.

Data Analysis

Independent variables were loading task (no-loading task control, L-counting, and continuous memory task), duration (10, 30, and 60 s), and period (loading task vs. no-loading task). Figure 1 shows that each loading task block was surrounded by two 60-s rest periods. During these periods, participants merely drove the car. When appropriate, they continued with the secondary task. Physiological and driving measures were assessed in these 60-s blocks too, but, as subjective ratings were given by the participants at the start of these blocks, these measures were analyzed only for the last 30 s. The average workload assessed in the two rest periods surrounding each loading task block was used as local control for that block.

The SWAT results were analyzed by adding the ratings on the three scales and neglecting the results of the card-sorting procedure. Moreover, the relative weight of each dimension was derived from the card sorting and combined with the ratings at the three scales in three different ways. The weights of the dimensions were computed (a) for the entire group, (b) for groups of prototypical participants, and (c) for individual participants. The card-sorting results showed that the group solution (option a above) was allowed because the coefficient of concordance, as computed by the analysis package provided with the SWAT, amounted to .82. The recommended minimal coefficient of concordance for group solutions is .78. Separate analyses on the SWAT workload results confirmed this. For this reason, only the analysis that was based on the solution for the entire participant group are reported below. The RSME

was analyzed by subjecting the raw ratings to an analysis of variance (ANOVA).

Analysis of the SCR signal was based on the weighted average interval between SCRs during the periods of elevated workload. The last SCR before and the first SCR after this period were also included in this analysis, as they marked intervals that partly overlapped with the loading task periods.

Results

Performance on the Loading Tasks:

L-Counting and CMT

On average, 5.3% of the L-counting targets were missed. This percentage was higher when the secondary task was carried out, 8.4% versus 2.3%; $F(1, 11) = 13.2, p < .01$. The CMT performance was indexed by the sum of the absolute differences between the number of detected and the number of presented target letters. Across participants, the number of detected targets deviated 6.4% from the number of presented targets.

General Workload Indicators

Statistical significance of the effects of the loading tasks are indicated in Table 1 for each dependent variable. Comparisons are made between both loading task conditions and the control condition in the first F/p column. The second F/p column in Table 1 shows significance of the local differences by comparing the dependent variables obtained during actual execution of the loading tasks and those measured during the rest periods before and after them. The third F/p column gives rough indications of the sensitivity of the dependent variables to the duration of elevated workload. These results indicate that steering frequency, secondary task performance, SWAT, RSME, and, to a lesser degree, HR variability, and SCR were sensitive to the workload induced by the loading tasks. In addition, levels of workload indicated by secondary task performance, SWAT, RSME, and steering frequency were affected by workload duration.

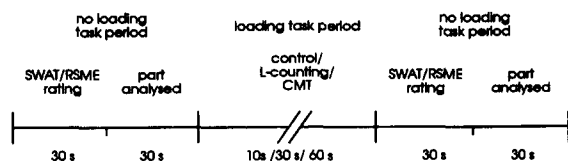


Figure 1. Schematic representation of rest and loading task periods while driving. CMT = continuous memory task; L = L-shaped light-emitting diodes; RSME = Rating Scale Mental Effort; SWAT = Subjective Workload Assessment Technique.

Driving Performance

Participants drove 84.6 kph on the average, but, as indicated in Table 1, speed was not affected by

Table 1
F Values and Levels of Statistical Significance With Respect to Global Loading Task Effects, Local Loading Task Effects, and Duration Effects for Each of Nine Dependent Variables

Task	Control vs. L-counting/CMT		Control vs. L-counting/CMT ×			
	<i>F</i> (1, 11)	<i>p</i>	Period		Duration ^a	
			<i>F</i> (1, 11)	<i>p</i>	<i>F</i> (1, 11)	<i>p</i>
Speed	0.2	> .10	1.2	> .10	0.2	> .10
Steering	18.5	< .001	6.9	< .05	1.0	> .10
Secondary task	38.9	< .001	33.1	< .001	6.8	< .05
SWAT	30.9	< .001			28.3	< .001
RSME	43.2	< .001			11.7	< .01
Interbeat interval	0.2	> .10	2.7	> .10	4.2	< .10
Heart rate variability	3.8	< .10	1.7	> .10	0.6	> .10
Eyeblinks ^b	0.6	> .10	0.3	> .10	7.0	< .05
Skin conductance response	4.1	< .10	8.6	< .05	0.2	> .10

Note. SWAT = Subjective Workload Assessment Technique; RSME = Rating Scale Mental Effort; CMT = continuous memory task.

^aThese planned comparisons involved 10 versus 60 s only. ^bControl versus CMT: L-counting (the number of L-shaped light-emitting diodes) excluded because eyeblinks and vertical eye movements could not be distinguished.

the loading tasks. The secondary (workload assessment) task did not affect speed, $F(1, 11) = 1.6$, $p > .20$.

Steering intervals were generally longer during the loading task period than in the no-loading task period (see Figure 2), $F(1, 11) = 51.4$, $p < .001$. However, Figure 2 suggests that steering intervals decreased more with L-counting than with the control task and the CMT. This was confirmed by pairwise planned comparisons; control task

versus L-counting: $F(1, 11) = 10.0$; CMT versus L-counting: $F(1, 11) = 6.1$, $ps < .05$. Steering intervals in the control condition and the CMT did not differ, $F(1, 11) = 1.0$, $p > .20$. In the loading task period, the average steering interval across all loading task conditions amounted to 3.6 s, which decreased to 3.0 s when the secondary task was also carried out, $F(1, 11) = 7.2$, $p < .05$. This indicates that the 3.3-s steering interval in the L-counting condition was not the lowest value possible. Hence, steering intervals were not subject to a floor effect and seem useful for indicating still higher levels of visual workload.

Workload Assessment Techniques

Secondary workload performance. Figure 3 and Table 1 show that secondary task performance dropped when a loading task was carried out, whereas performance remained unchanged in the control condition. Planned comparisons showed that this performance drop was significant for the L-counting and the CMT conditions separately, $F(1, 11) = 28.0$, $p < .001$, and $F(1, 11) = 12.3$, $p < .01$, respectively. Moreover, the difference between L-counting and the CMT was significant, $F(1, 11) = 25.9$, $p < .001$.

Figure 4 demonstrates that the percentage of detected targets decreased as L-counting and CMT

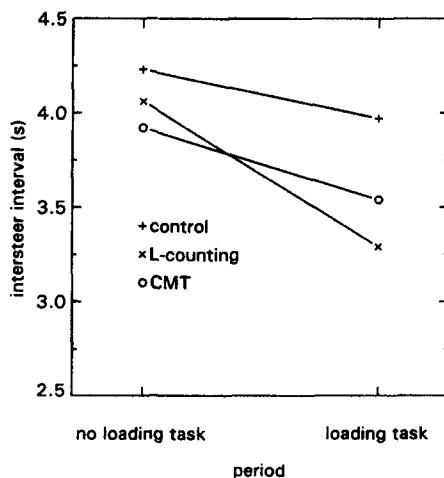


Figure 2. Intersteer interval as a function of loading task and period. CMT = continuous memory task; L = L-shaped light-emitting diodes.

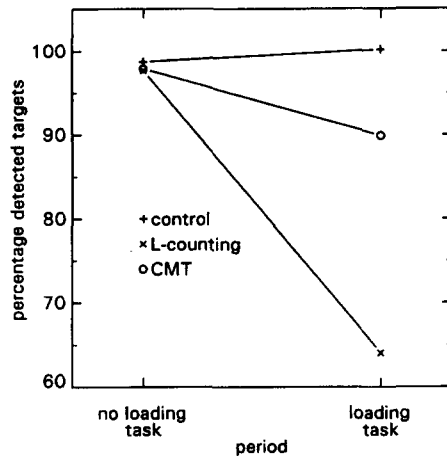


Figure 3. Secondary task performance as a function of period and loading task. CMT = continuous memory task; L = L-shaped light-emitting diodes.

lasted longer, but this was not the case in the control condition, $F(1, 11) = 6.8, p < .05$. Planned comparisons per duration confirmed this, as secondary task performance was lower at 60-s L-counting than at 10- and 30-s L-counting, $F(1, 11) = 6.6$ and $F(1, 11) = 8.9$, respectively, $ps < .05$, and secondary task performance was higher at 10-s CMT than at 30- and 60-s CMT, $F(1, 11) = 7.3, p < .05$, and $F(1, 11) = 12.4$, respectively, $p < .01$. It appears that timesharing a loading task and the secondary task became more difficult, as this situation took longer.

Subjective workload. As shown in Table 2, SWAT and RSME were both affected by loading

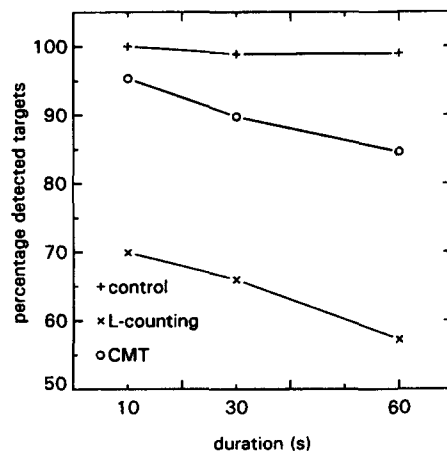


Figure 4. Effects of duration and loading task on secondary task performance. CMT = continuous memory task; L = L-shaped light-emitting diodes.

task and duration. Furthermore, Loading Task $\times$ Duration interactions were found for both subjective measures, $Fs(4, 44) > 4.0, ps < .01$, suggesting that the duration effect occurred with L-counting and CMT, but not in the control condition (see Figure 5). Pairwise comparisons of the three loading task conditions showed that SWAT and RSME were both sensitive to differences between control condition and L-counting and between control condition and CMT (see Table 2).

Table 2 also shows that both subjective workload measures produced virtually the same results with respect to duration. However, there was a difference in that SWAT differentiated between 10- and 30-s L-counting, which difference was not significant for RSME.

IBIs and HR variability. As indicated in Table 1, IBI and HR variability were not very sensitive to workload. Planned comparisons showed a marginally significant tendency of IBI to reduce when L-counting and CMT took longer, whereas this was not the case in the control condition, $F(1, 11) = 4.2, p = .06$ (see Figure 6). With 10-s L-counting, IBIs were longer than with 60-s L-counting, $F(1, 11) = 6.6, p < .05$. During the 60-s CMT

Table 2

Statistical Significance of the Differences Between Conditions as Indicated by the Subjective Rating Scales SWAT and RSME

Condition	SWAT <i>p</i>	RSME <i>p</i>
Loading task (main)	< .001	< .001
No-loading task vs. L-counting	< .001	< .001
L-counting vs. CMT	< .10	> .10
No-loading task vs. continuous memory task	< .001	< .001
Duration (main)	< .001	< .05
No-loading only ^a	> .10	> .10
L-counting only (s)	< .001	< .05
10 vs. 30 ^b	< .001	> .10
10 vs. 60	< .001	< .01
30 vs. 60	> .10	> .10
Continuous memory task only (s)	< .001	< .05
10 vs. 30	< .05	< .10
10 vs. 60	< .001	< .05
30 vs. 60	> .10	> .10

Note. SWAT = Subjective Workload Assessment Technique; RSME = Rating Scale Mental Effort; L-counting = number of L-shaped light-emitting diodes.

^aDuration effects per loading task tested with 10 versus 60-s planned comparisons. ^bDifferences between durations tested with Tukey post hoc tests.

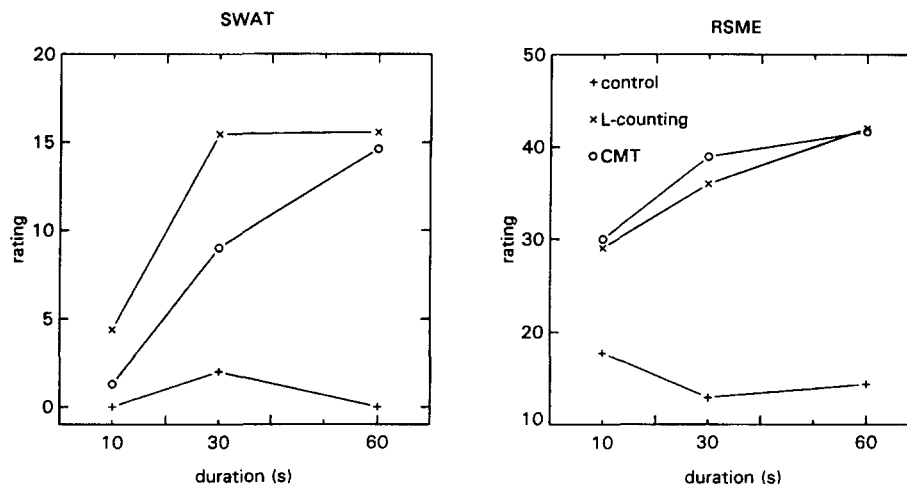


Figure 5. Subjective ratings as a function of loading task and duration. L = L-shaped light-emitting diodes; RSME = Rating Scale Mental Effort; SWAT = Subjective Workload Assessment Technique.

periods, IBIs were shorter than with 10- and 30-s CMT, $F(1, 11) = 19.8$, and $F(1, 11) = 19.0$, $ps < .001$, respectively

HR variability was lower in 60-s L-counting periods than in 60-s control condition periods; 0.049 versus 0.059, $F(1, 11) = 8.1$, $p < .05$. Likewise, HR variability was lower when the CMT was performed for 60 s as compared with 60-s control condition periods; 0.050 versus 0.059, $F(1, 11) = 5.6$, $p < .05$. Finally, HR variability was lower in the no-loading task periods of driving than while standing still; 0.057 versus 0.084; $F(2, 22) = 9.3$, $p < .05$.

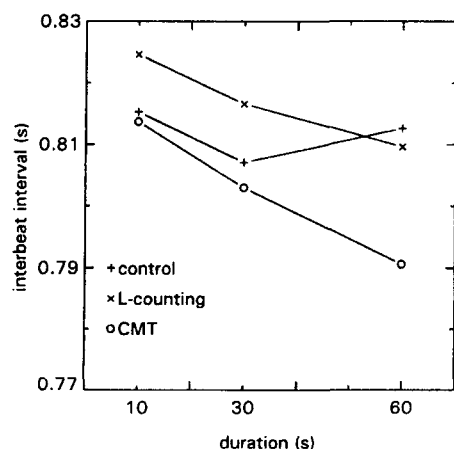


Figure 6. Effects of loading task and duration on interbeat intervals of the heart beat. CMT = continuous memory task; L = L-shaped light-emitting diodes.

Eyeblinks. L-counting was not included in the eyeblink analysis because eyeblinks could not be distinguished from the rapid vertical eye movements made toward the L-counting display. The average eyeblink intervals in the CMT and the control condition were not different; 6.7 versus 7.0 s; $F(1, 11) = 0.2$, $p > .20$. However, loading task period duration had different effects for the CMT and the no-loading task: in the no-loading task condition eyeblink intervals decreased with duration from 7.6 (10 s) to 7.3 (30 s) to 7.0 s (60 s), whereas in the CMT conditions, eyeblink intervals increased from 6.4 s to 7.0 to 8.2 s, $F(1, 11) = 7.0$, $p < .05$.

SCRs. Table 1 suggests that SCR responded to the loading tasks. Planned comparisons showed that the intervals between successive SCRs were shorter in the CMT condition than in the control and L-counting conditions; CMT versus control: $F(1, 11) = 18.9$, $p < .001$; CMT versus L-counting, $F(1, 11) = 8.8$, $p < .05$ (see Figure 7). Intervals between SCRs in the L-counting and control conditions were not different, $F(1, 11) = 2.4$, $p > .10$. There was a marginally significant difference between SCR intervals while standing still and when driving in the no-loading task period; 24.2 s versus 16.2 s; $F(2, 22) = 3.1$, $p = .06$.

Comparing Workload Measures

Correlations and factor analyses. Table 3 shows correlations between the various workload mea-

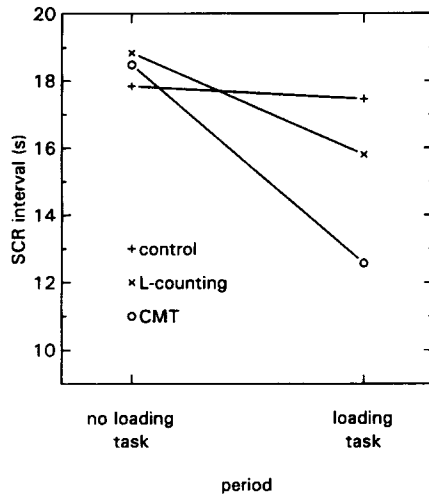


Figure 7. Effects of loading task and period on skin conductance responses (SCRs). CMT = continuous memory task; L = L-shaped light-emitting diodes.

asures in the control and the L-counting condition and between the workload measures in the control and the CMT condition. The individual SWAT dimensions time, effort, and stress have been

included in order to assess their diagnosticity. Table 3 confirms that steering intervals, secondary task performance, SWAT, and RSME responded in a way comparable with L-counting. A factor analysis indicated that two independent factors underlie L-counting workload. The first factor involved SWAT, the individual Time dimension of the SWAT, RSME, and Secondary Task Performance (explained variance 36%; factor loadings are $-.96$, $-.85$, $-.81$, and $.76$, respectively). The second factor depended mostly on SCR, Speed, and HR variability (explained variance 14%; factor loadings are $.66$, $-.60$, and $-.58$, respectively).

With respect to the CMT, SWAT, RSME, Secondary Task Performance and Steering Intervals showed significant mutual correlations. Again, the factor analysis revealed two underlying factors. The first relied mainly on SWAT, Time, RSME, and Effort (explained variance 31%; factor loadings $-.94$, $-.81$, $-.79$, and $-.73$, respectively). The second factor involved SCR, Eyeblink Intervals, and HR variability (explained variance 14%; loadings $.68$, $.67$, and $-.64$, respectively).

Sensitivity analyses. The sensitivity of the SWAT, RSME, secondary task performance, steer-

Table 3
Correlations Between the Various Measures for the Control and L-Counting Conditions and the Control and Continuous Memory Task (CMT) Conditions

Measure	1	2	3	4	5	6	7	8	9	10	11
1. SWAT	—	.88	.61	.70	.69	-.37	-.02	-.60	-.20	.07	-.08
2. Time	.77	—	.36	.49	.63	-.37	.05	-.55	-.16	.04	-.10
3. Effort	.85	.59	—	.05	.50	-.18	-.19	-.43	-.04	.18	.04
4. Stress	.66	.17	.34	—	.37	-.23	.07	-.33	-.23	-.05	-.10
5. RSME	.66	.53	.66	.30	—	-.28	.04	-.62	-.21	.04	-.15
6. Steer	-.26	-.35	-.21	-.04	-.23	—	-.13	.33	.03	-.13	.06
7. SCR	-.05	-.07	-.03	-.01	-.14	-.19	—	.07	-.07	.14	-.25
8. Secondary task	-.44	-.20	-.53	-.28	-.44	.10	.08	—	.03	-.03	.36
9. Speed	-.03	-.03	-.07	.04	-.02	-.05	.03	.16	—	-.36	.15
10. Interbeat interval	-.04	-.02	.13	-.20	-.07	-.22	.20	-.04	-.43	—	-.10
11. Heart rate variability ^a	.02	.12	-.02	-.06	-.09	.07	-.29	.36	.15	-.07	—
12. Eye blink rate ^a	.19	-.11	.12	.43	.10	.18	.34	-.18	-.08	-.09	-.30

Note. Each correlation is based on the average workload index per participant in the control condition and in the CMT or L-counting (number of L-shaped light-emitting diodes) loading task condition for each duration (i.e., $N = 12 \times 2 \times 3 = 72$ observations); significant correlations are boldface; blink rate is not indicated for L-counting because rapid vertical eye movements associated with scanning the L-counting display could not be distinguished from eyeblinks. L-counting conditions are in the upper right and CMT conditions are in the lower left. SWAT = Subjective Workload Assessment Technique; RSME = Rating Scale Mental Effort; SCR = skin conductance response.

^aNo comparable data in the L-counting condition.

ing intervals, and SCR intervals were compared in a Technique $\times$ Loading Task $\times$ Duration ANOVA after all data had been z-transformed. Figure 8 depicts the z scores for these promising workload measures. Table 4 indicates statistical significance of the differences in sensitivity between techniques. The secondary task was more sensitive to 10-s L-counting periods than any other task. Overall, SCR was a worse indicator for the workload of 60-s L-counting periods than SWAT, RSME, and the secondary task. Workload induced by executing the CMT for 10 s was indicated poorly by all measures. CMT performance of 30 and 60 s was indicated better by RSME than by steering intervals, SCR, and secondary task performance.

There are indications that workload might also have an effect on periods after workload elevations have ceased (Hancock, Williams, Manning, & Miyake, 1995; see also our Figure 1). This hypothesis was tested by comparing measures obtained in the no-loading task periods of the control condition, L-counting and CMT. This analysis was carried out for speed, steering intervals, secondary task performance, IBI, HR variability, and SCR intervals. It yielded a marginally significant effect for steering intervals only, $F(1, 11) = 4.4$, $p = .06$ (see Figure 2). Hence, the present study provides little support for the notion that relatively short workload periods have lasting effects.

Finally, Table 5 was produced to depict which measures are useful for finding differences at each workload duration. This table shows that secondary task performance, SWAT, and RSME were sensitive to the workload imposed by both the L-counting task and the CMT at all three durations

with the exception that the SWAT could not distinguish 10-s CMT from the control condition. Table 5 also shows that HR variability differentiated significantly between driving and standing still. SCR showed a marginally significant difference. The IBI did not indicate a difference between driving and remaining still.

Discussion

The present study investigated the sensitivity and the diagnosticity of nine common workload assessment techniques used to evaluate shorter and longer periods of elevated visual and mental workload. *Sensitivity* refers to the extent that a technique discriminates between different levels of workload. *Diagnosticity* indicates whether a technique responds differently to different types of workload.

The results were clearcut. Secondary task performance, SWAT, and RSME were sensitive to the visual 10-s workload peaks caused by L-counting. Secondary task performance, RSME, and, somewhat less reliably, SCR were sensitive to the 10-s mental workload peaks induced by the CMT. In addition, steering intervals and HR variability were sensitive to longer (30 and 60 s) periods of L-counting, whereas SWAT and HR variability were sensitive to longer expositions to the CMT.

The secondary task and steering intervals were more sensitive to L-counting than to the CMT; therefore, these techniques are diagnostic with respect to the type of workload. However, the data also showed that the performance reduction of the visually loading secondary task does not necessar-

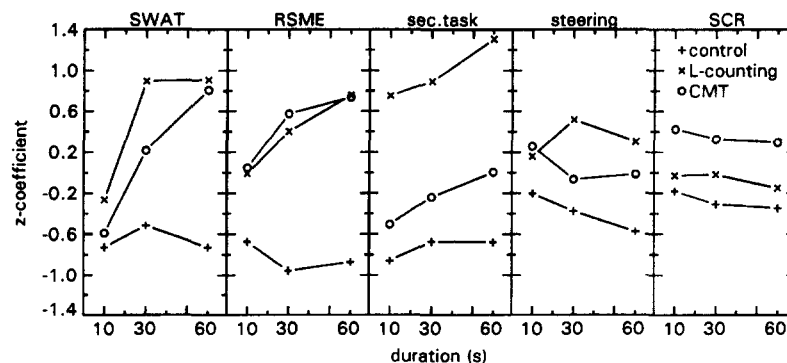


Figure 8. Z-transformed results of the five most sensitive indicators for workload as a function of loading task and duration. SWAT = Subjective Workload Assessment Technique; RSME = Rating Scale Mental Effort; sec. = secondary; SCR = skin conductance response.

Table 4

Comparisons of the Sensitivity of Pairs of Measures as a Function of Loading Task and Task Duration

Measure	L-counting ^a <i>p</i>			CMT <i>p</i>		
	10 s	30 s	60 s	10 s	30 s	60 s
SWAT vs. RSME	> .10	> .10	> .10	< .05	< .01	> .10
SWAT vs. steering	> .10	> .10	< .10	> .10	> .10	< .01
SWAT vs. SCR	> .10	< .01	< .01	< .10	> .10	> .10
SWAT vs. secondary task	< .01	> .10	> .10	< .05	> .10	< .05
RSME vs. steering	> .10	> .10	> .10	> .10	< .01	< .01
RSME vs. SCR	> .10	< .01	< .001	> .10	< .05	< .05
RSEM vs. secondary task	< .05	> .10	> .10	> .10	< .01	< .01
Steering vs. SCR	> .10	< .05	> .10	> .10	> .10	> .10
Steering vs. secondary task	< .01	> .10	< .05	> .10	> .10	> .10
SCR vs. secondary task	< .001	< .01	< .001	> .10	> .10	> .10

Note. Levels of significance are as indicated by Technique $\times$ Loading Task (L-counting or CMT vs. control) interactions in separate analyses of variance. CMT = continuous memory task; SWAT = Subjective Workload Assessment Technique; RSME = Rating Scale Mental Effect; SCR = skin conductance response.

^aNumber of times light-emitting diodes formed an L shape.

ily point to increased levels of visual workload alone because secondary task performance was also affected by the CMT. SCR responded only to mental workload, but its sensitivity was limited, given the fact that the statistical significance of these effects was "at the edge." In line with earlier results (O'Donnell & Eggemeier, 1986), the subjective techniques SWAT and RSME appeared not to be diagnostic with respect to visual and mental workload in that they were about equally sensitive

to L-counting and CMT. The observation that some measures were more sensitive to visual or mental workload underlines that workload is a multidimensional concept (Wickens, 1984b) and that, dependent on the question at hand, the diagnosticity of a workload assessment technique should be taken into account when designing a workload study.

We found that HR variability distinguished between driving with and without a 60-s loading

Table 5

Sensitivity of Each Workload Assessment Technique in Terms of Statistical Significance of the Differences Between the Loading Task and No-Loading Task Control Conditions for Separate Loading Tasks and Durations

Duration	L-counting <i>p</i> ^a			CMT <i>p</i>			Remaining still vs. driving only <i>p</i>
	10 s	30 s	60 s	10 s	30 s	60 s	
Speed	> .10	> .10	> .10	> .10	> .10	> .10	
Steering intervals	> .10	< .001	< .01	< .10	> .10	< .05	
Secondary task	< .001	< .001	< .001	< .01	< .01	< .001	
SWAT	< .05	< .01	< .001	< .10	< .01	< .001	
RSME	< .01	< .001	< .001	< .05	< .001	< .001	
Interbeat interval	> .10	> .10	> .10	> .10	> .10	< .10	> .10
Heart rate variability	> .10	> .10	< .05	> .10	> .10	< .05	< .01
Eyeblinks				> .10	> .10	> .10	
Skin conductance response intervals	> .10	> .10	> .10	< .05	< .10	< .05	< .10

Note. SWAT = Subjective Workload Assessment Technique; RSME = Rating Scale Mental Effort; CMT = continuous memory task.

^aNumber of times light-emitting diodes formed an L shape.

task, but it appeared not to be sensitive to 10- and 30-s periods of elevated workload. Moreover, HR variability discriminated between driving without a loading task and remaining still. This suggests that HR variability is sensitive to longer periods of elevated workload, but that this measure is not suitable for detecting workload elevations of less than 1 min.

As noted in the Method section, the SWAT was analyzed in four different ways. These analyses yielded virtually identical results; hence, the results of only one type of analysis were reported in subsequent analyses. Still, it is important to realize that even the simplest analysis, adding the three ratings, did about as well as the analyses that included individual differences, as obtained by the card-sorting procedure. This confirms the results of researchers who worked with homogeneous participant groups (see Moroney, Biers, & Eggemeier, 1995, for an overview), for the heterogeneous group of drivers in the present study. Moreover, the present data showed that the RSME could reliably detect the workload associated with 10-s CMT periods. This was not the case with the SWAT. In other words, it seems that in driving studies, subjective workload assessment techniques gain little from complex assessment and analysis methods such as those used in the SWAT. A simple one-dimensional rating scale appears sufficient for assessing subjective workload in driving. Future research should investigate whether there are specific task environments where complex scales like the SWAT and the National Aeronautics and Space Administration Task-Load Index (NASA-TLX; Hart & Staveland, 1988) are beneficial.

The highly limited sensitivity of eyeblinks to the CMT, together with findings in other studies that eyeblinks were sensitive to visual workload (e.g., Veltman & Gaillard, 1996; Wilson & Fisher, 1990), corresponds with earlier claims that eyeblinks are diagnostic with respect to the type of workload (Casali & Wierwille, 1984; Wierwille & Eggemeier, 1993). The EOG could not distinguish between rapid vertical eye movements toward the in-vehicle display and eyeblinks. Therefore, if drivers repeatedly glance at in-car displays, applied researchers might be better off with observation techniques to determine eyeblink frequency than with EOG.

An important theoretical result is that the factor analysis revealed two independent factors to under-

lie visual and mental workload. The first factor involved subjective workload ratings (SWAT, RSME, and the time pressure dimension of the SWAT) and, in case of L-counting, secondary task performance. The second factor included the physiological measures SCR, HR variability, and, in the CMT condition, eyeblink rate. This supports the suggestion in the introduction that subjective ratings and secondary task performance index information-processing load, or task complexity, and physiological measures indicate the effort invested in the task or, perhaps also, the stress caused by attending to a task unrelated to driving (cf. Casali & Wierwille, 1984; Hart & Wickens, 1990).

To help readers determine the most appropriate workload assessment technique for studies on visual and mental workload, we present in Table 6 a simplified and pragmatic overview of the presently used workload assessment techniques in terms of their sensitivity and diagnosticity.

It is important to note that the conclusions with respect to the secondary task should be considered with caution because intrusion of driving by the secondary task was measured only with respect to speed and steering frequency. Although speed did not show an effect of secondary task performance, steering frequency did increase. The data do not indicate the extent to which other parts of the driving task, such as braking reaction time and lane keeping, were affected. Various researchers (e.g., Noy, 1987) have argued that secondary tasks are likely to affect primary task performance. This stresses that participants should always and repeatedly be instructed to stop performing the secondary task whenever driving performance tends to deteriorate (Wickens, 1984a). The use of rewards and immediate performance feedback may be of value here. Wierwille and Gutmann (1978) found that primary task performance was affected more by a secondary task, as the primary task was less demanding. Therefore, when the secondary task technique is adopted, task priorities should be emphasized especially with a less demanding primary task in order to prevent underestimations in the lower workload regions.

A second limitation of the present study is that the loading tasks involved only one level of difficulty (i.e., absence vs. presence). This makes it hard to estimate the sensitivity of the present

Table 6
*Sensitivity and Diagnosticity of Visual and Mental Workload Measures Used
 in This Study and the Durations for Which These Qualifications Hold*

Measure	Sensitivity		Diagnosticity
	Visual	Mental	
Steering frequency	High ≥ 30 s	Poor	High ≥ 30 s
Secondary task ^a	High ≥ 10 s	High ≥ 10 s	Reasonable ≥ 10 s
SWAT	High ≥ 10 s	High ≥ 30 s	None
RSME	High ≥ 10 s	High ≥ 10 s	None
Skin conductance response	None	Reasonable ≥ 10 s	Reasonable ≥ 10 s
Heart rate variability	Reasonable ≥ 60 s	Reasonable ≥ 60 s	None
Interbeat interval	None	None	None
Speed	None	None	None
Eyeblinks	Yes ^b	None	Reasonable/High ^b

Note. SWAT = Subjective Workload Assessment Technique; RSME = Rating Scale Mental Effort.

^aQualifications depend on type of secondary task. ^bExpected on basis of other studies (e.g., Veltman & Gaillard, 1996; Wilson & Fisher, 1990).

workload assessment techniques to different levels of workload. In the extreme, the present techniques may have responded to the presence of the loading tasks per se, whereas the actual difficulty level was unimportant. Future research should assess the sensitivity of the techniques in this study to more than the two levels of workload reported in this article.

An important characteristic of the present study design was that the moments of workload elevation were known to the experimenter. However, workload assessment techniques will be less sensitive when the moments of elevated workload are not recognized and measures are assessed over much longer periods of time. Then, the occurrence of workload peaks may be entirely averaged out (Knowles, 1963; Wierwille, 1981). Determining whether workload peaks occur calls for application of these three methods: (a) executing a task analysis prior to the experiment and assuring that workload is captured at the appropriate moments; (b) in case of continuous techniques (secondary task performance and physiological measures) computing moving averages (Antin & Wierwille, 1984; Wierwille, 1981), and (c) asking participants to indicate workload peaks and their levels (Wierwille & Eggemeier, 1993).

With respect to use of subjective rating scales, it is important to note that these techniques usually do not instruct participants whether they should estimate peaks or averages across longer periods of time. The consequence of this omission is an

increase of individual differences because different participants will understand the rating task differently. Hence, the researcher should always indicate explicitly whether workload ratings should regard peaks or averages and to which period of time or to which situation the estimate should refer.

Another theoretical issue concerns the relationship between workload and performance. Tasks may differ considerably with respect to the probability that performance is affected by peak workload. For example, driving always involves some time pressure, and the occurrence of peak workload is probably more important for task performance than for tasks without time pressure such as in human-computer interactions. This suggests that researchers should adopt a workload assessment technique that is more sensitive to peak workload when task performance is more likely to be affected by peak workload. For driving research, assessment of peak workload seems crucial. Secondary task performance and subjective ratings appear to be appropriate techniques in this respect because they are sensitive to workload peaks.

In conclusion, the present study demonstrates that subjective ratings and secondary task techniques are good methods for assessing visual and mental workload peaks in driving. The relatively simple univariate RSME rating scale was somewhat more sensitive than the more complex SWAT. Steering frequency appears to be a sensitive measure for longer periods of visual workload, and

SCR appears to be sensitive to mental workload. The present results suggest that physiological measures such as SCR, HR variability, and eyeblink rate reflect invested effort rather than information-processing load and task performance. Because driving is a task in which peak workload may have serious ramifications for safety, studies on workload in driving should adopt workload assessment techniques that can evaluate the occurrence of peak workload. The present results show that secondary task performance and subjective ratings are useful in this respect. In task environments with less time pressure, this requirement may be less important.

References

- Antin, J. F., Dingus, T. A., Hulse, M. C., & Wierwille, W. W. (1990). An evaluation of the effectiveness and efficiency of an automobile moving-map navigational display. *International Journal of Man-Machine Studies*, 33, 581-594.
- Antin, J. F., & Wierwille, W. W. (1984). Instantaneous measures of mental workload: An initial investigation (pp. 6-10). *Proceedings of the 28th Annual Meeting of the Human Factors Society*. Santa Monica, CA: Human Factors Society.
- Boer, L. C., & Jorna, P. G. A. M. (1987). *Taskomat: Een batterij van informatieverwerkingstaken*. [Taskomat: A battery of information-processing task.] (Technical Report 1987 A2). Soesterberg, The Netherlands: Institute for Perception TNO.
- Brookhuis, K. A., de Vries, G., & de Waard, D. (1991). The effects of mobile telephoning on driving performance. *Accident Analysis & Prevention*, 23, 309-316.
- Casali, J. G., & Wierwille, W. W. (1984). On the measurement of pilot perceptual workload: A comparison of assessment techniques addressing sensitivity and intrusion issues. *Ergonomics*, 27, 1033-1050.
- Edelberg, R. (1972). Electrical activity of the skin: Its measurements and uses in psychophysiology. In N. S. Greenfield & R. A. Sternbach (Eds.), *Handbook of psychophysiology* (pp. 367-418). New York: Holt, Rinehart & Winston.
- Gaillard, A. W. K., & Wientjes, C. J. E. (1994). Mental load and work stress as two types of energy mobilization. *Work & Stress*, 8, 141-152.
- Hancock, P. A., Williams, G., Manning, C. M., & Miyake, S. (1995). Influence of task demand characteristics on workload and performance. *International Journal of Aviation Psychology*, 5, 63-86.
- Harms, L. (1986). Drivers' attentional responses to environmental variations: A dual-task real traffic study. In A. G. Gale, M. H. Freeman, C. M. Haslegrave, P. Smith, & S. P. Taylor (Eds.), *Vision in vehicles* (pp. 131-138). Amsterdam: North Holland.
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139-183). Amsterdam: North Holland.
- Hart, S. G., & Wickens, C. D. (1990). Workload assessment and prediction. In H. R. Boohar (Ed.), *Man-print: An approach to systems integration* (pp. 257-296). New York: Van Nostrand Reinhold.
- Hicks, T. G., & Wierwille, W. W. (1979). Comparison of five mental workload assessment procedures in a moving-base driving simulator. *Human Factors*, 21, 129-143.
- Hockey, G. R. J. (1986). A state control theory of adaptation to stress and individual differences in stress management. In G. R. J. Hockey, A. W. K. Gaillard, & M. G. H. Coles (Eds.), *Energetics and human information processing* (pp. 285-298). Dordrecht, The Netherlands: Nijhoff.
- Knowles, W. B. (1963). Operator loading tasks. *Human Factors*, 5, 155-161.
- MacDonald, W. A., & Hoffman, E. R. (1980). Review of relationships between steering wheel reversal rate and driving task demand. *Human Factors*, 22, 733-739.
- Michon, J. A. (Ed.). (1993). *Generic intelligent driver support*. A comprehensive report on GIDS. London: Taylor & Francis.
- Moroney, W. F., Biers, D. W., & Eggemeier, F. T. (1995). Some measurement and methodological considerations in the application of subjective workload measurement techniques. *International Journal of Aviation Psychology*, 5, 87-106.
- Noy, Y. I. (1987). Theoretical review of the secondary task methodology for evaluating intelligent automobile displays. *Proceedings of the 31st Annual Meeting of the Human Factors Society* (pp. 205-209). Santa Monica, CA: Human Factors Society.
- O'Donnell, C. R., & Eggemeier, F. T. (1986). Workload assessment methodology. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance: Vol. II. Cognitive processes and performance* (pp. 42.1-42.29). New York: Wiley.
- Reid, G. B., & Nygren, T. E. (1988). The Subjective Workload Assessment Technique: A scaling procedure for measuring mental workload. In P. A. Hancock, L. Freeman, C. M. Haslegrave, P. Smith, & S. P. Taylor (Eds.), *Vision in vehicles* (pp. 131-138). Amsterdam: North Holland.

- cock & N. Meshkati (Eds.), *Human mental workload* (pp. 185–218). Amsterdam: North Holland.
- Stern, J. A., Boyer, D., & Schroeder, D. (1994). Blink rate: A possible measure of "fatigue." *Human Factors*, 36, 285–297.
- Taylor, D. H. (1964). Drivers' galvanic skin response and the risk of accident. *Ergonomics*, 7, 439–451.
- van Dellen, H. J., Aasman, J., Mulder, L. J. M., & Mulder, G. (1983). Time domain versus frequency domain measures of heart rate. In J. F. Orlebeke, G. Mulder, & L. J. P. van Doornen (Eds.), *Psychophysiology of cardiovascular control: Models, methods and data* (pp. 353–374). New York: Plenum Press.
- Veltman, J. A., & Gaillard, A. W. K. (1996). Physiological indices of workload in a simulated flight task. *Biological Psychology*, 42, 323–342.
- Verwey, W. B. (1991). *Towards guidelines for in-car information management: Driver workload in specific driving situations*. (Technical Report IZF 1991 C-13). Soesterberg, The Netherlands: TNO Institute for Perception.
- Verwey, W. B. (1993). How can we prevent overload of the driver? In A. M. Parkes & S. Franzen (Eds.), *Driving future vehicles* (pp. 235–244). London: Taylor & Francis.
- Verwey, W. B., & Janssen, W. H. (1989). Driving behavior with electronic in-car navigation aids. *Proceedings of the Road Safety in Europe Conference* (pp. 81–97). Gothenburg, Sweden: VTI.
- Vivoli, G., Bergomi, M., Rovesti, S., Carrozzi, G., & Vezzosi, A. (1993). Biochemical and haemodynamic indicators of stress in truck drivers. *Ergonomics*, 36, 1089–1097.
- Wickens, C. D. (1984a). *Engineering psychology and human performance*. Columbus, OH: Merrill.
- Wickens, C. D. (1984b). Processing resources in attention. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 63–102). London: Academic Press.
- Wierwille, W. W. (1979). Physiological measures of aircrew mental workload. *Human Factors*, 21, 575–594.
- Wierwille, W. W. (1981). Instantaneous mental workload: Concept and potential methods for measurement. *Proceedings of the International Conference on Cybernetics and Society* (pp. 604–608). New York: Institute of Electronics and Electronic Engineers.
- Wierwille, W. W., & Eggemeier, F. T. (1993). Recommendations for mental workload measurement in a test and evaluation environment. *Human Factors*, 35, 263–281.
- Wierwille, W. W., & Gutmann, J. C. (1978). Comparison of primary and secondary task measures as a function of simulated vehicle dynamics and driving conditions. *Human Factors*, 20, 233–244.
- Wilson, G. F., & Fisher, F. (1990). Classification of flight segment using pilot and WSO physiological data. *Proceedings of the 34th Annual Meeting of the Human Factors Society* (pp. 109–111). Santa Monica, CA: Human Factors Society.
- Zeier, H. (1979). Concurrent physiological activity of driver and passenger when driving with and without automatic transmission in heavy city traffic. *Ergonomics*, 22, 799–810.
- Zijlstra, F. R. H. (1993). *Efficiency in work behavior*. Unpublished doctoral dissertation, Technical University, Delft, The Netherlands.

Received May 3, 1995

Revision received February 26, 1996

Accepted March 1, 1996 ■