Birgit Milius & Anja Naumann (Eds.):

Rail Human Factors

Proceedings of the 2nd German Workshop on Rail Human Factors

March, 8th and 9th, 2016
Stadthalle Braunschweig
Birgit Milius & Anja Naumann (Eds.):

**Rail Human Factors**

**Proceedings of the 2nd German Workshop on Rail Human Factors**

**March, 8th and 9th, 2016, Stadthalle Braunschweig**

**ISBN 978-3-937655-39-0**

In cooperation with

**ITS automotive nord e. V.**
Hermann-Blenk-Str. 17
38108 Braunschweig
Tel.: +49 531 231721-0
Fax: +49 531 231721-19
E-Mail: info@its-an.de
Internet: www.its-an.de

**Institut für Eisenbahnwesen und Verkehrssicherung**

**Deutsches Zentrum für Luft- und Raumfahrt**

German Aerospace Center
Birgit Millius & Anja Naumann (Eds.):
Rail Human Factors
Proceedings of the 2nd German Workshop
March, 8th and 9th, 2016, Stadthalle Braunschweig
Braunschweig: ITS automotive nord 2016
125 pages, 73 illustrations, 15 tables
ISBN 978-3-937655-39-0, € 89,00

Documentation:
Ergonomics and usability of advisory and assistance systems – Automation, train control
systems – Fatigue, monotony, mental workload – Risk and situation awareness – Human error
and human reliability – Incident investigation – Safety culture – Human factors integration (e.g.,
in organizational processes) – Rules and standards

The proceedings at hand include the contributions of the speakers of the Second German
Workshop on Rail Human Factors 2016. The workshop is organised by ITS automotive nord in
cooperation with the Institute of Railway Systems Engineering and Traffic Safety (IfEV),
Technical University of Braunschweig and the Institute of Transportation Systems, German
Aerospace Center.

© ITS automotive nord e. V. – Braunschweig 2016

All rights reserved. No parts of this publication may be reproduced, stored in a retrieval system or
transmitted in any form or by any means: electronic, magnetic tape, mechanical, photocopying,
recording or otherwise, without written permission from the copyright holders.

Printed in Germany ISBN 978-3-937655-39-0
Comparison of real-time relative workload measurements in rail signalers

Rob van Broekhoven¹, Aron Wolf (Willy) Siegel¹, Jan Maarten Schraagen¹,² and Matthijs L. Noordzij¹

¹University of Twente, Department of Cognitive Psychology and Ergonomics, Enschede, The Netherlands
[R.F.G.vanBroekhoven@student.utwente.nl, A.W.Siegel@, J.M.C.Schraagen@, M.L.Noordzij@utwente.nl]
²TNO Earth, Life, and Social Sciences, Soesterberg, The Netherlands
Jan_Maarten.Schraagen@tno.nl

Abstract
This exploratory field study investigated the weak resilience signals of workload in a rail traffic control room. The goals of this research are to see whether real-time system information of a rail control post can be used to predict workload of a rail signaler in real-time [1], and to further improve this method. In order to investigate this question, three workload measures were used. The first was the subjective Integrated Workload Scale, the second was a physiological measurement of electrodermal activity and the third was behavioural observation. For two cases the subjective workload was compared to the system information algorithm and the two other workload measures. The results show that the system information for communication, manual actions and switch cost are discriminating for workload.

Keywords: Workload, IWS, electrodermal activity, rail control room, rail signaler

Introduction
It is Wednesday 14:00 h. at a rail control room in Alkmaar, The Netherlands. A rail signaler throws a glance at his screen and makes some adjustments in the rail traffic planning. It is a calm shift and the train traffic runs as normal. At 16:23, a train driver calls in to report that the train has hit an object. The driver has stopped the train to check out what happened, as the procedure prescribes. As a response to the incoming call, the rail signaler notifies the decentralized traffic manager and rail signaler of the adjacent area of the stop location of the train. Next, he proceeds to inform the rail signaler of another adjacent post about the situation. Because the situation is rather unclear, the rail signaler calls all the approaching trains and orders them to stop. Each call includes exact prescribed actions and migration to avoid miscommunication. After seven minutes (16:30) the inspecting rail driver reports that he did not find anything that could explain the sound he heard and that there is no sign that the train has hit a person. Therefore, the rail signaler gives permission to drive again. The local co-workers, the decentralized traffic manager and the rail signaler from the other post are informed and the restrictions are cancelled. The rail signaler calls all related trains to abrogate the restrictions and informs them that they may start driving again. He requests to remain vigilant around the reported area. This case presents the effects of one train stopping for 7 minutes with the consequences of a workload increase for more than half an hour. Around 17:00, the last telephone call was conducted. While the events unfolded, the rail signaler had to monitor and act on different trains and events. The rail signaler was constantly switching between incoming calls from train drivers, informing co-workers, being updated by co-workers, anticipating on all new incoming trains in the area, manually rerouting these trains and informing all involved train drivers by telephone. This case describes a possible urgent and alarming situation where lots of different actions are necessary and a lot of different people need to communicate. Does the workload increase? Are rail signalers aware of their own (perceived) workload? Resilience engineering studies how socio-technical systems (STS) deal with unexpected and unforeseen circumstances, such as described in the case above [2]. Siegel and Schraagen [1] proposed that dealing with such circumstances in a resilient fashion requires STS’s to focus on so-called ‘weak resilience signals’. Weak resilience signals are signals that indicate an possible degradation of the STS without immediately triggering a predefined alarm. An example of a weak resilience signal could be a change in experienced workload that is not noticed or is not recognized as an alarming signal. For an organisation, both Madni and Jackson (2009) and Hollnagel (2009) state that the level of resilience is not merely a given factor, but an ability that can be developed to make the organization more flexible and proactive. Important factors in developing resilience are the ability and opportunity to anticipate, monitor, respond and learn from situations [4].
To improve resilience in a railroad setting, Siegel and Schraagen [1] developed a real time support system presenting weak resilience signals to increase the ability to respond to events and learn from them. Weak
resilience signals provide this ability without the need for escalation or accident. One of the weak resilience signals described by Siegel and Schraagen [1] is the relative increase or decrease of subjective and objective workload. Presenting workload weak resilience signals was done by presenting rail signallers with changes of their subjective workload and objective workload configured from their information system. Subjective workload was operationalised by a one-dimensional workload scale designed for rail signallers. This scale is called the Integrated Workload Scale (IWS; [5]). Objective workload was operationalised by means of an algorithm based on the model of cognitive task load (CTL[6]). The cognitive task load for a certain task is based on three dimensions: task complexity, task duration and task switching. The more complex, the longer the duration or the more switching between different tasks, the higher the cognitive task load. Siegel and Schraagen [1] also developed an algorithm, derived from log data of the traffic system, resulting in a measure called the external cognitive task load (XTL; Siegel and Schraagen [1]). The XTL is based upon four main measurable tasks of the rail signaller: monitoring, plan mutations, manual actions and communication by telephone. Because these measures are taken from system information, there could be a discrepancy between the behaviours that the system data would predict and the actual behaviour of the rail signaller. For example, an automatic mutation in the planning can change to something else, or can change back to the original planning without the rail signaller’s awareness. This will have no effect on the executed behaviours of the rail signaller, but the system will register activity. Therefore, this study will compare workload as measured by means of the XTL with other workload measures. In this way, this study investigates whether the results of the XTL correspond to other workload measures.

However, the literature is not consistent about the exact definition of workload (e.g., [7]). The problem is that there is no exact empirical definition and no physical unit to measure workload. Still, there is a whole range of methods, attempting to measure workload. Those methods generally focus on different facets of workload, such as self-report questionnaires (NASA-TLX; [8], heart rate variability [9] and EEG [10]). There is consensus, however, that at least three components are important for measuring workload. These components are subjective, physiological and performance measures [7]. Therefore, the current study will use three different ways of measuring workload, corresponding to these three components.

First, for the subjective measure, we used the Integrated Workload Scale (IWS; [5]). The IWS consists of a 9-point one-dimensional scale on which the rail signallers could indicate their perceived workload for a certain period of time. The IWS is specifically designed to measure rail signallers’ subjective workload and gives an insight in their perceived cognitive workload.

Second, for the physiological measure, we used electrodermal activity (EDA). Electrodermal activity is an online physiological measure of workload and there is consensus that it at least reflects a general measure of arousal or stress [11]. The EDA is expressed in skin conductance (SC) units [12]. In the EDA measurement, there are several parameters that can be extracted. Some parameters are related to the phasic, short lived Skin Conductance Responses (SCR), others are related to tonic, slow changes in the average level of the skin conductance level (SCL). The EDA measurement directly reflects activity of the sympathetic nervous system without being affected by parasympathetic activity [12] and it is a non-intrusive measurement that minimizes motion artefacts [13]. As the EDA measurement is not intrusive, the rail signallers will not be disrupted in their work.

Third, for the performance measure we used behavioural observation, enabling to make a good comparison with the XTL [1]. The behavioural observation will focus on the executed behaviours, forming the basis for a behavioural performance measure. We expect that certain behaviours correlate with the XTL measure.

The current research builds upon a previous study by Siegel and Schraagen [1] by adding behavioural observations and EDA. Those measurements will be compared with the algorithm used. Furthermore, these measurements can be used to further calibrate this algorithm. This exploratory field study attempted to answer two research questions. The first is whether the four workload measurements described (IWS, XTL, EDA and behavioural observations), support each other in the identification of changes in observed workload. The second question is whether the objective workload measure employed by Siegel and Schraagen [1] can be compared and complemented with the measurements used.

Methods

Participants & Procedures

The observations took place at a rail control post responsible for the area to the north of Amsterdam. The post was located in the city of Alkmaar. The rail control post consisted of four workstations (WS) with four rail
signallers on active duty, one backup rail signaler for calamities, one decentralized train traffic manager and one team supervisor. This observational study focused on one of the four work stations of the railroad control post. The 10 (9 male and 1 female) rail signallers that participated were between 22 and 52 years old ($M = 37.6; SD = 11.12$). The participating rail signallers had experience from half a year up to 34 years ($M = 11.8; SD = 11.01$). The protocol guiding the observations included an oral recorder consent, due to cultural constraints and at the specific request of the post management, and was approved by the ethics committee of the University of Twente. Before the observations started, the instructions and goals of the study were explained. When the participants were ready and everything was clear, they were asked to wear the EDA-sensor and were informed that the behavioural observations would start in a few minutes. The IWS measurement was running during the whole day and evening. The EDA measurement, as well as camera monitoring, was conducted during the day provided participants were willing to wear the EDA sensor and agreed to be recorded. In total, 34 hours of EDA measurement and 26 hours of behavioural observations were recorded. The camera monitoring was done to capture possible unique events and to look back for specific behaviours. Coding of observed behaviour was restricted to half an hour before the subjective IWS measure indicated “Some spare time” or higher. This was done for practical reasons in analysing all recorded material. During and between shifts, it was possible for the rail signallers to rotate positions. If this happened, the EDA-sensor was retrieved and data were extracted and logged before the EDA-sensor was passed on to the next rail signaler. Camera and behavioural observations continued, but a change of shift was marked in the video file. When the shift was coming to an end, the participants were asked for any remarks about the shift and were thanked for their participation.

Measurements

*Integrated Workload Scale (IWS)*

Siegel and Schraagen [1] (Figure 1) developed a tool based on the Dutch validated IWS [14]. The IWS would pop up for 30 seconds on the rail signaler’s work station. This was repeated every five minutes. In this way, we received an automatic and continuous rating of the IWS and a longer duration of a higher IWS will be referred to as a “stretch” of increased subjective workload as defined in Siegel and Schraagen [1]. To maintain a high response rate, it was possible for the rail signallers to open and fill in the IWS during the whole five minutes. It was also possible to adjust the last response they had given. This gave the rail signaler the opportunity to primarily focus on handling the situation, while still having the ability to fill out the IWS. If no response was given, the last value was copied under the assumption that there was no change of experienced workload.
External Cognitive Task Load

The External Cognitive Task Load (XTL) was calculated from real-time data retrieved from the operational control system [1]. The XTL was adjusted by adding a one to the formula used by Siegel and Schraagen [1]. This was done to achieve the same range as the IWS (from 1 to 9). The algorithm was based on the number of automatically executed plan rules in 5 minutes per workstation (monitoring, mon), the number of mutated plan rules in 5 minutes per workstation (plan mutations), the number of non-executed plan rules in 5 minutes per workstation (manual actions, man), and the percentage of seconds spoken through the telephone of 5 minutes per workstation (communications, com). The constant k’s were initialized with 1 and adjusted during post-processing, optimizing the relation with IWS: \( K_{\text{mon}} = 0.4, K_{\text{plan}} = 0.9, K_{\text{man}} = 1.2 \) and \( K_{\text{com}} = 1.5 \).

\[
\text{XTL}_{\text{WS}} = K_{\text{switchWS}} \left( K_{\text{mon}} \frac{\text{Mon}_{\text{WS}}}{\text{Mon}_{\text{max}}} + K_{\text{plan}} \frac{\text{Plan}_{\text{WS}}}{\text{Plan}_{\text{max}}} + K_{\text{man}} \frac{\text{Man}_{\text{WS}}}{\text{Man}_{\text{max}}} + K_{\text{com}} \cdot \text{Com}_{\text{WS}} \right) + 1
\]

1 ≤ XTL_{WS} ≤ 9

The XTL formula of Siegel and Schraagen [1] has been altered by adding a 1, causing the XTL values to be between 1 and 9, just like the values retrieved from the IWS.

The switch cost was taken into account by the number of activations that is composed from: 1) the number of delayed trains, 2) the number of telephone calls, and 3) the number of incidents reported in 5 minutes per workstation divided by the maximum number of activations in a 5 min time slot. The XTL gives a general relative cognitive task load configured from system output each five minutes. It will also provide a relative load of each of the four categories (monitoring, planning, manual and communication) which can be used to look at specific components in the XTL formula.

Behavioural cognitive Task Load

The Behavioural cognitive Task Load (BTL) was calculated in a similar way as the XTL. The BTL is based on the model of Neerinck [6] to be able to compare the variables of both measures with each other. The difference between the XTL and the BTL is that the information for the XTL comes from the ProRail information system, whereas the information from the BTL comes from executed behaviours of the rail signaler. Behaviours were selected based on observations, interviews with rail signallers and the four different categories of the XTL. Behaviours were observed using the Observer XT (version 11) and were rated based on how long (s) a behaviour was executed and how many switches occurred between different behaviours in five-minute time frames. This was done by observing for how long rail signallers showed behaviours that were linked to
observation, manual actions, planning behaviour, communication with team members and making telephone calls with others outside the rail traffic control post. These categories were further specified, taking into account different behaviours and implementation locations (Table 1; Figure 2). A differentiation was made, for example, between telephone calls originating from different parts of the socio-technical system. More specifically, a call from a bridge operator is likely to cause a low increased workload because the waterway bridge is manually controlled with one button. On the other hand, an incoming alarm call is more likely to increase workload because it needs immediate action.

Table 1: Overview categorized behaviours BTL

<table>
<thead>
<tr>
<th>Monitor</th>
<th>Planning</th>
<th>Manual</th>
<th>Communication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fast and global glance on the screens</td>
<td>Manual plan</td>
<td>Railway occupation screen</td>
<td>Local communication</td>
</tr>
<tr>
<td></td>
<td>screen</td>
<td></td>
<td>Communication through telephone</td>
</tr>
<tr>
<td>Railway occupation screen</td>
<td>Plan screen</td>
<td>Overview screen</td>
<td>Decentralized traffic regulator</td>
</tr>
<tr>
<td></td>
<td>monitor</td>
<td></td>
<td>Bridge</td>
</tr>
<tr>
<td>Overview screen</td>
<td></td>
<td>Writing report</td>
<td>Co-RS, specific case.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Co-RS-in other post</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td></td>
<td>Co-RS or other co-worker, general but work related.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Train driver</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Alarm</td>
</tr>
</tbody>
</table>

Figure 2: Screen one is the occupation screen, screen two is the planning screen, the three screens under number three are the overview screens

The formula of the BTL is based on the time(s) that behaviours in the categories (mon, plan, man, com) were observed. The constant k’s were initialized with 1.

\[
BTL_{ws} = B_{switchWS} \left( k_{mon} \cdot mon \left( s \right) + k_{plan} \cdot plan \left( s \right) + k_{man} \cdot man \left( s \right) + k_{com} \cdot com \left( s \right) \right)
\]

Again, the factor ‘switch cost’ from Neerincx (2003) was integrated. The switch cost for BTL was based on the number of switches in 5 minute intervals divided by the maximum observed number of changes in behaviour. The maximum behavioural switches observed during the study were 60 switches in 5 minutes.

Electrodermal activity
Rail signallers were asked to wear the Affectiva QTM sensor. This is a wrist worn, watch-like sensor that measures EDA with 1 cm diameter Ag-AgCl dry electrodes at the ventral side of the wrist. EDA data were pre-processed with a Continuous Decomposition Analysis (CDA) as implemented in Ledalab [15], which requires MATLAB (Mathworks, Natick, MA, USA). From the EDA, an estimate of the skin conductance level (SCL) as well as the overlaying phasic activity (occurrence and amplitude of SCR’s) can be acquired. The phasic activity, coming from classical Trough-to-Peak analysis, was reported (threshold for an SCR amplitude was set at .03 μS; Boucsein (2012)). As recommended by Boucsein (2012), visual checks were performed on plots of skin conductance data to identify failed measurements, “non-responding” (indicated by an absence of SCRs in a given measurement) and incorrect classification of SCR’s. Data from these problematic measurements were removed from further analysis. The SCL and SCR parameters were expressed in 5 minute intervals to allow for comparisons to the XTL and IWS the values.

Results

Data collection and case comparison

In order to compare the methods with each other, we took IWS as a baseline to make a distinction between low (IWS, 1-2) and high (IWS, 3-9) workload. We chose for IWS as a baseline because it has an uni-dimensional scale and because the IWS is used and validated for rail controllers [5,14]. However, occurrences of incidents or high workload were rare during the study. Therefore, behavioural observation was only further analysed around IWS elevations. During the observations, the IWS rose 14 times above “minimal effort (2)” and there was only one period of “very busy”(6)/”extremely busy”(7). If three of the IWS elevations the pattern showed a clear stretch in the IWS and the data collected from the other measurements were usable. Two of these cases (briefly describe below) contained sufficient data points for further statistical analyses.

Short Description of case 1 and 2

Case 1
A train driver is calling in thinking he hit a person and is going to check it. The train signaller informs co-workers and starts informing train drivers to stop the train or slow velocity as prescribed. After a few minutes the train driver reports he could not find anything and that it must have been something else. The train signaller informs co-workers and train drivers that they can start driving again. The short time it took to stop and get going again had over an half an hour delay involving all trains on the trajectory.

Case 2
The rail signaller is informed by mail that trains need to reduce velocity between a certain trajectory to a maximum of 40 km/h. Colleagues are informed and all approaching trains for this trajectory are called and informed according to procedures.

IWS results

The average IWS during the day of the main event was “minimal effort” with a small deviation (M = 2,06; SD = 1,13). The IWS pattern for the cases further analysed have a stretch lasting 10 or more minutes. The IWS pattern for the two cases are presented in Figure 3. In the other cases, the IWS scores did not become this high or the duration was too short. For further analyses, the IWS pattern will be used as a reference for the other methods and a distinction will be made between low IWS (0-2) and high IWS (3-9).
EDA results

The EDA data were visually inspected for any non-responders. All participants that seemed to provide usable EDA data were further analysed. Statistical analyses were conducted for the two cases using a MANOVA, comparing different EDA measurements (SCR, Amplitude, SCL) during the period of high IWS with the corresponding measurements during a period of low IWS. We found significant differences between periods of high and low IWS for all three measures in Case 1 (Figure 4, case 1). First of all, we found the SCR to be significantly different for periods of high IWS and for periods of low IWS (F(1,18) = 8.58, p < .009). The SCR signals occurred significantly more for periods with a high IWS (M = 87.4; SD = 18.96) than for periods with a low IWS (M = 63.60; SD = 17.68). The amplitude was significantly (F(1,18) = 8.59, p < .009) higher for periods with a high IWS (M = 27.73 μSiemens (μS); SD = 7.90) than for periods with a low IWS (M = 17.08 μS; SD = 8.33). Also the SCL was significantly different for periods of high compared to periods of low IWS (F(1,18) = 11.18, p < .004). Again, the SCL was significantly higher for periods of high IWS (M = 25.94 μS; SD = 5.37) than for periods with low IWS (M = 17.27 μS; SD = 6.19). These results show that the three EDA measures can discriminate between high and low IWS in Case 1.

For case 2, only the SCL was significantly different (Figure 4, case 2; F(1,19) = 1.66, p < .02) for periods of high IWS (M = 0.05 μS; SD = 0.08) compared to periods of low IWS (M = 0.65 μS; SD = 0.62). The results for the SCR and Amplitude were not significant for case 2. Moreover, the effect for SCL in case 2 is incongruent with the results of case 1. In case 2, the SCL is significantly higher for periods with low subjective workload.

BTL results

To corroborate the scoring system used, two of the researchers scored a sample of half an hour of observations. The inter-rater reliability between the two observers had 85% agreement (Cohen’s kappa = 0.82) in number of seconds per behaviour.
For the behavioural observation results, we performed a similar MANOVA comparing the four BTL categories, number of switches between behaviours and observed behaviours during periods of high IWS with the corresponding measurements during periods of low IWS. For case 1 the factor communication differs significantly between periods of high and low IWS ($F(1,18) = 4.74, p < .04$). When looking at the subcategories of communication (figure 5), we see that there is a significant difference for communication through telephone with a train driver ($F(1,18) = 10.70, p = .004$). This means that this behaviour occurs more during periods of high IWS ($M = 87.75$ s (out of 300); $SD = 75.38$) than during periods of low IWS ($M = 8.86$ s (out of 300); $SD = 11.48$). Also, the local communication with the decentralized traffic manager was significantly different ($F(1,18) = 4.54, p = .05$) during periods of high IWS ($M = 9.53$ s (out of 300); $SD = 13.52$) compared to periods of low IWS ($M = 0.38$ s (out of 300); $SD = 1.20$). This means that communication through the telephone with a train driver and local communication with a decentralized traffic manager are significantly higher in a high IWS situation than in a low IWS situation.

![BTL Case 1](image)

**Figure 5: BTL observed behaviours for case one for high and low IWS. Significant differences are indicated with (*).**

For case 2 the four BTL categories communication ($F(1,19) = 17.85, p < .001$), manual ($F(1,19) = 11.23, p < .003$), planning ($F(1,19) = 5.85, p < .05$) and monitoring ($F(1,19) = 21.70, p < .001$) were significantly different between high and low IWS. Also the number of switches between behaviours was significant ($F(1,19) = 36.73, p < .001$) with more switches in 5 minutes for high IWS ($M = 35.43; SD = 11.33$) than for low IWS ($M = 12.36; SD = 6.30$). On behavioural level (figure 6), communication through telephone with a train driver was significantly different ($F(1,19) = 10.36, p < .005$) for high IWS periods ($M = 63.7$s (out of 300); $SD = 76.10$) compared to low IWS periods ($M = 0.00$s (out of 300); $SD = 0.00$). Also local (case specific) communication with colleagues was significantly different ($F(1,19) = 8.08, p < .01$) with more communication in high IWS periods ($M = 16.22; SD = 21.94$) than in low IWS periods ($M = 0.00$s (out of 300); $SD = 0.00$). Also manually writing ($F(1,19) = 10.68, p < .004$) was significantly different for periods of high IWS ($M = 16.96; SD = 19.95$) compared to periods of low IWS ($M = 0.00$s (out of 300); $SD = 0.00$). Monitoring planning screen was also significant ($F(1,19) = 10.95, p < .004$), with more monitoring during high IWS periods ($M = 17.82$s (out of 300); $SD = 7.33$) compared to low IWS periods ($M = 7.27$s (out of 300); $SD = 6.68$). Also monitoring the overview screen was significantly higher ($F(1,19) = 19.59 p < .001$) during high IWS periods ($M = 38.89$s (out of 300); $SD = 18.05$) compared to low IWS periods ($M = 0.00$s (out of 300); $SD = 0.00$).
periods ($M = 8.73$ s (out of 300); $SD = 6.68$). In conclusion, these results show that some specific behaviours were able to discriminate between periods of high and low IWS.

![BTL Case 2](image)

**Figure 6:** BTL observed behaviours for case two for high and low IWS. Significant differences are indicated with (*).

**XTL results**

For the XTL data, we performed a MANOVA for low versus high IWS with five factors (Mon, Plan, Man, Com and Switchcost). For Case 1, the factor communication differed significantly ($F(1,18) = 11.20, p < .004$) between periods with high IWS ($M = 0.25; SD = 0.20$) and periods with low IWS ($M = 0.03; SD = 0.06$). This means that for periods with high IWS.

For Case 2, the factor communication was significantly different ($F(1,19) = 8.58, p < .009$), being higher during periods of high IWS ($M = 0.25; SD = 0.33$), compared to periods of low IWS ($M = 0.00; SD = 0.00$). Manual was also significantly different ($F(1,19) = 9.00, p < .006$), with more manual data in high IWS ($M = 0.10; SD = .13$) compared to low IWS ($M = 0.00; SD = 0.00$). Switch was also significantly different ($F(1,19) = 5.09, p < .04$) with more switches during periods of high IWS ($M = 0.07; SD = 0.06$) than during periods of low IWS ($M = 0.01; SD = 0.04$). In conclusion, these results show that communication, manual and switch cost discriminated between high and low IWS.
**Discussion**

This study investigated whether the four workload measurements (IWS, XTL, EDA and BTL) supported each other in the identification of changes in observed workload, and whether the XTL algorithm can be confirmed and complemented.

The results show that EDA is a good discriminator between high and low IWS values in case 1, which is in line with the consensus that electrodermal activity is an online physiological measure of workload that at least reflects a general measure of arousal or stress [11]. This effect was not found in case 2 except for the SCL which was opposite but small. This discrepancy in case 2 could be explained by the smaller change of IWS, which did not pass the physiological arousal or stress threshold. The EDA measurement results show that the EDA seems to be a promising method to use in measuring workload in rail signallers. The method is not intrusive with their work and it is theoretically possible to process the data in real time (although this was not the case in this research). However, the current experiment was relatively short and the observed periods were relatively calm, so further research is necessary.

BTL shows that different behaviours occur with high versus low IWS. Mainly the category “communication” seems to be important. Looking at different behaviours, “telephone communication with train driver” and “contact with the decentralized” came back in both cases. When compared with the XTL, these effects reoccurred partially. The effects on the XTL, however, are less pronounced. The reason for this might be that, in the XTL, no distinction was made between with whom the telephone communication took place. This information seems to be important for interpreting these results, considering the BTL data. Also the factors “manual” and “switch cost” seem to differentiate between high and low workload. The BTL observations show that there is a high correlation between subjective and behavioural patterns, but that this highly depends on the behaviour in combination with other factors. For example, the same behaviours (communication/calling on the telephone) can have a different impact on experienced workload if the context or communicating partner is different. For the XTL, it would be desirable to make a differentiation for different categories or interactions in the socio-technical system. For example, calling with a train driver has a greater correlation on experienced workload that calling with a bridge operator.

Finally, The XTL formula in the investigated cases shows a differentiating ability in both communication and manual actions. This shows that the XTL and in particular the parameters “communication” and “manual” could differentiate between high and low workload. However for manual action the effects are not congruent and should be examined in perspective of more cases. Also Switch cost seems to show a trend (although not significant in case 1). For case 1 this could be explained by lag of the IWS/XTL. If the last high IWS for case 1 is removed the XLT switch cost is also significant. The XTL could be further improved in further research by differentiating the input data for XTL for the different categories. In this way, for example, a distinction could be made through the other party in a telephone call. These steps will make the XTL more sensitive and will create a better match between performance and experienced workload.

Overall, the current research shows that real-time observation of subjective measures using IWS and XTL can be done and are corroborated by EDA and behavioural observation. IWS, XTL, EDA and BTL are capable of...
making distinctions between high and low experienced workload. Further research and specifications are necessary to determine and validate which of the system’s data have a high predictive validity and which do not. The current research contributes to a better understanding of measuring workload of rail signallers by showing that system information can be used to give a relative indication of the workload of the rail signaler.

References


