

## University of Groningen

### The measurement of drivers' mental workload

de Waard, Dick

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*

Publisher's PDF, also known as Version of record

*Publication date:*

1996

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

de Waard, D. (1996). *The measurement of drivers' mental workload*. [Thesis fully internal (DIV), University of Groningen]. s.n.

**Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

**Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

Workload assessment is of interest to many applied settings ranging from VDT (visual display terminal) data-typing to space travel. Each area has its own specific problems. Much research was and still is performed in the area of aviation, in which, in particular, military flight studies have been carried out. In these studies pilots have to perform very complex tasks under extreme conditions and the selection of pilots is so stringent that in general only healthy young men and women are capable of realising these tasks. This selected group of people also serves as subject in aviation workload studies, where extreme G-forces and heart rate values of up to 160 bpm are no exception (see, e.g., Roscoe, 1993). These types of environments have a clear influence on physiological measures (e.g., on HRV quality, see Jorna, 1993).

No such forces are encountered in ground travel, though it could be argued that the human perceptual information system is unfit for the high speeds of travel possible in modern cars on motorways. Actual practice shows that this is not true and that many people are able to perform this task daily without negative consequences. Sometimes, however, our information processing system reaches its limit and things go wrong. In most of those cases a human error has occurred (Smiley & Brookhuis, 1987), resulting in a traffic-law violation that led to an accident (Rothengatter, 1991). In that case driving speed may have turned out to be too high to deal with safely. The selection criteria for driving are also far less strict than those applied in (military) flight, and, as a result, the population behind the wheel is far more diverse in capabilities. These factors, among others, make workload research in traffic an area with its own specific problems. Results booked in this area of research may benefit a large group of people.

Most workload measures used in traffic research have been developed, and tested, in the laboratory and in other applied settings such as the workplace (see e.g., Meijman & O'Hanlon, 1984) and aviation. The exceptions to these are the primary-task measures, for driving; the vehicle parameters.

A useful experimental design in traffic research is to compare task performance in an experimental (e.g., mental load) condition with performance under baseline conditions. A difference in performance can then be attributed to the experimental manipulation. Recently, however, Brookhuis (1995a, 1995b) has proposed critical levels of performance for different primary-task measures. These critical values can be considered performance margins as discussed in chapter 4. The criteria are not workload redlines, since they indicate the point at which performance should be considered to be affected, and thus indicate a shift from the A to the B region. Most of the measures' critical levels

have been linked to unsafe behaviour, e.g., a level at which the likelihood that the vehicle leaves the traffic lane increases to a major extent (see Brookhuis 1995ab). In the following evaluations the absolute criteria will be included.

Evaluation of workload measures on their characteristics in traffic research will mainly be restricted to work that my colleagues and myself have performed. Self-report measures, primary-task and secondary-task measures, and physiological parameters have been used in these field studies. From these studies, specific road sections or conditions were selected with increased task demands. The studies will be divided into two categories, studies that include an increase in complexity and studies in which driver state is affected. The first category can be further divided into two sub-categories, an increase in road complexity versus an increase in task complexity, i.e. the addition of a secondary task. Differential sensitivity of a selection of measures to mental load in relation to demand are of primary interest in the evaluations.

### **Selected sections or conditions**

From one simulator and six field studies, experimental and baseline conditions or road segments (sections) were selected and the sensitivity of workload measures were compared between conditions and load categories. Sections were selected based upon expected effect of stressors or environment on workload, i.e. a selection based upon task demand. The following baseline and load conditions were selected:

#### *Complexity studies - environment:*

- (1) From the '**weaving section study**' (appendix 1), a study performed on the A28-motorway, driving over the combined entrance/exit road-section was selected as experimental condition and compared with a baseline control section. In appendix 1 the load condition is referred to as 'ACC 2' (section ZL in the Dutch report). The baseline control section was a road segment with no entrance or exit and is indicated in the appendix as section 'CTR1' (C1 in the Dutch report). All subjects drove these sections two times, once without eye-movement registration equipment, once with the equipment mounted on their heads (indicated as 'c' for CEMRE, Continuous Eye Movement Registration Equipment).
- (2) In the '**noise barrier study**' the same eye-movement equipment was used in one condition. Driving over a road section near a noise barrier was used as experimental condition and compared with driving along the same road section in the opposite direction, where no such barrier was present. The mid-part of the noise barrier was far closer to the motorway than the begin and end part. Driving on the motorway along the barrier created the impression that the screen 'approached' the car.

- (3) In the '**road layout study**' (appendix 2) the effect of a changed road design of an 'A'-class road on driving behaviour, in particular on speed choice, and on mental load was tested. The baseline consisted of driving on an ordinary A-road section that either preceded or followed an experimental section. All roads had a speed limit of 80 km/h and were single carriageways with two lanes, separated by a white line. The experiment included roads in two environments: a road leading through a forest (Wr, woodland road) and a road leading through open moorland (Mr, moorland road).

*Complexity studies - additional task:*

- (4) The '**car-phone study**' (appendix 3) was carried out both on a quiet motorway and on a busy four-lane ringroad. In the experimental conditions the drivers had to perform a difficult memory task, the PASAT, Paced Serial Addition Task (Gronwall & Sampson, 1974), while operating either a hand-held or a hands-free telephone set. In the experiment subjects drove and handled the car-phone five days a week for a total of three weeks. For the present comparisons, only data collected during the first week in which driver workload is likely to have been highest, were used.
- (5) The '**tutoring**' or DETER (Detection, Enforcement and Tutoring for Error Reduction) study (appendix 4) is the only simulator study included in the comparisons. In this study, drivers had to complete four trials in a driving simulator where they drove through built-up areas, on A roads and on dual carriageways. The middle two trials, where an enforcement and tutoring system provided the subjects with feedback about detected violations, were compared with the first and last trial, when no feedback about violations was given. The tutoring messages and the required behavioural adaptation were suspected of increasing mental load.

If baseline performance in the above-mentioned studies is assumed to be in region A2, performance in the load condition with increased demand can be expected to be mainly situated in the A3 region (see figure 2), and perhaps in the neighbouring left-hand section of the B-region (Table 3). In some of the conditions, in particular the conditions that included the use of the CEMRE, mental load may have been additionally increased. The CEMRE reduced the visual field and subjects were therefore required to make additional head movements (see also appendix 1). However, in none of the studies is demand expected to be excessive.

Two studies in which driver state was affected were added. In the three load-conditions of the two studies, task difficulty was increased because driver state that was non-optimal as a result of the use of alcohol, a sedative drug or fatigue that followed lengthy driving. Performance in the load condition of these studies is expected to be, on average, situated in the A1 or D region (figure 2). The actual,

individual region of performance, however, will depend upon individual capacity, experience and goals set for performance.

*Driver state studies:*

- (6) In the ‘**DREAM**’ (Driver Related Evaluation And Monitoring) study (appendix 5) the effects of legally-allowed levels of Blood Alcohol Concentration ( $BAC \leq 0.5 \%$ ), and the effects of fatigue (2.5 hours of driving, indicated in appendix 5 as ‘vigilance’), were compared with baseline performance (the first hour of the last-mentioned condition). Driving on a busy four-lane ringroad and on a monotonous motorway were included in the study.
- (7) Finally, in the ‘**antihistamine study**’ the effects of a new-generation antihistamine (Ebastine) were compared with placebo and an active-drug control, Triprolidine. The active drug, which has a sedative effect, was chosen as the experimental condition and its effects were compared with the effects of placebo. In both conditions subjects had to drive on a busy four-lane ringroad and on a four-lane motorway.

Table 3. Traffic studies that are referred to in the figures in the following sections: region is the a priori and thus expected region of task performance as shown in figure 2. ‘condition indicated’ designates how the condition is referred to in the figures’ legends, while the number of subjects is indicated under ‘N’. References with a # are listed in full as appendix to this thesis.

<u>study</u>	<u>test environment</u>	<u>selected load condition(s)</u>	<u>condition indicated</u>	<u>region</u>	<u>N</u>	<u>references</u>
<i>complexity, environment:</i>						
<b>Weaving section</b>	On-the-road	combined entrance/exit entrance/exit + Eye mark.	Weav Weav(c)	A3 A3-B	52	De Waard (1991)# De Waard et al. (1990)
<b>Noise barrier</b>	On-the-road	Noise barrier Noise barrier + Eye mark.	NoiseB NoiseB(c)	A3 A3-B	22	Jessurun et al. (1990)
<b>Road layout</b>	On-the-road	Woodland Road, exp. Moorland Road, exp.	Wr Mr	A3 A3	28	De Waard et al. (1995)# Jessurun et al. (1993)
<i>complexity, task:</i>						
<b>Car Phone</b>	On-the-road	phone, motorway phone, ringroad	Pmw Prr	A3 A3-B	12	Brookhuis et al. (1991)# Brookhuis et al. (1989)
<b>Tutoring</b>	Simulator	warning messages	Tut	A3	27	De Waard et al. (submitted)# De Waard et al. (1994)
<i>state:</i>						
<b>DREAM</b>	On-the-road	Alcohol, motorway Alcohol, ringroad Fatigue, motorway Fatigue, ringroad	Alc(mw) Alc(rr) Fat(mw) Fat(rr)	D-A1 D-A1 D-A1 D-A1	20	De Waard & Brookhuis (1991a)# De Waard & Brookhuis (1991b) Brookhuis & De Waard (1993) Thomas et al. (1989)
<b>Anti-histamine</b>	On-the-road	Triprolidine, motorway Triprolidine, ringroad	Tri (mw) Tri (rr)	D-A1 D-A1	15	Brookhuis et al. (1993) De Vries et al. (1989)

In table 3 the above-mentioned studies are listed. In the following sections and figures the different selected conditions will be referred to as indicated in the column 'condition indicated'. 'N' denotes the number of subjects that completed the tests.

Table 4. Measures used in each study (■). Measures will be explained in the next chapter. Alcohol and fatigue were conditions in one study, the DREAM study. SECOND. = secondary.

GROUP:	SELF-REPORT	PRIMARY TASK PERFORMANCE	SECOND. TASK PERFORMANCE	PHYSIOLOGICAL
Measure:	R R A S E C M C T E L- I A V	S S T D D L L S C P T W	D M E E I Y L R E A R M Y O O R V	H H .1 E E R R 0 M E V G G H z
Study				
Weaving S.	<input type="checkbox"/> ■ <input type="checkbox"/>	■ ■ <input type="checkbox"/>	<input type="checkbox"/> ■ ■	■ ■ ■ <input type="checkbox"/> <input type="checkbox"/>
Noise Barr.	<input type="checkbox"/> ■ <input type="checkbox"/>	■ ■ <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> ■	■ ■ ■ <input type="checkbox"/> <input type="checkbox"/>
Rd Layout	<input type="checkbox"/> ■ <input type="checkbox"/>	■ ■ ■	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	■ ■ ■ ■ <input type="checkbox"/>
CarPhone	■ <input type="checkbox"/> <input type="checkbox"/>	■ ■ <input type="checkbox"/>	■ ■ <input type="checkbox"/>	■ ■ ■ <input type="checkbox"/> <input type="checkbox"/>
Tutoring	■ <input type="checkbox"/> <input type="checkbox"/>	■ ■ <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	■ ■ ■ <input type="checkbox"/> <input type="checkbox"/>
Alcohol	■ <input type="checkbox"/> ■	■ ■ ■	■ <input type="checkbox"/> <input type="checkbox"/>	■ ■ ■ <input type="checkbox"/> ■
Fatigue	■ <input type="checkbox"/> ■	■ ■ ■	■ <input type="checkbox"/> <input type="checkbox"/>	■ ■ ■ <input type="checkbox"/> ■
AntiHistam.	■ <input type="checkbox"/> ■	■ ■ <input type="checkbox"/>	■ ■ <input type="checkbox"/>	■ ■ ■ <input type="checkbox"/> <input type="checkbox"/>

Which workload measurement method was used in which study can be seen in table 4. Three self-report scales were used of which two were unidimensional (RSME and Activation scale). The third scale, the activation scale of the RECL (Road Environment Construct List, see below), is based on multiple Likert-scales. As primary-task performance measures, the *SD* of the lateral position (SDLP), the *SD* of the steering-wheel movements (SDSTW), and the Time-to-Line Crossing (TLC)-measure were used. Mirror checking and delay in speed adaptations to a lead car's speed changes in a car-following task are listed under secondary-task performance as embedded tasks. A genuine secondary task, performance on the PASAT, was only applied in the car-phone study. Three heart-rate measures are listed under physiology, average heart rate (HR), the modulation index of heart rate variability in the time domain (HRV) and variability in the frequency domain, in the 0.10 Hz band (.10 Hz). Activity of the facial corrugator muscle was used in one study, while ongoing EEG activity was used as physiological measure in the alcohol and fatigue (vigilance) conditions of the DREAM experiment.

The evaluation of measure sensitivity to workload, and in particular to differences in sensitivity to increased load in terms of

affected state opposed to increased complexity, will focus on the measures that were available in most studies, i.e.,

as self-report measures:

- RSME (Effort rating scale)
- Activation Scale

as primary-task performance measures:

- SD of the lateral position (SDLP)
- SD of the steering wheel movements (SDSTW)

as physiological measures:

- Average Heart Rate (HR)
- Heart Rate Variability in the time domain (HRV)
- Heart Rate Variability in the 0.10 Hz frequency domain (0.10 Hz)

For the sake of completeness not only the studies mentioned in table 3 will be evaluated, but other studies that were carried out in traffic and were found in the literature will, as far as possible and relevant, also be included in the next chapters.

---

## 5.1 Self-report measures

In this chapter, experience with driver self-report workload ratings will be described. The Dutch RSME and the originally German Activation scale will first be treated. The activation scale of the RECL is 'an odd one out', but is included because it was the only self-report rating that is available from the Weaving section, Noise Barrier and Road Layout studies (see table 4). Results obtained by others with the Task Load Index and SWAT are discussed under other self-report measures.

### **RSME, Rating Scale Mental Effort**

In traffic research, the RSME (Zijlstra & Van Doorn, 1985, Zijlstra & Meijman, 1989) was used in the car-phone study and in the simulator experiment, and effects are compared with effects of the sedative antihistamine Triprolidine and the effects of alcohol and time-on-task (Car-Phone, Tutoring, Antihistamine and DREAM respectively in table 3). In figure 3 the absolute scores on the RSME scale of these four studies are indicated. Baseline ratings of effort of driving are compared with ratings of effort while driving and using a car-phone (load), driving without (baseline) vs. with (load) a switched-on enforcement and feedback system (Tutoring), and driving under placebo vs. under the influence of Triprolidine (load). The effects of 0.5 ‰ alcohol and fatigue (2.5 hours of driving) could not, due to the experimental design, be compared with baseline ratings, which could

not be collected. In figure 4 the change in scale values of the load condition opposed to baseline is indicated for the studies that included such a condition. All ratings were collected after completion of the driving task.

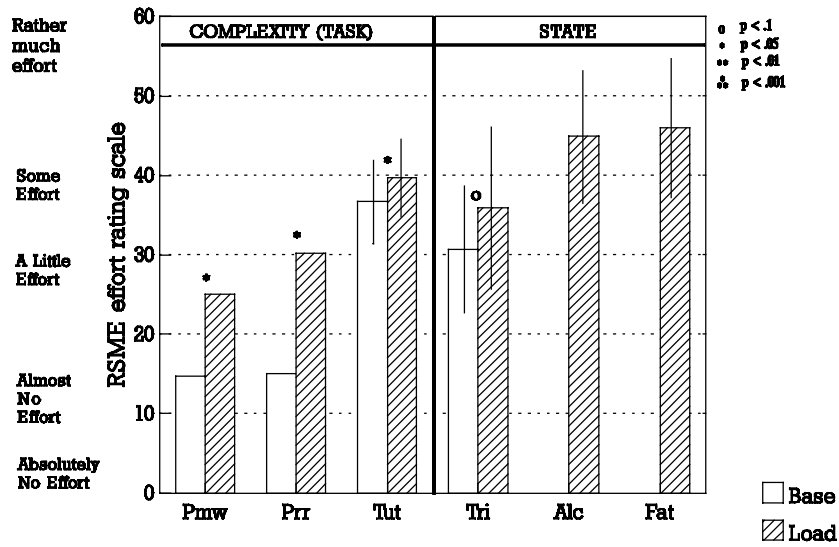


Figure 3. Average ratings of exerted effort on the unidimensional RSME of baseline driving and car-phone use (both on the motorway, Pmw, and on a busy ringroad, Prr), driving with and without an enforcement & tutoring system (Tut), driving under placebo and Triprolidine (overall rating, Tri), and driving on the motorway under the influence of alcohol (Alc(mw)) and while fatigued (Fat(mw)). If available, the 95% confidence interval is indicated.

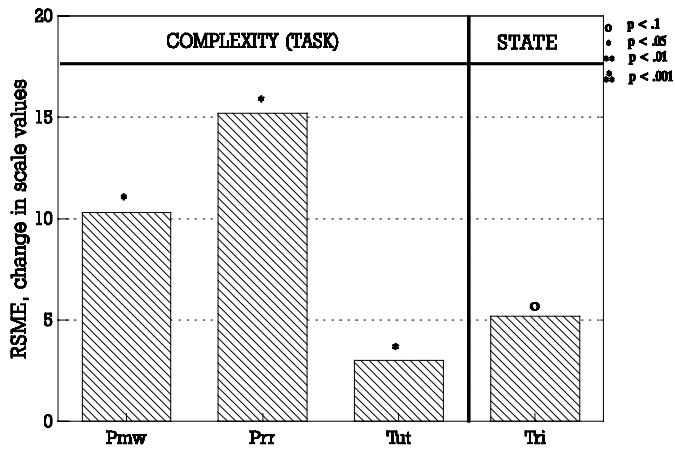


Figure 4. Average change in ratings of exerted effort on the unidimensional RSME in the case of car-phone use (Pmw and Prr), driving with an enforcement and tutoring system (Tut), driving under influence of Triprolidine (Tri), all compared with the baseline (control, placebo) measurement.



In all cases the RSME was able to distinguish between task-load situation and baseline. An increase in effort was reported in the case of car-phone use and as a result of the behavioural adaptation required by the enforcement system. The sedative effect of Triprolidine also resulted in an increase in effort exerted. Between the Tutoring and Car-phone study important differences in baseline values were found. These differences may reflect differences between the subjects who participated, but it is more likely that they reflect differences between the baseline tasks. As mentioned previously, the effects of task load were compared with baseline driving. For the Tutoring experiment, baseline driving included handling a simulator car and driving through a varied area, while in the Car-phone experiment an instrumented vehicle had to be driven through traffic. Judging from the absolute scores, the latter task is less effortful. Recently, support for this statement was found in a study in which the same subjects performed the same task both in traffic and in a simulator (De Waard & Brookhuis, in press). Driving in the simulator required more effort as measured with the RSME.

### **Activation scale**

Bartenwerfer's activation scale was used in two studies that are listed in table 4. In the DREAM experiment, however, no baseline ratings could be collected<sup>6</sup>. The effect of Triprolidine on reported activation level estimated over the whole journey, was not significant. The application of the activation scale to traffic research has mainly been limited to drug research. An indication of the measure's sensitivity to affected driver state can be obtained by looking at the results from these 'drugs & driving' studies. In figure 5 the change in scale values, compared with placebo, is listed for drugs as measured in five on-the-road studies. The average placebo value over all studies was 131, which is just below the reference point 'I am solving a crossword puzzle' (see appendix B for the scale). Data regarding antidepressants, hypnotics, analgesics, tranquillizers and antihistamines have been taken from Louwerens et al. (1983), Volkerts et al. (1984), Brookhuis et al. (1985a), Volkerts et al. (1987) and De Vries et al. (1989), respectively.

The most pronounced effect on reported activation level was the reduction found in the antidepressant study. One hypnotic reduced reported activation level, while the analgesic showed a dose-related effect. This last effect was not in the expected direction, activation level increased with an increase in dose of this drug for pain-treatment. However, in that study performance measures did not decline with increasing dose either, and nor did reaction-time performance in a laboratory task (see Brookhuis et al., 1985a).

---

<sup>6</sup> Average rating for Alc(mw) was 130.5 and for Fat(mw) 139.0.

## Bartenwerfer's activation scale

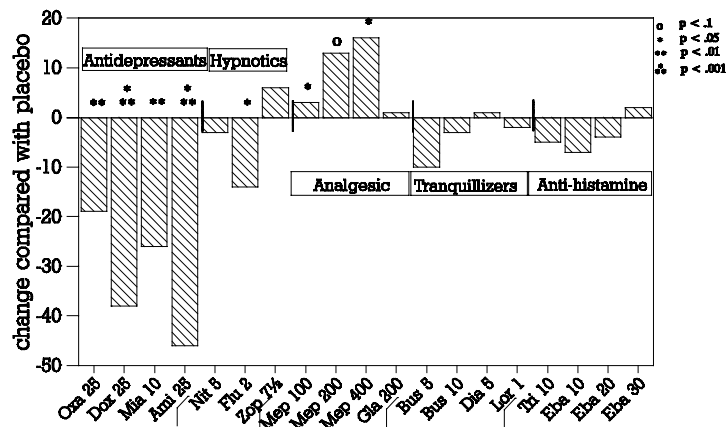


Figure 5. Average change in rated activation in five drug studies compared with the placebo conditions

No effects on the scale were found in the tranquillizer and antihistamine studies. On the basis of these studies it seems that the scale is fit to be used for effects on subjectively experienced effects on the Central Nervous System. The relation of the scale to mental workload in general is, at present, hard to assess. It seems likely that the scale is of particular use in the areas further away from optimal performance, hence in the D and C regions.

### Other self-report measures used

*RECL, Road  
Environment  
Construct List*

The Road Environment Construct List (RECL, Steyvers, 1993, Steyvers et al., 1994) was developed to measure appraisal of road environments. The RECL is a three factor scale. Each of the sixteen items load on one of three factors. The factors are: 'Hedonic value', which denotes the aesthetic appraisal of the road and its environment, 'Perceptual variation' denoting the heterogeneity in the road environment, and 'Activation value' denoting the extent to which the road and environment are considered to be activating. The latter factor may be useful for workload measurement in a traffic environment.

The RECL was used in studies in which the RSME was not used and therefore the Activation value of the RECL is included in the evaluation on usefulness as an indicator of driver activation. Though the driver is asked to evaluate the road and its environment, an activating effect of the environment could be related to road-environment demands and might therefore influence driver mental activation.

Although the trend in scores of baseline and load conditions in the road-layout experiment was in the direction of increased load, differences between the two conditions on both roads (Wr and Mr) were not significant. In the two motorway studies no baseline measurements were taken. However, two other conditions of these experiments could be compared: driving without and with ('c') eye-movement equipment mounted on subjects' heads. In both studies subjects did not rate the activating influence of the environment different as a result of the equipment (see table 5).

Table 5. Average rating on the Activation scale of the RECL. Baseline measurements were only collected in the road layout study.

	Baseline	Load	Significance (t-test)
Weaving Section	-	3.6	] ns
Weaving Section (c)	-	3.7	
Noise Barrier	-	3.6	] ns
Noise Barrier (c)	-	3.7	
Road Layout Wr	4.2	4.5	ns
Road Layout Mr	3.2	3.8	ns

### Other self-report measures in other studies

*TLX, Task Load Index*

In none of the studies listed in table 3 was the NASA Task Load Index (TLX) used. A few on-the-road studies reporting the use of this self-report measure were found. Fairclough et al. (1991) used the RTLX (Byers et al., 1989) in a dual-task performance study. They found an increase in overall workload in the dual task condition, which consisted of driving plus having a conversation, compared with single-task performance, which was normal driving. The RTLX was also used in another study performed in the same vehicle (Vaughan et al., 1994). In the experiment RDS (Radio Data System) messages had to be attended to. The messages were presented to subjects in three conditions in a within-subjects design: 1. auditory, 2. auditory and continuously visible on a display, and 3. auditory and temporarily (15 s) visible on a display. Overall RTLX mental workload rating was lowest for condition 2, auditory plus visual constant. The RTLX factors 'mental effort' and 'time pressure' showed a similar effect (the lowest rating for condition 2, the highest rating for condition 1 and slightly less high for condition 3). The results found in this dual task study illustrate the diagnosticity of the RTLX in the reflection of higher scores on the time-pressure factor in the case of auditory messages and no or quickly disappearing visual information.

In a simulator study in which the effects of a hands-free car-phone were tested Alm & Nilsson (1994) found an effect of the car-phone task on all subscales of the TLX. An interaction between car-phone use and driving-task difficulty (in terms of driving a straight

opposed to a winding road) was only found on the frustration subscale, and not on the mental-demand or operator-effort subscales.

## SWAT

The SWAT was used in simulator and on-the-road experimental tests of the GIDS system (Janssen et al., 1994). The system gave support to the driver by route guidance messages, and with respect to speed, collision avoidance and lane keeping (simulator trials only). Judging from the SWAT-reference that was provided in the text, an adapted version was used in which the card-sort section was left out. The authors report the overall mental workload index, which is defined as the addition of three 3-point scales (time stress, mental effort and psychological stress) resulting in a sum-scale range from 3 to 9. SWAT ratings differed between integrated and non-integrated GIDS support both in the simulator trials and in the on-the-road tests. The difference between integrated and non-integrated support was that support was only scheduled according to demand in the first condition. Scheduling includes, for instance, postponing an incoming phone call in the event that a lead vehicle brakes suddenly.

In an on-the-road experiment Verwey & Veltman (1995) found that summational SWAT ratings were equally sensitive to increases in workload as ratings on the RSME. Inclusion of the card-sort task for SWAT did not yield more accurate workload estimates.

### **Properties of self-report measures**

Sensitivity, selectivity, diagnosticity, validity and primary-task intrusion are of major importance for a measure of driver workload. These properties were assessed as adequately as possible on the basis of the above-described experiments. The region in which the measure was found to be sensitive is indicated under sensitivity, and region-sensitivity has to be considered the prime property.

The RSME is designed to reflect operator effort. In the car-phone and tutoring experiments the RSME was found to be sensitive to task-related effort, while in the antihistamine study the rating scale was sensitive to state-related effort. Accordingly, when performance is in Region A1 and A3/B the RSME can be expected to reflect driver mental effort. The drug studies showed that the activation scale is in particular sensitive to an affected driver state as a result of (highly) sedative medicine such as hypnotics and antidepressants. Increased activation levels, e.g., as a result of the use of amphetamine (Sanders, 1983), can be expected to be reflected in higher activation scores, but as yet, there is, to my knowledge, no evidence available from empirical studies to support this prediction.

Diagnosticity for the two unidimensional scales is low unless they are applied per task dimension as proposed by Zijlstra & Meijman (1989). Selectivity is difficult to assess as the main other factor to which the scales could be sensitive, physical workload, is very restricted in driving. Reliability is high, as sensitivity to mental

workload in the different studies is high. Primary-task intrusion is low as long as the ratings are asked after completion of the task. Since hardly any equipment is required for collection of the measures the implementation requirements are low. No problems in operator acceptance have been encountered, so informal evidence supports high operator acceptance. In table 6 the results are summarized.

Table 6 Summary of properties of self-report workload measures.

Property	Measure	
	RSME	Activation
sensitivity (Region)	(D-)A1, A3-B	D, (B-C)
diagnosticity	low	low
selectivity	prob. high	(?)
Reliability	high	high (?)
primary-task intrusion	low	low
implementation requirements	low	low
operator acceptance	high	high

## 5.2 Primary-task performance measures

Parkes (1991) defined the primary task of the driver as maintenance of safe control over the vehicle. One of the major subtasks in vehicle control is lateral position control. Therefore, a measure of driving deviations from the centre of the lane is a good means to assess primary-task performance in car driving. Lateral deviation, or more specifically the *SD of the Lateral Position* (SDLP), has been shown to be a sensitive performance measure (e.g., Hicks & Wierwille, 1979, O'Hanlon et al., 1982, O'Hanlon, 1984, Brookhuis et al., 1985b, Green et al., 1993b). The task of keeping a vehicle between the lines of a lane is largely a psychomotor task involving eye-hand coordination. The term 'tracking-ability' is sometimes applied to it (e.g., Stein et al., 1987), stressing the strong resemblance to the laboratory task.

### Standard Deviation of the Lateral Position

In figure 6 the average (right-hand lane) SDLP in baseline and load conditions is displayed, while in figure 7 the change in *SD* of the lateral position compared with baseline is shown. This relative measure was added to neutralize the differences in baselines between studies, which are likely to have been caused by differences between roads/road segments, season (weather) and so on. The absolute value of the SDLP in the Tutoring experiment is omitted from figure 6 because in the experiment the road width and test environment were very different from the on-the-road tests. In both figures the critical impairment levels

(Brookhuis, 1995a, 1995b) are indicated, while in figure 7 the impairment in lateral position control found in an 'alcohol calibration study' (Louwerens et al., 1987) is included.

An increase in SDLP, i.e. an increase in swerving, was found near the noise barrier (but only in the condition without eye-movement measurement), and as a result of alcohol (Alc(mw)) and prolonged driving (Fat(mw)). A decrease in the SDLP in the mental load condition was found in conditions in which subjects handled a car-phone (Prr and Pmw), when the enforcement system was switched on (Tut), and on the experimental road-layout (Wr and Mr). In some cases, the relative short section that was selected as load condition could have had an effect on SDLP. Near the noise barrier, for instance, the average lateral position on the road moved to the left. In the road-layout experiment the road surface and effective road width had been reduced, forcing drivers into more accurate lane-keeping. The effect of lane width on tracking performance was also found in a pilot study performed in a driving simulator (Green et al., 1993b), they found an increase in SDLP with increases in lane width. Taking these factors into account leaves only primary-task performance *decrements* on the SDLP measure as a result of alcohol and prolonged driving. In the Tutoring and Car-phone experiment primary-task performance under mental load, as measured by SDLP, even improved, while the sedative drug Triprolidine and driving on the Weaving Section did not lead to a significant increase in SDLP.

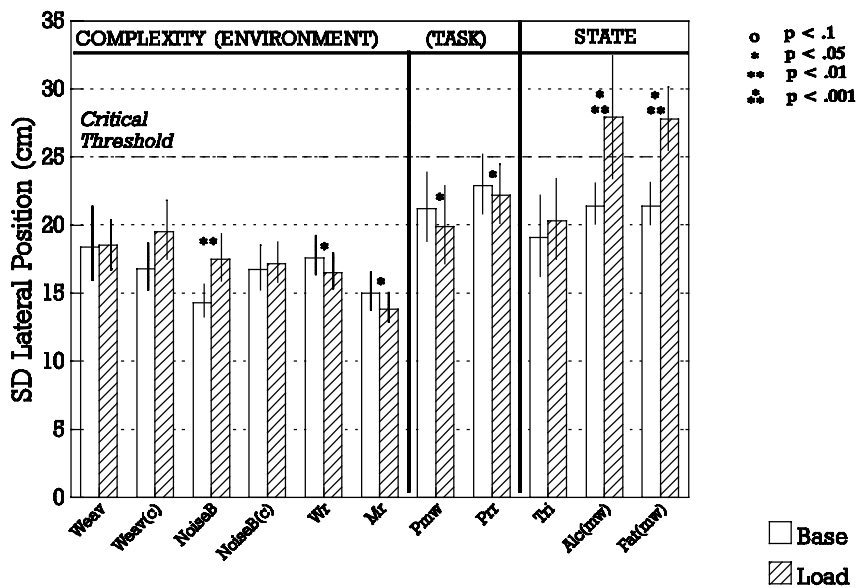


Figure 6. Standard deviation of the lateral position under baseline and mental load conditions. The studies from which the conditions were selected are listed in table 3. The indicated absolute threshold indicates driver impairment (see Brookhuis, 1995ab). The 95% confidence interval is also indicated.

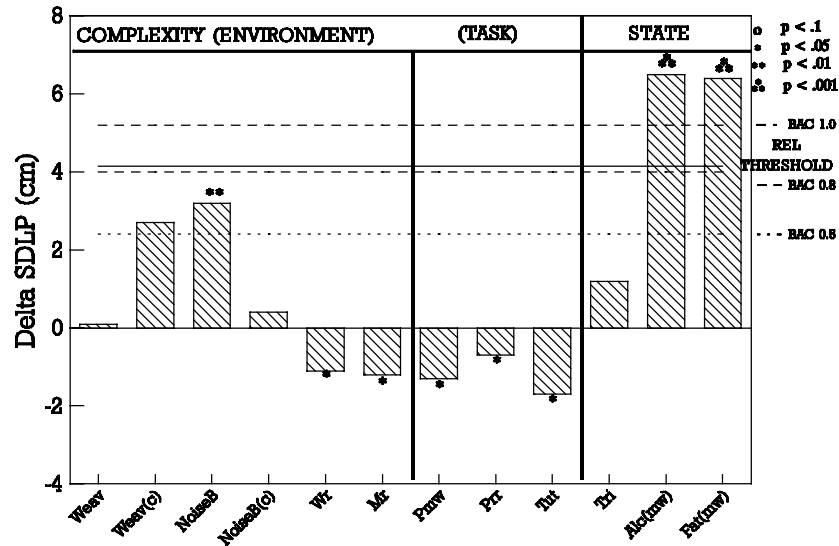


Figure 7. Change in standard deviation of the lateral position under mental load compared with baseline measurements. The studies from which the conditions were selected are listed in table 3. The indicated relative threshold denotes driver impairment (see Brookhuis, 1995ab). The indicated BAC (Blood Alcohol Concentration) values are impairment levels as found in an 'alcohol calibration study' (see Louwerens et al., 1987)

### Standard Deviation of the Steering wheel movements

Related to the SDLP, but closer to one of the main sources of swerving, is the driver's steering behaviour. Due to relatively low-attentional driving demands, or due to attentional demands of additional tasks, drivers do not pay continuous attention to the lane-tracking (steering) task. This results in steering 'holds', i.e. periods without steering-wheel movements (see Macdonald & Hoffmann, 1980, Godthelp et al., 1984). Several steering measures have been developed, from relatively simple measures, such as the number of zero-degree crossings of the steering-wheel or steering-reversal rate (McLean & Hoffmann, 1975), to more complex measures involving frequency analyses (McLean & Hoffmann, 1971, Blaauw, 1984) and compound functions (Fairclough, 1994). Steering-reversal rate (McLean & Hoffmann, 1975, Macdonald & Hoffmann, 1980) and the *SD* of the steering-wheel movements, always measured on straight road segments, are frequently used performance measures that are not complicated to calculate. In the figures 8 and 9 the ( $\Delta$ ) *SD* of the steering-wheel movements (*SDSTW*), on sections with hardly any or no curvature is shown. Again the critical impairment level (Brookhuis, 1995ab) is displayed in both figures. In three studies the *SDSTW* increased in the load condition, in two studies to a level above the absolute impairment criterium. The elevated *SDSTW* at the experimental road-layout (Mr)

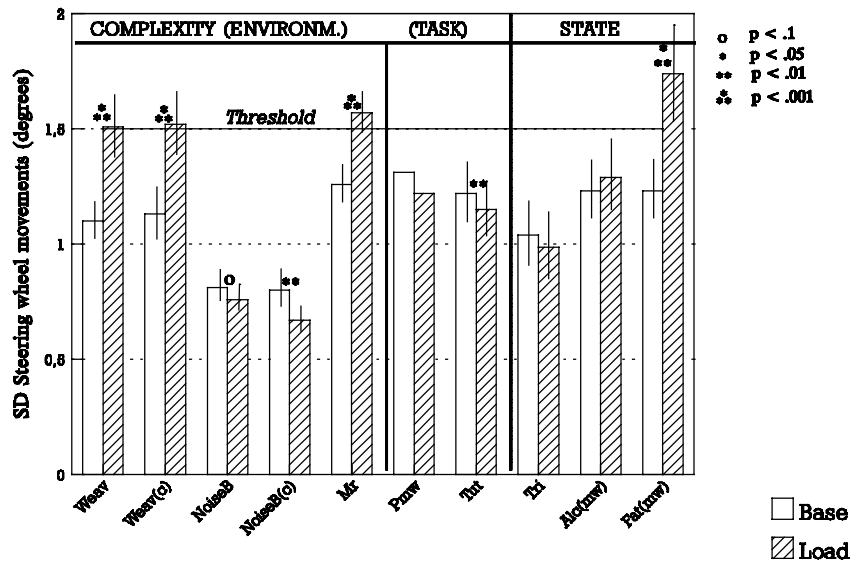


Figure 8. Standard deviation of the steering-wheel movements under baseline and mental load conditions. The studies from which the conditions were selected are listed in table 3. The critical threshold level indicates driver impairment (see Brookhuis, 1995ab). If available the 95% confidence interval is displayed.

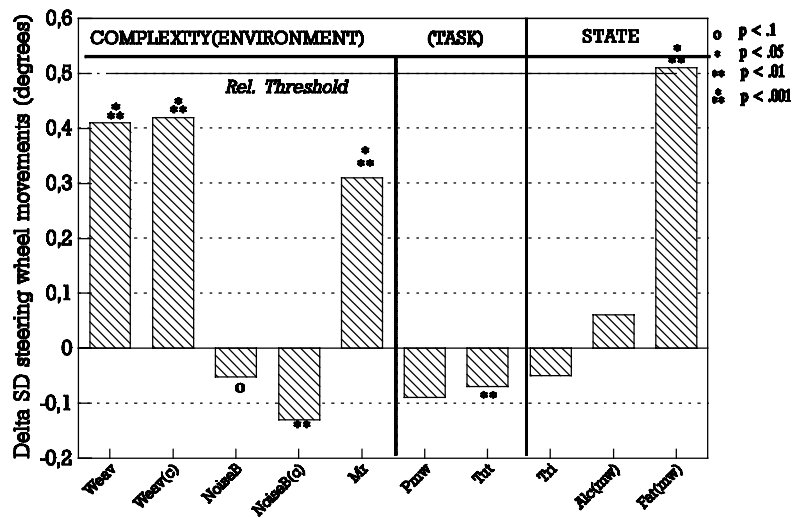


Figure 9. Change in standard deviation of the steering wheel movements under mental load compared with baseline measurements. The studies from which the conditions were selected from are listed in table 3. The indicated relative threshold indicates driver impairment (see Brookhuis, 1995ab).



was unexpected. However, the two selected road segments may have differed slightly in curvature. The experimental road section was somewhat more curved. Road curvature may have had a similar effect on the SDSTW in the experiment that focused on the Weaving Section. In the other experiments completely straight or even the same motorway sections were compared with each other, which in general is to be preferred. In the simulator (Tut) and Noise Barrier study a significant decrease in SDSTW was found. A decrease in SDSTW may be indicative of increased steering effort, and thereby of more accurate steering, e.g., as a result of road environmental demands.

#### A combined statistical test

The statistical power of the individual tests can be increased by combined testing of the effects found in the different experiments. If it is assumed that it is the same parameter that is affected in the different studies (and that that parameter is mental workload), then the effects found in the studies can be tested in combination by (Snijders, 1995):

$$z = \frac{\sum_{i=1}^k \alpha_i \theta_i}{\sqrt{\left( \sum_{i=1}^k \alpha_i^2 SE_i^2 \right)}}$$

with  $k$  = the number of experiments  
 $\theta_i$  = the estimated effect in experiment  $i$   
 $SE_i$  = standard error in experiment  $i$   
 $\alpha_i$  = weight of experiment  $i$ .

$z$  is tested in a standard-normal distribution, with  $H_0: \theta = 0$ .

This test was applied to the SDLP and SDSTW measures. The following results were found:

---

*SD of the lateral position:*

Complexity <sup>a,b</sup> :	$z = -0.29, NS$
Complexity (environment) <sup>a,c</sup> :	$z = +1.87, p < 0.05$
Complexity (task) <sup>b</sup> :	$z = -2.63, p < 0.005$
State <sup>c</sup> :	$z = +4.51, p < 0.0005$

<sup>a</sup> = Without  $W_r$  and  $M_r$  due to reduced road width.

<sup>b</sup> = Weighted,  $\alpha = 1$  for  $P_{rr}$  and  $P_{mw}$ ,  $\alpha = 2$  for Tut.

<sup>c</sup> = All conditions equal weights.

---

The road layout experiment was excluded from the tests as changes in SDLP cannot be solely attributed to changes in mental workload, but are combined with effects of reduced road width. In the Complexity tests the two car-phone conditions were weighted to balance effects with the single condition of the tutoring experiment.

Increased complexity in terms of a change in environment as opposed to additional tasks have dissimilar effects on SDLP. Increased task complexity concur with reduced SDLP, while increased environmental demands coincide with an increase in SDLP. Tested together as effect of ‘complexity’ levels out effects and renders a nonsignificant result. These results will be further discussed in chapter 5.5.

---

*SD of the steering wheel movements*

Complexity <sup>a</sup> :	$z = +3.86, p < 0.0005$
Complexity (environment) <sup>d</sup> :	$z = +8.35, p < 0.0005$
Complexity (environment) <sup>a</sup> :	$z = +4.93, p < 0.0005$
Complexity (task) <sup>e</sup> :	-
State:	$z = +16.45, p < 0.0005$

<sup>a</sup> = Without Mr due to reduced road width in load condition.

<sup>d</sup> = Mr is included in this test with  $\alpha_{Mr} = 2$ , while  $\alpha_{Weav} = \alpha_{Weav(c)} = \alpha_{NoiseB} = \alpha_{NoiseB(c)} = 1$ .

<sup>e</sup> = Not tested, only standard error information from one study (Tut) available

---

An increase in complexity of the environment and a decreased driver state both lead to a significant increase in the SD of the steering wheel movements. Increased task complexity reduces the SD of the steering wheel movements. These results will, together with the effects on SDLP, be discussed in chapter 5.5.

**Other primary-task performance measures**

*Time-to-line crossing*

While the SDLP and SDSTW mainly reflect performance at the control level, one level higher, at the manoeuvring level of performance, the *Time-to-Line Crossing* (TLC, Godthelp 1984) is a measure of driver primary-task performance. TLC is a continuous measure that represents the time required for the vehicle to reach either the centre or edge line of the driving lane if no further corrective steering-wheel movements are executed. TLC reflects the amount of time drivers can neglect path errors. Due to the measure’s skewness, in general minimum, median or 15% TLC values are calculated (Godthelp et al., 1984, Godthelp, 1988). TLC is expected to reflect driving strategy and in particular occlusion strategy (time spent not looking at the road). With increases in mental-load, smaller TLC values can be

expected; a more demanding task is likely to decrease the amount of time spent looking at the road.

In table 7 median and minimum TLC, as well as the change in TLC relative to baseline are depicted for the DREAM (for TLC see De Waard & Brookhuis, 1991b) and road-layout study. In the vigilance condition a decrease in TLC was found. This is in accord with the increase in number of steering-wheel holds that was found as a result of time-on-task (De Waard & Brookhuis, 1991b). In the road-layout study the layout of the road had been changed significantly. Drivers were more or less forced to drive close to the centre line and as a result the left-hand TLC decreased, while the right-hand TLC increased. This measure actually reflects the time required to reach an *imaginary* edge line, as the line had been removed! As a result, interpretation of the TLC measures in terms of mental load measures is not useful with data of the road-layout study.

Table 7. Median and minimum time-to-line crossing (s) in baseline and mental load conditions. Change in TLC denotes the change from baseline to load. Significant results have been printed in **bold**.

condition	left hand Median TLC		left hand Minimum TLC		right hand Median TLC		right hand Minimum TLC	
	base	load	base	load	base	load	base	load
Mr	4.95	<b>4.27</b>	2.35	<b>1.91</b>	3.10	<b>3.59</b>	1.19	<b>1.50</b>
Alc(mw)	5.31	5.42	1.89	1.83	3.79	3.98	1.58	1.69
Fat(mw)	5.31	<b>3.95</b>	1.89	<b>1.71</b>	3.79	<b>3.11</b>	1.58	<b>1.30</b>

	Median TLC Change in TLC		Minimum TLC Change in TLC	
	left	right	left	right
Mr	<b>-0.68</b>	<b>0.49</b>	<b>-0.44</b>	<b>0.31</b>
Alc(mw)	0.11	0.19	-0.18	0.11
Fat(mw)	<b>-1.36</b>	<b>-0.68</b>	<b>-0.07</b>	<b>-0.28</b>

### Results with respect to TLC, SDLP and SDSTW from other author's studies

Riedel (1991) also used the TLC measure in a drug study in which subjects performed a driving task on the road. He found a maximum *increase* in median TLC (undifferentiated to line) of 0.15 s in the Triazolam condition, while baseline median TLC on the motorway was 4.69 s. SDLP in the same condition increased with 6.6 cm to 30.7 cm. On the basis of these data, he concluded that SDLP is the most sensitive measure for driver impairment.

The effect of Blood Alcohol Concentrations on SDLP as found by Louwerens et al. (1987) have been indicated in figure 6. The sensitivity of the measure to an affected driver state as a result of the

use of hypnotics are summarized in Brookhuis (1995b). Significant increases in SDLP starting at 2.5 up to 7 cm are reported.

Van Winsum et al. (1989) compared steering-wheel movements of drivers who navigated from a map with the steering-wheel movements of drivers who were guided vocally. They found no effect on steering-wheel movements in the more demanding map condition. This result may be related to the urban road environment. It is likely that the use of most primary-task control indices (SDLP, SDSTW) is confined to non-urban environments. In urban traffic most steering-wheel movements will be related to longitudinal and lateral tracking demands (Wildervanck et al., 1978).

Green et al. (1993a) compared driving behaviour and self-report ratings of difficulty of route guidance messages using three different interfaces. Only slight differences in *SD* of steering wheel movements were found, the largest *SD* of steering wheel movements were measured when the information was displayed in the instrument panel (1.1°), followed by a simulated Head-up display (1.0°). The *SD* of the steering wheel movements were smallest (0.9°) for auditory presented information. Ratings of difficulty of use of the route guidance information while driving that were given after the test rides (Green et al, 1993a, p.82) followed the same pattern, the lowest difficulty rating being given for the auditory information. However, memory load in the case of auditory route guidance was largest. In all three conditions route guidance information was additionally combined with information regarding vehicle state and traffic information that was presented to the driver in the instrument panel at a different location. This additional information could have interacted with the route guidance messages and therefore a relation between type of interface and mental load is hard to assess accurately.

Fairclough (1994) measured steering-wheel movements in a study in which subjects drove under the influence of low amounts of alcohol, and under placebo conditions. Just as in the DREAM study (see figure 8) he found an increase in the standard deviation of steering-wheel movements of drivers with a BAC up to 0.5 ‰.

### **Other primary-task measures in other studies**

Apart from the above-mentioned *accuracy* measures in vehicle control, sometimes *speed* measures are used in the assessment of primary-task performance. An example of a speed measure is the time that is required to finish a route. Both Jordan & Johnson (1993) and Fairclough et al. (1991) found the time required to complete a route to be significantly longer in the load condition in which subjects had to adjust a stereo or had a conversation, compared with normal driving along the same route. The measure can be indicative of a strategic choice for a lower driving speed to compensate for high information load, and accordingly lead to a decrease in mental load. Similar compensatory strategies are reported for slower decision making and

slower action performance in elderly drivers (Brouwer & Ponds, 1994). Brown et al. (1969) also found an increase in time required to finish driving a circuit as a result of the use of a car-phone, while Van Winsum et al. (1989) found the same effect -an overall lower driving speed- when they compared map navigation with vocal-route guidance. However, the application of the measure is rather rough, and in non-controlled environments, e.g. in on-the-road studies, the measure is susceptible to disturbance factors such as traffic density. The use of speed measures as a sensitive indicator of increased mental load seems, therefore, to be the most reliable in laboratory and simulator experiments.

### **Properties of primary-task performance measures**

Lane-keeping in experienced drivers is to a large extent determined by automatic, control-level processing. Consequently, measures of accuracy in lane-keeping, such as the SDLP and SDSTW, would not be expected to be sensitive to variations in mental workload in the A-region. The different experiments, however, show that this is not the case, both SDLP and SDSTW being sensitive measures. A likely explanation for this is that there is no 'pure' automatic and controlled behaviour, but that aspects of automatic behaviour remain influenced by controlled processing (Schneider and Fisk, 1983). Strategy sets performance margins and the inaccuracies that are allowed. This also clarifies why improvement on these primary-task performance measures is possible. Increased task demands can lead to increased driver effort, which increase primary-task performance if under baseline conditions inaccuracies are allowed. This issue will be further discussed in chapter 5.5.

Although improvement in primary-task performance measures is possible, in general, affected task performance implies reduced task performance, and this is the case in the D, B and C regions. As task performance is at a minimum level in the C-region, performance measures will no longer vary with changes in demand in that region. Sensitivity of the SDLP and SDSTW is highest in the B and D regions. In studies in which driver state was reduced, a decrease in SDLP and SDSTW was found. The same is true for the increased environmental demand studies. Diagnosticity of the measures is low, although the difference in direction of the effect as found between Complexity environment vs. Complexity task may be an indication of differential sensitivity. Selectivity is hard to assess on the basis of the driving studies reported above. Hardly any physical effort is required in driving, and emotional stress, for instance, was not tested. It is quite possible that the measures are affected by these factors and therefore selectivity is expected to be relatively low. Sensitivity to mental workload as found in the different tests results in a 'high' rating for reliability. The implementation requirements for the measurement of steering wheel movements are low. A potentiometer mounted on the steering wheel

column with a measurement range of  $90^\circ (\pm 45^\circ)$  and a resolution of  $0.1^\circ$  is adequate for accurate measurement of movements on noncurved road sections. For the measurement of the vehicle's lateral position more complex equipment is required. A useful device is the so-called 'lane tracker', which resembles a camera but the interior consists of an array of diodes that are sensitive to differences in light intensity. The camera is directed towards the road delineation (see appendix 5). A relatively cheap but labour-intensive solution is to make video registrations of the road scene (De Waard & Steyvers, 1995). The advantage of the latter technique is that it can also be applied on roads without delineation. In the future progress in camera techniques will probably facilitate automatic detection of road delineation or road shoulder. Operator acceptance of the measures is high because registration is unobtrusive. Table 8 provides an overview of primary-task measures' properties.

Table 8. Summary of properties of primary-task workload measures.

Measures		
property	SDLP	SDSTW
sensitivity (Region)	D, B	D, B
diagnosticity	low	low
selectivity	(low)	(low)
reliability	high	high
primary-task intrusion	none	none
implementation requirements	high	low
operator acceptance	high	high

### 5.3 Secondary-task performance measures

If no specific instructions are given it is not clear which task is given priority. In heavy traffic the conversation with a passenger will probably be disrupted to maintain driving performance while in quieter environment and during a very interesting conversation driving performance will be affected (Wickens, 1984). Moreover, while the division between primary and secondary tasks may be very clear-cut for most laboratory tasks, this is not the case in driving. In traffic, behaviour is quite often related to the manoeuvre that is performed. Monitoring of rear traffic can be crucial if an overtaking manoeuvre is planned. In those cases the task of looking into mirrors and over one's shoulder cannot be called 'secondary'. Task integration can also blur the transition from primary to secondary task. A good example of dual-task integration is *car-following*. In heavy traffic this task will be added to the primary task of lateral and longitudinal vehicle control. It is the addition of a task, but the added task is not artificial. The experience of

various subtasks as a 'single task' is in particular likely if the subtasks are related or coherent (see, e.g., Korteling, 1994ab). Viewed in this way, car-following performance could be an embedded secondary task. However, a condition for a task to be termed embedded is that it is given lower priority than the primary task. It is not certain that car-following is given lower priority than lane-keeping. Perhaps a useful description of a secondary task in traffic research is that the task does not have to be performed continuously. In this way, the primary task remains restricted to speed and lateral vehicle control. Secondary tasks are non-continuous tasks, i.e. headway keeping can only be performed in case a lead vehicle is present and looking into the mirrors is performed at intervals. The definition is weak, but so is the separation of primary and secondary tasks in traffic.

#### *Car-following*

At the Traffic Research Centre a car-following task for use in real traffic has been developed (Brookhuis et al., 1994). In the task, a lead car's speed fluctuations have to be followed by the driver of an experimental vehicle. This task is designed to be sensitive to impairment of performance in attention and perception, while lane-tracking is merely sensitive to performance on eye-hand coordination. In terms of the hierarchical model of car driving (Janssen, 1979, Michon, 1985, see also chapter 1) the lane-tracking parameters (SDLP, SDSTW) reflect performance at the control level, while car-following parameters reflect performance at the manoeuvre level. The main parameter in car-following performance is the delay in reaction to speed changes of the lead vehicle. We (Brookhuis et al., 1994) obtain this measure by performing a coherence analysis on the speed signals of the lead and the following car. Apart from delay (calculated as 'phase shift' between the two speed signals in the frequency domain) two other parameters are computed, which both give an indication of 'how well' the car-following task is performed. Coherence is a measure of the accuracy of car-following performance, while the modulus indicates the amount of overreaction to speed changes by the following car (Porges et al., 1980).

The car-following task was included in the car-phone, DREAM and antihistamine studies. Delay increased in conditions in which a car-phone was used (+23%), after alcohol consumption (+19%), and in the condition in which Triprolidine had been taken (+42%). Time-on-task (Fatigue) did not affect delay, but coherence slightly decreased in this condition.

#### *Mirror checking*

Mirror checking is another good example of an embedded secondary task that is specific for car driving. Two variables can be distinguished in mirror checking: frequency and duration. Total *duration* of mirror checking was measured both in the Weaving Section and the Noise Barrier study. In the Noise Barrier study, however, only data related to the load condition were available. In this condition no more

than 2.7% of the total time was spent looking in the mirrors. In the Weaving Section study, the difference in mirror-looking time between load (10.6%) and control (10.2%) was not significantly different. In an in-vehicle navigation study, Fairclough et al. (1993) compared driving performance and visual attention while navigating from map vs. from a text-LCD screen. They found a decrease in duration of fixations in the rear-view mirror in the higher demand (i.e., map) condition. In another study, reported in the same paper, glance frequency (but not glance duration) in the rear-view mirror was decreased in the condition in which internal vehicle 'checking behaviour' of a display was higher. The authors' conclusion was that glance duration and glance frequency are representative for different aspects of driver behaviour. Duration appears to be sensitive to difficulty of information intake, while glance frequency represents visual activity in terms of checking behaviour, both inside (e.g., speedometer checking) and outside (e.g., mirror checking) the vehicle.

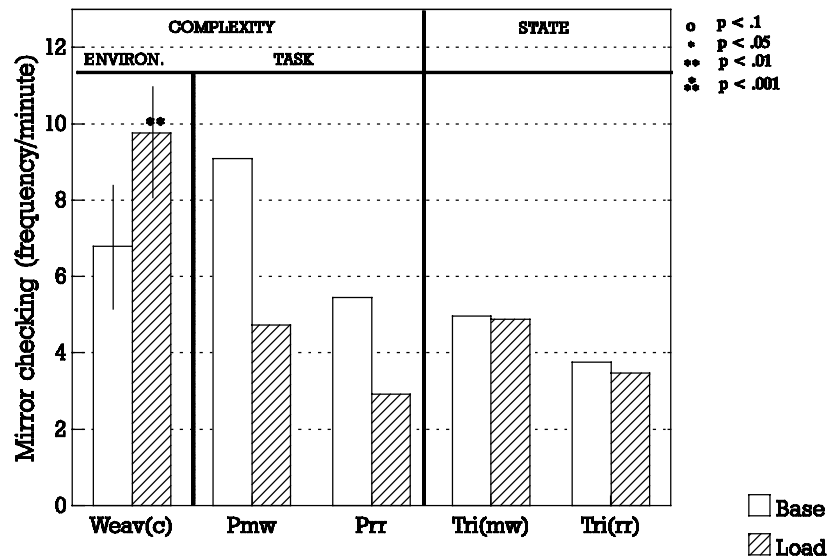


Figure 10. Frequency of interior and outside mirror checking in the car-phone and antihistamine study. In both studies motorway and ringroad sections were scored. If available the 95% confidence interval is indicated.

In the car-phone and antihistamine study, mirror checking frequency was scored from video, in both studies separately for the (quiet) motorway and (busy) ringroad. In the Weaving Section study the CEMRE-condition could be used to assess mirror-scanning frequency. As can be seen in figure 10, frequency of mirror-looking is reduced in the load condition of the car-phone study. The main effect of car-phone was not significant, but the interaction between road type and phone was. The larger effect of load on the motorway may be responsible for



this. No effect of load was found in the antihistamine study, only the effect of road type (again ringroad vs. motorway) was significant. In both studies mirror-looking frequency was lower on the more traffic-dense ringroad, where a car-following task had to be performed. In the Weaving Section study a significant *increase* in frequency of mirror checking was found in the load condition. This is particularly important because no difference in *duration*, i.e. proportion of the total time, of looking into the mirrors between the load and baseline conditions was found. Again the road environment may be responsible for the increase. The load section of the motorway was a combined entrance/exit with vehicles merging in and out of traffic, while the control section did not contain any entrances or exits. An increase in mirror-checking frequency and 'behind traffic monitoring' is important near entrances, even if no change-of-lane is planned, owing to the possible need of an evasive manoeuvre to the left-hand lane.

Rear mirror checking was also affected in the study reported by Van Winsum et al. (1989). In an unfamiliar environment, frequency of looking into the rear view mirror was reduced in the higher workload condition. Frequency of fixations seems most useful for workload assessment, though only if workload demand is not low. Fixation duration may be useful to assess certain aspects of task difficulty, in particular legibility, layout and amount of information (Fairclough et al., 1993).

#### *Additional tasks*

An actual additional task that had to be performed simultaneously to driving was the *PASAT*, the *Paced Serial Addition Task* (Gronwall & Sampson, 1974). The task itself is a demanding combination of a memory load and an addition test. This secondary task was used in the car-phone study, where the stimuli (digits) were presented over the phone. The task was used to create a fixed, heavy information-processing load on the subjects, more or less comparable to a difficult conversation. There was no control condition in which the task was performed without having to drive a car and/or use the car-phone. No significant differences in performance between the two road classes, motorway and ringroad, were found.

Earlier, at the end of the 60's, Brown et al. (1969) had studied the effects of telephoning on car driving performance by having subjects drive a car and perform gap-acceptance tests which were combined with a reasoning test. Subjects had to judge the correctness of sentences in relation to pairs of letters, e.g. "A follows B, -BA" (answer: True). Any impairment in driving performance could be attributed to divided attention; there was no need for the subjects to manually operate the car-phone. No effects of the additional task on primary-task vehicle-control measures were found, with the exception of an increase in time that was required to complete the circuit. Performance on the secondary task, however, was poorer in the

condition in which the task was combined with driving. Both reaction time and the proportion of errors increased. Gap-acceptance performance was also reduced by the additional task.

Verwey (1993b) carried out an experiment in which 48 subjects drove an instrumented vehicle over rural and inner-city roads while as secondary task they performed a visual detection task or an auditory addition task. While driving, subjects were guided by vocal messages issued by the experimenter. The experiment was a between-subject study with as factors: age (young vs. old), secondary task (auditory addition task vs. visual detection task), route familiarity (2 levels) and traffic density (2 levels). Subjects were instructed to give priority to the primary task of driving (Subsidiary Task Paradigm). Single-task performance of the secondary task while standing still was poorest for the elderly (79% opposed to 88% correct for the young). When driving, the older subjects' secondary-task performance (73% correct) was affected, while the younger subjects' performance did not decline (87% correct). Familiarity and traffic density had little effect on performance, while large differences on secondary-task performance were found between road situations. Between similar situations, i.e. between comparable road characteristics, no differences on secondary-task performance were found. In the study primary-task performance was only measured by assessment of speed control. Since different road segments had different speed limits, conclusions regarding primary-task performance are restricted. However, subjects unfamiliar with the road drove slightly slower and may therefore have reduced workload by adapting primary-task performance.

Brouwer et al. (1991) and Van Wolfelaar et al. (1990) have used an elegant 'driving-simulation' task. It was not the task environment that was elegant, but the way in which the level of primary-task performance was adapted to individual capability. By individually adapting the level of single task performance they succeeded in obtaining an equal task *difficulty* for all subjects. The primary task was a compensatory lane-tracking task. Added to this task was a visual analysis task. Van Wolfelaar (1990) added a third task to these, subjects had to respond to visual stimuli presented in the periphery. Although the simulator and the tasks that were used are more similar to laboratory tasks than to actual driving, the advantage of equal single-task difficulty for all is that divided attention problems can be studied taking into account differences in individual capability and allocation strategy. Results show that elderly are less successful in dividing attention in dual-task performance.

### **Properties of secondary-task measures**

If it is assumed that performance of a secondary task uses up 'spare capacity', then secondary-task measures could be performance measures that are sensitive in the A region. However, most secondary tasks interfere (to a varying extent) with primary-task performance and

task instruction alone cannot determine which task receives priority. Embedded tasks are regarded as the best secondary tasks. Even though it still is not certain what priority the embedded task receives, at least primary-task intrusion is low. In car driving, measurement of car-following performance and mirror checking can supply embedded task measures. Delay in car-following was found to be a sensitive measure in the sedative antihistamine, alcohol and car-phone conditions. Sensitivity of this measure can accordingly be expected in the D/A1 and A3/B regions of performance. Frequency of mirror checking was found to be sensitive in the Weaving Section study, while the measure also differed between motorway and ringroad-driving. This measure is sensitive in the A3/B regions, while the frequency was not affected in the antihistamine study, and sensitivity in the A1/D regions requires further examination. Duration of glances in the mirror was not sensitive in the Weaving Section study, and no conclusions with respect to sensitivity of this measure in regions of performance can be drawn.

Both delay in car-following and mirror checking can reflect performance at the manoeuvre level. Diagnosticity of the latter measure to visual demand is moderate to high (Fairclough et al., 1993). Delay in following a lead vehicle was found to be sensitive in car-following conditions in all studies and seems to be a sensitive and reliable measure. Mirror checking frequency showed a similar sensitivity in the motorway and ringroad conditions of the antihistamine and car-phone studies, and reliability is accordingly rated high. Primary-task intrusion when using embedded secondary tasks is low. However, when studying car-following behaviour and more or less natural variations from a lead car have to be followed, task priorities may become somewhat blurred. Primary-task intrusion and operator acceptance when registering mirror-checking behaviour depends upon measurement technique. The CEMRE is an intrusive device while video registrations made by small cameras can remain completely unnoticed by the subjects. Implementation requirements in terms of instrumentation and time/equipment required for analysis are high for all measures. An overview of secondary-task performance measures' properties is presented in table 9. Mirror checking measures are based on a limited number of studies, and were measured with different techniques.

Apart from quantification of task performance in measures such as the SDLP or the frequency of mirror scanning, task performance could be rated by an observer. This method is sometimes used, but suffers from other methodological problems, such as training of the experimenter. If applied correctly, and if observers are well trained, results could add to the previously discussed primary and secondary task-performance measures. Critical incidents, law violations and lateral position errors are measures of driving performance and have been used as such in task-performance assessment (e.g., Pohlmann & Traenkle, 1994). In particular, complex behaviour, such as the

occurrence of critical incidents, or behaviour in a complex driving environment can be easier, or more accurately, detected and judged by an observer than captured in a single performance measure.

Table 9. Summary of properties of secondary-task workload measures.

property	Measures		
	Delay in car-following	Mirror Checking Duration	Frequency
sensitivity (Region)	D-A1,A3-B	?	A3-B
diagnosticity	low - moderate	moderate (?)	mod.-high
selectivity	moderate (?)	?	moderate (?)
reliability	high	?	high
primary-task intrusion	low-moderate	low	low
implementation requirements	high	high	high
operator acceptance	high	high	high

## 5.4 Physiological measures

### Heart rate measures, ECG

Heart rate measures have been, and still are, very popular as in-vehicle registered physiological measures. The attractiveness of ECG is obvious, electrodes are easy to attach and distortion by physical movements is limited with car drivers, who simply have no other choice than to remain seated while driving.

Although heart period is measured and used as input for statistical analyses, the more popular ‘average heart rate’ during baseline and load condition is shown in figure 11. Note that the load condition is compared with a (similar) *baseline condition*, and not with the rest measurement. Compared to rest, driving (in both baseline and load condition) significantly elevates heart rate in all conditions. In the Noise Barrier experiment, average heart rate decreased in the load condition, while in the simulator (Tut) and antihistamine studies no effects of load compared with baseline measurements were found. An increase in heart rate (or a decrease in heart period or IBI) was found in both conditions of the Weaving Section study, and as a result of telephone use. On the adapted road leading through the woods (Wr), HR marginally significantly increased. Low amounts of alcohol increased HR on the motorway (Alc(mw)), an effect that is in accordance with findings of Mascord et al. (1995). The active drug Triprolidine did not affect heart rate frequency significantly, but average HR was prominently decreased as a result of time-on-task (the fatigue or ‘vigilance’ condition as it is indicated in appendix 5). These effects can even be better seen in figure 12, where the difference in

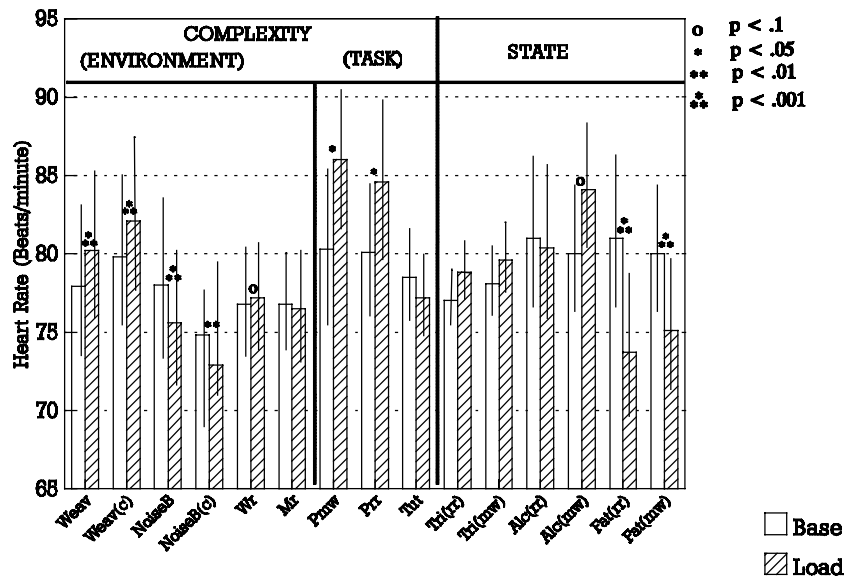


Figure 11. Average heart rate during baseline driving and during mental load. The 95% confidence interval is also indicated.

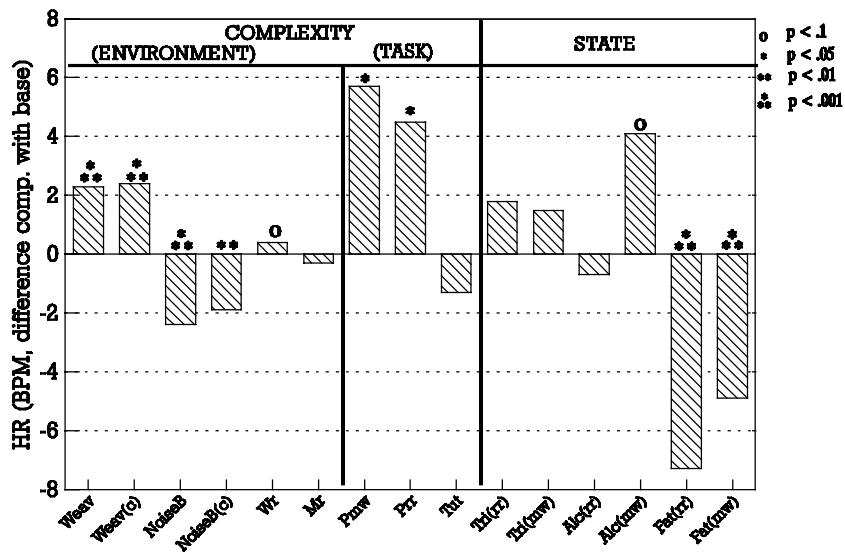


Figure 12. Difference in average heart rate during mental load compared with baseline driving.

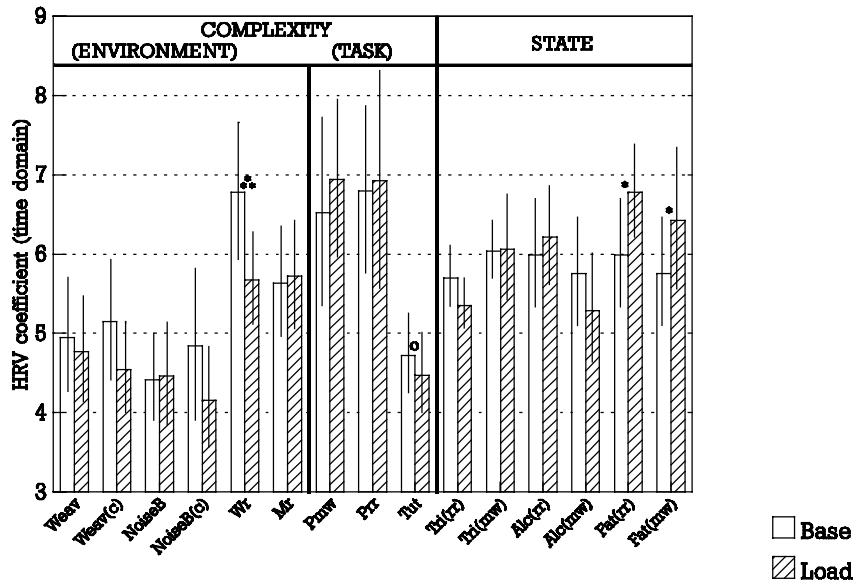


Figure 13. Standardized heart rate variability in the time domain during baseline driving and during mental load. The 95% confidence interval is also indicated.

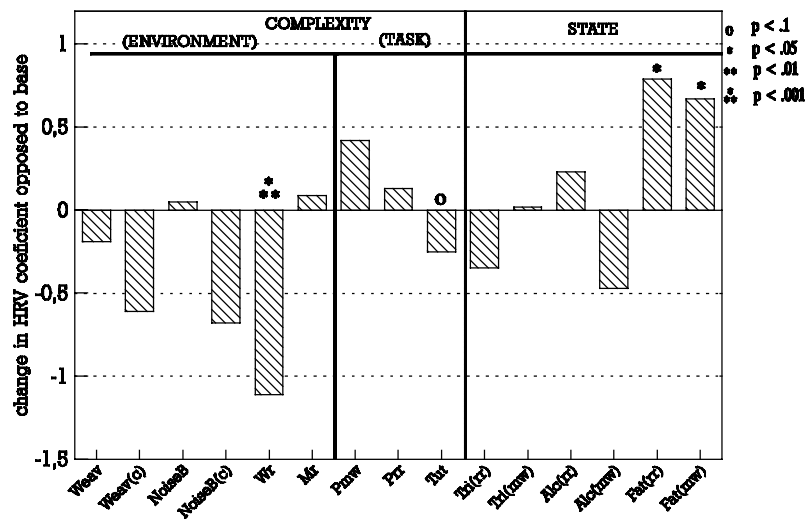


Figure 14. Difference in HRV during mental load compared with baseline driving.

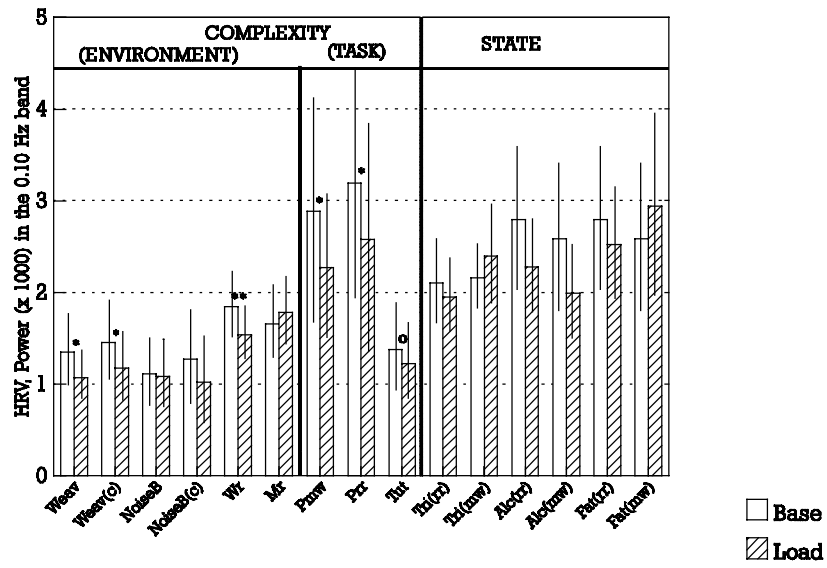


Figure 15. Energy in the 0.10 Hz frequency band of heart rate variability during baseline driving and during mental load. The 95% confidence interval is also indicated.

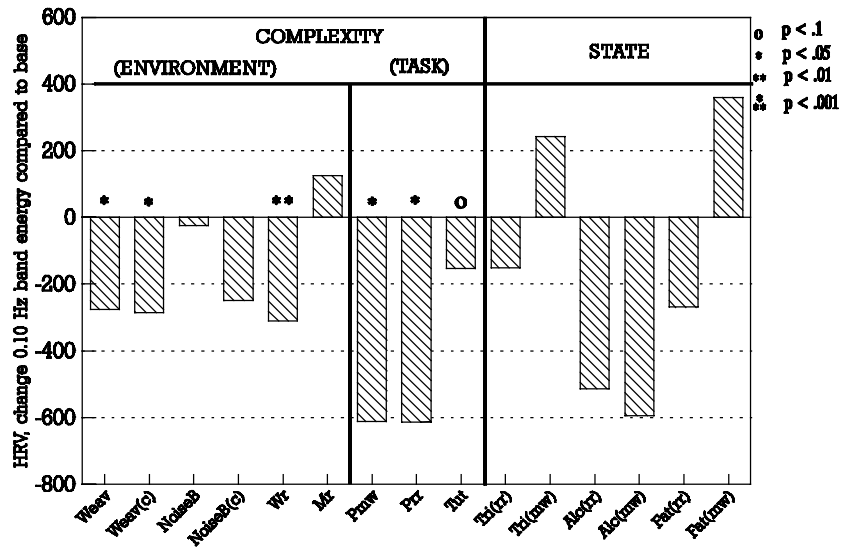


Figure 16. Difference in energy in the 0.10 Hz frequency band of heart rate variability during mental load compared with baseline driving

beats per minute of the load condition compared with the baseline condition are shown.

The Variation Coefficient (HRV), the standardized time-domain variability-measure of heart rate, is shown in figures 13 and 14. A significant decrease in variability was found in the DETER simulator study, and on the adapted Woodland road. A decrease was also found as a result of time-on-task; a finding in accord with Mascord & Heath (1992). The decrease in variability on the motorway as a result of an average Blood Alcohol Concentration of 0.5 ‰ was not statistically significant.

Compared to the time-domain variability measure, the frequency measure of 0.10 Hz variability is clearly more sensitive to the mental load manipulation (figures 15 and 16). Driving over the weaving section (Weav and Weav(c)), using the car-phone (Pmw and Prr), driving with feedback from the enforcement and tutoring system (Tut), as well as driving over the adapted road layout (Wr only) all reduced power in the 0.10 Hz variability band. The 0.10 Hz component-power is said to decrease as a result of relatively low levels of alcohol (see Gonzalez Gonzalez et al., 1992). The results regarding the Alcohol condition in the DREAM study (see figure 15) are in the expected direction, but not statistically significant.

Heart rate's idiosyncratic nature as well as high initial values can become very prominent in the spectral analysis and power computations that are required for determination of the 0.10 Hz HRV component. For this reason, energy in the 0.10 Hz frequency band is sometimes expressed as relative energy change compared with rest measurement (e.g., L.J.M.Mulder, 1988, Heino et al., 1996). For the studies in which a rest measurement was available, the additional change in 0.10 Hz HRV energy in the baseline and load conditions are displayed in table 10. In this table, the difference between baseline and load is also shown. Apart from 'size' differences, no large dissimilarities with figures 15 and 16 are apparent, with the exception of the Weaving Section and Noise Screen conditions in which the base-load difference is prominently reduced, or changes into a HRV increase in the non-CEMRE condition. The differences expressed as proportional change are, for reasons of lower inter-subject variability, probably more reliable than the absolute differences as shown in figure 16.

Heart rate profiles are a fairly recent development to monitor heart rate (variability) at a more continuous level. In the Weaving Section study (appendix 1), heart rate and 0.10 Hz-component heart rate variability were calculated and linked to specific road segments. Data chunks of 30 s were used as input and a resolution of 10 s was reached. With this technique, a more continuous index of the parameters can be obtained. In the Weaving Section study, changes in HR(V) during driving seem to reflect mental effort. Effects on parameters were tested



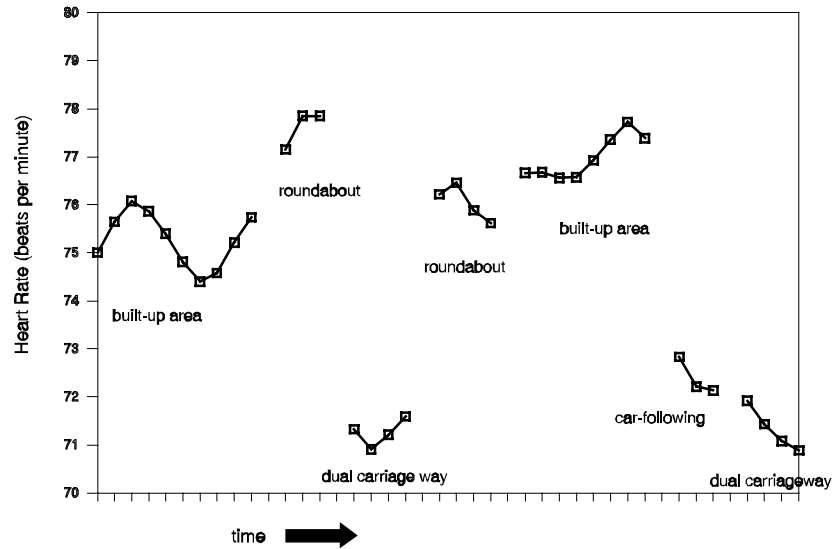


Figure 17. Average heart rate (N = 22) while driving in a simulator, during a trial in which subjects received feedback about detected law violations. Data were taken from the Tutoring (DETER) experiment (see appendix 4)

by comparing individual scores on an experimental section where load was suspected, with the scores on a section directly before this section (see appendix 1). The profile method was also applied in the simulator

Table 10. Change in energy in the 0.10 Hz frequency band of heart rate variability expressed as proportional change compared with rest measurements during baseline driving and during driving under mental load

study	Rest	Base	Load	Additional Change Load
<b>complexity (environment)</b>				
Weav	100%	-51%	-55%	- 4%
Weav(c)	100%	-30%	-53%	-23%
NoiseBarrier	100%	-18%	- 4%	+14%
NoiseBarrier(c)	100%	- 1%	-13%	-12%
Wr	100%	- 8%	-26%	-18%
Mr	100%	-19%	-11%	+ 8%
<b>complexity (task)</b>				
Pmw	100%	+13%	-14%	-27%
Prr	100%	+19%	-12%	-31%
Tut	100%	-13%	-23%	-10%

study and in figure 17 and 18 respectively, average heart rate and change in 0.10 Hz HRV energy are indicated; both averaged over 22 subjects. The third trial (see appendix 4) in which subjects received

feedback if violations were made, was selected for the figures. Thirty-second segments of data were used as input while the chosen step size again created a 10 second resolution. The different road environments are indicated in the figures. Clearly visible are the reductions in average heart rate frequency while driving over the dual carriageways and the increase in heart rate while driving around the roundabouts, and in the built-up areas. Figure 18 supports the idea that heart rate variability provides a reliable reflection of mental effort associated with different tasks. It can be seen that waiting for a red traffic light coincides with increases in variability, while driving on a roundabout corresponds to decreases in heart rate variability. The effects found in the simulator are very similar to effects found in an early on-the-road test of car driving, reported in Mulder (1980). Traffic density and traffic complexity were found to have a clear relation with reduced 0.10 Hz heart rate variability.

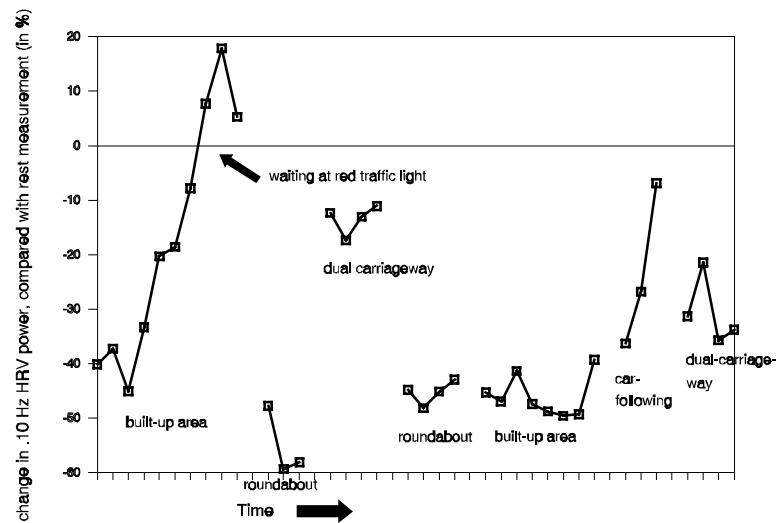


Figure 18. Change (in percentage) in 0.10 Hz HRV energy compared to the rest measurement. The same condition as in figure 17 was selected

### a combined statistical test

Again the effects found in the different experiments were tested in combination. The overall effect of complexity on heart rate is a significant decrease in IBI (an increase in HR). The driver state test is largely dominated by the effect of fatigue in HR. The total effect of reduced driver state is a reduced heart rate. Due to the direct effect of alcohol on heart rate this result has to be regarded with caution. The test on heart rate variability in the time domain (the variation coefficient) shows that HRV is reduced under increased (environment) complexity, but not as a result of increased complexity due to additional tasks.

---

*Inter-beat-intervals*

Complexity <sup>b</sup> :	$z = -2.73, p < 0.005$
Complexity (environment) <sup>c</sup> :	$z = +0.20, \text{NS}$
Complexity (task) <sup>b</sup> :	$z = -3.34, p < 0.0005$
State:	$z = +1.73, p < 0.05$

*Heart rate variability (Time domain)*

Complexity <sup>b</sup> :	$z = -1.95, p < 0.05$
Complexity (environment) <sup>c</sup> :	$z = -3.05, p < 0.005$
Complexity (task) <sup>b</sup> :	$z = +0.37, \text{NS}$
State:	$z = +0.65, \text{NS}$

*0.10 Hz component of heart rate variability*

Complexity <sup>b</sup> :	$z = -3.23, p < 0.0005$
Complexity (environment) <sup>c</sup> :	$z = -2.53, p < 0.01$
Complexity (task) <sup>b</sup> :	$z = -2.29, p < 0.025$
State:	$z = -1.16, \text{NS}$

<sup>b</sup> = Weighted,  $\alpha_{\text{pr}} = \alpha_{\text{pmw}} = 1, \alpha_{\text{Tut}} = 2$ .

<sup>c</sup> = All conditions equal weights

---

Only driver fatigue has a significant effect on HRV (increase), the total test of reduced driver state is not significant. Finally, spectral energy of heart rate variability in the 0.10 Hz frequency band is consistently and significantly reduced in the increased complexity conditions, but is not significantly affected in the test of the effect of a reduced driver state. This last aspect is very important and supports Mulder's idea (G. Mulder, 1995) that the 0.10 Hz component is sensitive to task-related effort and not to state-related effort.

**Other physiological measures used***ElectroEncephalo  
Gram (EEG)*

Ongoing EEG was more frequently used as indicator of driver state than as indicator of driver workload. The two are, however, not unrelated. As argued by some authors (Schneider et al., 1984, Kantowitz, 1992a), fatigue, e.g. as a result of the time spent performing a task, will be accompanied by a decreased arousal level and a reduced capacity, or a reduced willingness to spend resources (Meijman, 1991), and may therefore increase mental load. Ingestion of sedative drugs can be expected to result in the same effect. Brookhuis et al. (1985b, 1986) have found major increases in alpha and theta energy that were related to decreased driver activation caused by the use of antidepressant drugs. During prolonged train (Thorsvall & Åkerstedt, 1987) or truck driving (Kecklund & Åkerstedt, 1993), the driver's activation level as indicated by energy in the alpha and/or theta band was found to decrease rapidly. We (De Waard & Brookhuis, 1991a, appendix 5) have used the relative energy parameter  $[(\alpha + \theta) / \beta]$  as indicator of driver state and

found a significant increase on the parameter with time-on-task. When, after two hours of non-stop driving, subjects returned to a busy ringroad and had to follow a lead car, activation level increased again. Clearly, the increased task demands on the ringroad, increased mental load. It seems that EEG frequency analysis are most useful as an indicator of tonic driver activation, and can be included in workload research for these purposes.

*Electromyogram  
(EMG)*

Facial EMG of the corrugator supercilii muscle was measured in the road-layout experiment (appendix 2). An effect of driving vs. rest, and of the two different road environments was found, while no effect of mental load as a result of the experimental road-layout was found. As the 0.10 Hz component of heart rate variability was sensitive to the (expected) difference in workload between the experimental and control road, and EMG activity of the corrugator was not, it is suggested that these measures may be tapping different dimensions of task load (see appendix 2). To my knowledge, no experimental field-studies that further examine the differential sensitivity to workload of these two measures have as yet been performed.

*Eye movements*

The number and duration of eye fixations on instruments or in the mirrors while driving (see for mirror scanning also the section on secondary-task performance) may well be indicative for driver strategy. Rockwell (1988) found more glances instead of longer glances at a radio that had to be adjusted while driving. The strategy for most of the complex tasks was to take a series of glances of 1.25 s until the task was completed. Only if information could not be extracted in a glance, e.g. due to legibility, drivers could be tempted to increase glance duration. A minority of glances of up to 3 s were found when adjusting the stereo. Rockwell (1988) argues that glances of this duration are a threat to traffic safety, in particular in car-following situations.

In the Noise Screen and Weaving Section studies, fixation time (as proportion of the total looking time) was determined for various categories. Parkes (1991) refers to this measure as 'glance allocation'. In the studies initially eye movements were scored in various categories that were later combined into larger categories. Three categories were analyzed:

- traffic relevant fixations: looking straight forward, at other traffic, at the blind spot
- traffic irrelevant fixations: fixations on the other carriageway (which is irrelevant for motorway driving), the road environment, noise barriers, in the air
- mirrors & dashboard ('other points of focus')

The opportunity to look at, for driving, irrelevant stimuli will increase with decreases in workload (low time-pressure). This is partly

comparable to the path-neglect time in TLC (see under primary-task performance measures). A more demanding task *environment* requires an increase in the time spent looking at the road. In particular, the time spent looking at, for task performance relevant, objects, such as other traffic participants, road signs, road layout, etcetera, will increase. This includes looking in the mirrors. If it is not the road environment that requires additional attention but a device inside the car, it may have the opposite effect. Less time will be spent looking at relevant objects in the traffic environment.

In the Weaving Section study a reduction in time spent looking at the dashboard (speedometer) was found in the mental load condition (see table 11), while in the NoiseBarrier study, only data regarding the load condition were available. It is therefore difficult to draw conclusions on the basis of one study only. In addition to this, fixation time was scored in these analyses, and not fixation frequency, which is additionally required to assess driver strategy (Rockwell, 1988). In the Weaving Section study, fixation frequency on relevant objects increased in the load condition. While fixation time increases significantly with 6%, the number of fixations on traffic relevant objects is elevated with 13.2 fixations, an increase of 56%. Data from this study thus indicate larger sensitivity for fixation frequency compared with fixation duration expressed as proportion looking time. Scanning behaviour in which more glances instead of longer glances are taken (cf. Rockwell, 1988) could account for this difference in measure sensitivity.

Table 11: Proportion fixation time (%) and number of fixations per minute (fix/min) per category for the Weaving Section study (base and load) and the Noise Barrier study (load only). Eye movements were scored from video registrations made with the CEMRE equipment ('c' - condition only). Significant results have been printed in **bold**.

Study: condition: Fixation category	NoiseBarrier	Weaving Section		Weaving Section	
	load	base	load	base	load
	(%)	(%)	(%)	(fix/min)	(fix/min)
Relevant	76	72	<b>78</b>	23.4	<b>36.6</b>
Mirrors	3	10	11	6.8	<b>9.8</b>
Dashboard	7	6	<b>3</b>	4.8	4.7
Non-relevant	15	12	<b>8</b>	9.4	<b>6.9</b>

### Results from other studies

In other studies found in the literature similar effects of mental load on ECG measures are reported as were found in the studies listed in table 3. Zeier (1979) measured heart rate in heavy city traffic while subjects drove a car with manual transmission, a car with automatic transmission or were just passengers. Both average heart rate and HRV (time domain) differed significantly between the manual-transmission condition and the other two conditions. Driving with automatic

ECG

transmission or riding as a passenger did not lead to a significant difference in heart rate measures.

Egelund (1982) concluded that the 0.10 Hz component was an indicator of driver fatigue. Although average HR decreased with time-on-task, Egelund found that HR was, just as time-domain-HRV, not sensitive to fatigue in this study. Janssen & Gaillard (1985) concluded that the 0.10 Hz component of heart rate variability was a more sensitive measure in mental load assessment than the  $P_{300}$  amplitude in ERPs in their on-the-road study.

Fairclough et al. (1991) found an effect on HR of car-phone use. Average heart rate while performing a secondary task presented through a hands-free phone was found to be higher compared with the same task presented by an experimenter that accompanied the driver in the passenger seat. The authors give two possible explanations for the effect, either additional effort is required in the phone condition due to lack of cues in conversation, or unfamiliarity with cellular mobile phones aroused the subjects (cf. the practice effects found by Brookhuis et al., 1991). Van Winsum et al. (1989) found an effect of mental load on average HR and on the 0.10 Hz component of HRV. They found navigation based on a map to be more effortful than navigation by vocal messages, as measured by a decrease in power in the 0.10 Hz component band of HRV.

Janssen et al. (1994) did not find significant effects on the 0.10 Hz component in an on-the-road study in which a control group and two groups that received driver support were compared. The trend in the displayed figure, however, indicated decreased variability with driver support, a situation that could be comparable to the DETER Tutoring study. The authors suggested that the measure's insensitivity could be due to sensitivity to 'an averaged workload level'. If so-called heart rate (variability) profiles had been determined, a more detailed picture might have emerged in that study.

### *EMG*

One of the facial muscles that has been found to be sensitive to workload, is the frontalis (e.g., Van Boxtel & Jessurun, 1993). Zeier (1979) did not find an effect on EMG frontalis-activity of driving a car with automatic vs. manual gear transmission. However, he did find an effect of driving vs. being a passenger, the latter leading to lower muscle tension. The findings of Zeier (1979) support the idea that facial EMG activity taps a different dimension than (the 0.10 Hz component of) heart rate variability. Both in Zeier's study and the road layout study, EMG and HRV were differentially sensitive to workload. In addition, the two muscles that were measured, corrugator and frontalis, might also differ in selectivity. Jäncke (1994) found that the frontalis is not sensitive to emotional evaluation, while the corrugator is. A practical constraint of measurement of the corrugator in driving are the electrode positions that may interfere with the visual field.

### *ERPs*

Measurement of Event Related Potentials (ERPs) has mainly been restricted to laboratory experiments. An exception to this are the studies reported by Janssen & Gaillard (1985). In two studies subjects had to drive an instrumented car through three road environments: through the city, over rural primary-roads and over motorways. During these rides they had to perform a secondary, auditory, Sternberg task. EEG was measured and P<sub>300</sub> amplitude and its latency to task-relevant stimulus presentation (a secondary task) was determined. In the first experiment P<sub>300</sub> amplitude was decreased and latency increased as a result of task load. City driving caused the largest increase in latency, surprisingly followed by motorway driving. In addition, motorway driving decreased P<sub>300</sub> amplitude most, while amplitude was equally decreased during city and rural primary road driving, compared with rest measurements. In the second, similar study, city driving was left out. No effects on the P<sub>300</sub> were measured in this experiment. The authors report large individual differences and significant variance in the ERP data. They relate the remarkable position of motorway driving compared with the other conditions to the self-pacedness of the driving task. Complexity of the selected motorway section may, however, have had an effect on task demands (e.g., driving of a clover-leaf was included).

### *EDA*

In different studies Electrodermal Activity (EDA) has been related to the traffic environment (for an overview see Fairclough, 1993). Michaels (1962) reports an increase in EDR amplitude with an increase in traffic density, while Brown & Huffman (1972) report an increase in SCL if there is more traffic and there are more traffic lanes. Most in-vehicle studies have been performed in the sixties and focused on the effect of traffic environment on driver's EDA. In the seventies, Zeier (1979) measured EDA with electrodes positioned on the inner side of the left foot. He compared the effect of three conditions on psychophysiological measures, driving a car with manual transmission, with automatic transmission or being a passenger in a car. Effects on Skin Conductance level were not significant, but SCR (Skin Conductance Responses) were most numerous while driving the car with manual transmission. Least SCR were measured in the condition where subjects were passengers.

EDA is not only sensitive to all SNS activation, it might also be susceptible to physical movements. This last aspect is particularly relevant in car driving where EDA generally is measured on the palm of the hand, while both hands have to be used in steering. In mental workload research EDA might be useful to assess overall SNS activation level, but movements artifacts are a possible source of disturbance.

### *Hormones*

There are not many mental load studies that include the evaluation of hormone levels. In general, the measurement of hormone

levels is restricted to situations in which the driver's occupation is very demanding. Examples of this type of stress research are the studies regarding city-bus drivers (Mulders et al., 1988) and coach drivers (Raggatt & Morrisay, submitted). One exception to the long-term impact studies was found, in a study reported by Zeier (1979) examining the effects of driving in heavy city traffic were examined. Adrenaline levels were found to be higher when driving a car as opposed to being a passenger. In addition, driving with manual transmission also led to higher adrenaline levels than driving with automatic transmission. No differences were found on noradrenaline levels.

### **properties of physiological measures**

Background EEG is sensitive as an indicator of operator state, hence in region A to D. Average heart rate and heart rate variability in the time-domain are useful indicators of overall operator arousal level, i.e. in region D/B. The 0.10 Hz component however, is sensitive to task-related effort. It seems -as Mulder (1980) supposed- that the measure is sensitive to the Defense Response (Sokolov, 1963). The defense response is associated with a cardiovascular pattern of increased blood pressure, heart rate and stroke volume, decreased blood flow to renal, intestinal, and skin vascular beds, and increased skeletal muscle blood flow (Johnson & Anderson, 1990). The pattern is similar to responses evoked by stressful stimuli producing arousal in preparation for fighting. The defense response is coupled to increased sympathetic and reduced vagal activation, reflecting task-related effort and is accordingly connected to A3-region performance. Sensitivity of eye movements also seems to be highest in case of region A3 performance. Moreover, eye movements are related to visual demand, making it the highest diagnostic measure of table 12. Selectivity of EEG is low, operator state is reflected. The ECG measures differ in selectivity; HR and HRV are affected by many influences (respiration rate, physical effort) while this is less true for the 0.10 Hz component. Background EEG is a highly reliable, between-tests, measure for operator state, but individual differences (e.g., in the production of  $\alpha$ -waves) weaken this qualification. The many tests in which ECG measures were found to be sensitive to workload result in a reliability that is rated high. Primary-task intrusion when taking EEG and ECG measures is low once the electrodes have been attached. Measurement of eye movements may interfere with primary-task performance if cornea reflection is registered with the aid of a CEMRE. Intrusion is low if the driver's face is registered on video or in case of registration of EOG. Implementation requirements are high for most physiological measures, as special equipment such as sensitive amplifiers are required. For spectral analysis, for example, precise, i.e. 1 ms resolution R-top detection is required (L.J.M.Mulder, 1992). Special software is also needed. Only when average heart rate and HRV are determined, are implementation



requirements less stringent. Finally, operator acceptance is inversely related to intrusiveness of measure registration. In table 12 the properties of different physiological measures are summarized.

Table 12. Summary of properties of physiological workload measures.

<b>Measures</b>					
<b>property</b>	EEG back- ground	ECG HR	ECG HRV	ECG .10 Hz	Eye movements fixations/min
sensitivity (Region)	D-A2	D,B	D,B	A3	A3(?)
diagnosticity	low	low	low	low	high
selectivity	low	low	low-mod.	mod-high	?
reliability	mod-high	high	high	high	?
prim-task intrusion	low	low	low	low	<sup>1</sup>
implementation req.	high	moderate	moderate	high	high
operator acceptance	moderate	high	high	high	high-moderate <sup>1</sup>

<sup>1</sup> depends upon measurement technique

## 5.5 Discussion

Driving a vehicle is a task that demands continuous adaptation to a changing environment. A large part of the subtasks that have to be performed, such as lateral position control and speed maintenance, are tasks that are largely performed automatically at the control level, with hardly any driver effort. Representatives of performance measures at this level are the SDLP and steering wheel measures. At irregular intervals the control-level tasks are extended to include manoeuvre tasks, such as overtaking of other vehicles and following of leading cars. These tasks are not automated and require the driver's attention. Indicative measures of performance at this level are delay in car-following and the frequency of mirror checking.

A deteriorated driver state has been separated from increased task complexity as sources of increased workload. The effect of a deteriorated driver state and the increase in task complexity on primary-task performance might, however, appear to be the same. The primary-task parameter SDSTW changes in conditions of increased task complexity (e.g., Weaving Section study) and as a result of time-on-task. However, in combination with self-report ratings and physiology, a more differentiated picture emerges.

The pattern of measure sensitivity that emerges from the key studies (listed in table 3) is as follows: increased complexity, both in environment and in task, has an effect on the self-report scale RSME, and on the ECG. Task complexity vs. increases in environmental complexity seem to differentially affect the SDLP and SDSTW.

Additional tasks lead to a decrease in SDLP and SDSTW, while an increase in complexity of the environment increases both measures. An affected driver state resulting from the consumption of alcohol or sedative drugs does not affect heart rate variability as much as increases in complexity do. Time-on-task mainly affects the average heart rate level and the driver's EEG. Ratings on the self-report scale RSME and activation scale are more sensitive to changes in driver state. Secondary-task performance, in particular the embedded task of car-following, is sensitive to both sources of increased workload.

Region of performance remains a very important factor, as an increase in a primary-task parameter such as the SDLP can be the result of being overloaded as well as of driver deactivation. It seems that all deviations from optimal performance, both as a result of increased and decreased demand, can be traced by the combination of performance parameters and self-report and/or physiological indices. The moment task demands increase and the driver has to try harder, i.e. has to invest effort, heart rate variability in the 0.10 Hz band will decrease. The 0.10 Hz component is in particular sensitive to the defense response when task demands increase, and the driver exerts task-related effort. The changes on this parameter as a result of state-related effort and driver deactivation are less conclusive. Though the effects are large in terms of size, they fail to reach the 5% level of significance. Only Egelund (1982) reports significant changes on this parameter as a result of fatigue. The self-report scale RSME has more general sensitivity to driver effort, irrespective of whether it concerns state-related effort or task-related effort. It seems that these two measures, in combination with a primary-task performance measure, are the most useful to assess mental workload in the complete A region.

In most of the experiments listed in table 3, peak loads (Verwey & Veltman, 1995) play only a limited role. Workload during the car-phone conversation, while driving over the Weaving Section or over the adapted road layout; in all three conditions overall workload was increased. Only driving with the tutoring device could lead to peak loads at the moment messages are issued. However, on the basis of conversations with subjects after completion of the experiment it seems that the increase in mental workload in this experiment is more related to continuously intensified monitoring of the road environment and speedometer, than to information processing peaks at the moment of warnings. In sum, sensitivity of measures as reported above is sensitivity to overall workload, but no conclusions with respect to sensitivity to peak loads can be made on the basis of these experiments.

*Task inter-pretation, goal setting*

In the Car-phone study, Road layout and Tutoring experiments, an improvement on one of the primary-task measures, the SDLP, was found in the load condition. Since the effect of load in the three studies should be positioned in optimal performance section of the inverted-U, in the A3 region, no effect on primary-task parameters is expected. The

task environment may imply that higher performance is required and the improvement in performance may be the result of increased effort (as measured by a reduction in 0.10 Hz heart rate variability and an increased RSME score). In principle, the primary-task measure could therefore also be used for the assessment of workload in the A3 (and possibly also the A1) region. The best description of performance measures in these regions would then be 'no change or improvement in primary-task performance measures'. Finding an improvement in primary-task performance is paradoxical. Optimal performance is defined as the best performance, so no improvement is expected. In many laboratory tasks this is reasonable; in the field, however, conditions exist that allow for inaccuracies in primary-task performance during performance in the A-region. Unless subjects are given the strict instruction to drive in the centre of a lane and to try to steer as accurately as possible, improvement in primary-task performance can occur. A wide motorway lane, or the wide lanes used in the simulator experiment, do not necessitate accurate steering. *Goal setting or Task interpretation is an important factor and the need to perform at the highest level possible is in general absent in driving* and in field experiments. An improvement in performance was also found on the SDSTW-measure, in the load conditions of the Noise barrier and Tutoring studies. A similar explanation could be given for the improvement in lane-keeping performance, namely increased effort as indicated by physiological and self-report measures in both conditions results in increased primary-task performance.

Predicting the effects of tasks on driver mental workload is very difficult. Firstly, there are individual differences in goal setting and these differences vary from route choice to steering accuracy. Driving is to a large extent a self-paced task. If demands are too high, a slower driving speed can be chosen so as to be better able to deal with these demands. An elderly driver may prefer to make a detour so that he or she can drive over familiar roads thus facilitating the task environment. Once the task goals have been set, the task that has to be performed -the task demands- determine task complexity. How difficult a task is, however, depends upon capability (which may be lower for the elderly driver as just described), state and context. A novice driver will require more effort for vehicle control than an experienced driver. Driving performance itself can be related to externally set performance margins, critical levels, such as the margins proposed by Brookhuis (1995ab). Nevertheless only relative measures can give a further indication of mental workload. Strictly speaking, workload can only be determined per individual. It is always task X performed by individual Y (who is in a certain state) that leads to performance in Region Z. However, not all individuals are all that different and people often use similar strategies for performance of the same tasks. So, even though not all individuals set exactly the same goal, there are margins that are considered

acceptable. Heavy swerving and leaving the motorway lane is not considered acceptable by most drivers. Task demands can accordingly be defined in terms of maintaining the vehicle between the lines of the driving lane. For experienced young drivers it is not likely that there is much difference in (e.g. self-reported) effort required for the basic task of lateral and longitudinal vehicle control. This makes a link between a certain task and a region of task performance possible. In table 3 expectations about the region of performance for the different driving tasks have been specified.

Nevertheless, the most important factor in the measurement of workload is to assess changes in mental workload. Performance with the use of any device, in any environment or state under investigation, should be compared with baseline performance, driving without the use of the device, under 'normal' or standard conditions or while being sober. Changes in mental workload (measures) give a clear indication of what the effects of the changed demands are, incorporating at the same time changes in strategy or altered goals. This is, after all, the way people deal with changes in task demands in real life.

