SUBJECTIVE MENTAL WORKLOAD OF DUCTH TRAIN DISPATCHERS:
VALIDATION OF IWS IN A PRACTICAL SETTING

Wilms, M.S., MSc. Eur.Erg.[1] & Zeilstra, M.P., MSc. Eur.Erg.[1]

Intergo, consultancy in human factors and ergonomics, P.O. Box 19218
NL-3501 DE, Utrecht, Holland, E-mail: wilms@intergo.nl

This study shows the validation of IWS in a practical setting and compares the results of the subjective workload observation method with objective workload measures. This study showed that IWS is sensitive to several environmental and task-related factors. IWS is a useful tool in assessing mental workload in the real-world setting, because of its sensitivity to fluctuations of mental workload related to several environmental, task and personal variables and its sensitivity to underload. No significant correlation was found between IWS and objective workload.

Introduction

Nowadays, human factors in railway operations is receiving much more attention than in the past. The rail network comprises a complex system faced with several competing pressures. On the one hand the railway network is expanding, the amount of passengers increase, companies wish to run more trains etc. On the other hand there is the demand for a reliable, high quality service that guarantees the safety of staff and passengers. One of the key players in railway operations are train dispatchers. Their primary task is to control the direction and order of traffic in a situation characterized by a limited infrastructure, timetables and organizational rules. Delays, conflicts and accidents need to be avoided and, when present, resolved quickly, efficiently and safely. Train dispatchers making errors can have serious consequences, possibly resulting in delays or even dangerous situations for passengers and railway employees.

An important area of human factors is mental workload. Mental workload is considered to be an important factor in individual performance within complex systems and both mental underload as well as mental overload is associated with decreased performance (Lysaght et al., 1989), increased amount of errors (Desmond & Hoyes, 1996) and decreased efficiency of the entire system. The

Dutch rail network manager ProRail, responsible for infrastructure management, acknowledges the importance of optimal levels of workload for its train dispatchers in order to ensure high quality performance and systems safety.

ProRail has 3 instruments to investigate workload of train dispatchers. Subjective mental workload is measured using the Integrated Workload Scale (IWS). The objective amount of workload is measured using Task Weighing™. The third instrument measures the work environment (Figure 1).
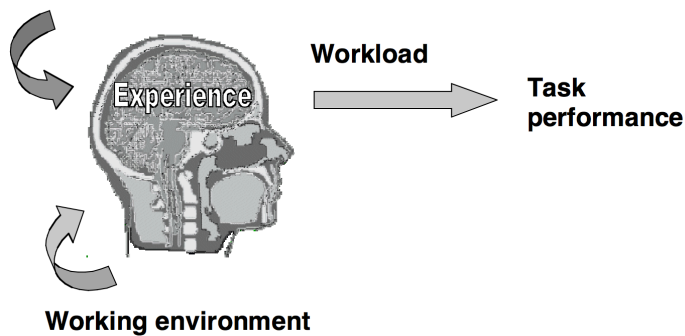
Task demands

Figure 1 The model of workload (based on ISO 10075, 1991)

The main goal of this study is the validation of IWS in a practical setting. In laboratory studies IWS has shown sensitivity to fluctuations in mental workload related to changes in the environment, such as increased number of trains or unforeseen delays and disturbances (Pickup et.al., 2005). Another goal of this study is to compare results from the IWS study with results from TaskWeighing™. In a previous study results from IWS and TaskWeighing™ seemed to be in agreement, but were not tested statistically (Zeilstra, Bruijn and Van der Weide, 2009). Investigating a potential correlation between the IWS and TaskWeighing™ provides further inside in how different methods may interact. This could influence future development and implementation of tools for measuring mental workload.

Methods

In this study 394 hours of observations of train dispatchers spread over 26 workstations in the Netherlands were done; 336 observations involved male participants with a mean age of 46.3 years (SD = 7.4) and 58 observations involved female participants with a mean age of 41.1 years (SD = 5.7). The mean number of years experience was 14.9 (SD = 8.7) for males and 10.2 (SD = 5.6) for females. Participation was on a voluntary basis. All statistical tests were performed with Statistical Program SPSS.

IWS

IWS is a self-report tool and consists of a 9-point scale (1 = not demanding, 9 = work too demanding). While working train dispatchers verbally reported their subjective level of workload to the researcher at 5-minute intervals during a one-hour session. During each 5-minute interval the researcher registered the time spend on 7 predefined activities: monitoring of trains on the computer screen, communication by phone, communication with the planner, communication with colleagues, reading and writing, computer interaction and a category of other activities (not work related activities). Also, several environmental (perturbations, workstation type and time period) and personal (education, familiarity with the workplace, age, number of years experience and gender) variables were registered.

TaskWeighing™

TaskWeighing™ is a tool to evaluate and predict workload based on task demands. Several mathematical formulas were developed and used to calculate mental workload per task based on task demands of train dispatchers and the relative load of the task. This results in a number of workload points, without unit, per task. Mental workload was then calculated by adding the number of workload points together for all tasks carried out in one hour. TaskWeighing™ has been used to calculate workload during rush hour without significant delays and/or disturbances and a disturbed rush hour over 26 work stations.

Absolute and relative frequencies were calculated from all IWS scores. Mann-Whitney U tests and Kruskal-Wallis tests were performed to examine effects of environmental and personal variables on mental workload scores. Spearman's correlation coefficient was used to correlate the average time spend per activity and the number of train movements with IWS scores. Mental workload scores were determined by taking the median IWS score per train dispatcher.

Effects of environmental and personal variables on underload and acceptable workload levels were assessed with use of the Chi-Square test. Two categories were made. The 'underload category' consisted of train dispatchers with a median IWS score of 1 and the acceptable workload category consisted of train dispatchers with a median IWS score higher than 1 and lower than 9. Overload, defined as a median IWS score of 9, did not occur frequently enough to be included in this analysis.

Effects of environmental and personal variables on time spend on activities was examined by separate ANOVAs.

Results of IWS and TaskWeighing™ were compared by determining the median IWS score per work station and correlating this score with the number of workload points calculated by TaskWeighing™. Spearman's correlation coefficient was used. Enough data was available to compare the IWS and TaskWeighing™ during undisturbed rush hours, but not enough data was present for disturbed rush hours.

Results

IWS scores

The total number of IWS scores and their relative frequencies are displayed in Figure 2.

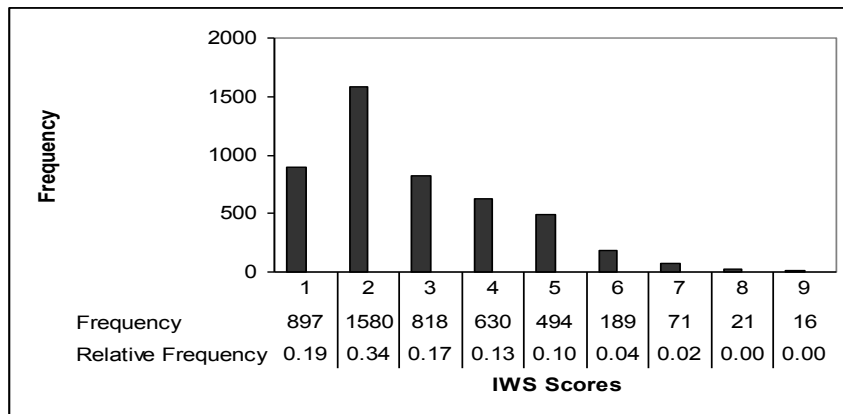| IWS Scores	1	2	3	4	5	6	7	8	9
Frequency	897	1580	818	630	494	189	71	21	16
Relative Frequency	0.19	0.34	0.17	0.13	0.10	0.04	0.02	0.00	0.00

Figure 2 Absolute and relative frequencies of all IWS scores (N = 4716)

IWS scores and Environmental variables

Three Kruskal-Wallis tests were done with Perturbations, Workstation Type or Time Period as independent variable and median IWS scores per train dispatcher as dependent variable. IWS scores were significantly affected by perturbations, $H(2) = 61.5$, $p < .001$ and workstation type, $H(3) = 14.24$, $p < .01$, but were not significantly affected by time period, $H(2) = 2.36$, $p = .31$. Post-hoc Mann-Whitney U tests were performed for workplace and perturbations. After inspection of the data three Mann-Whitney U tests were performed for each variable, because no significant differences were expected between nodes, busy nodes and very busy nodes. A Bonferroni correction was applied so all effects are tested at a .0167 level of significance. The results are presented in Table 1.

Table 1: Results of the post-hoc Mann-Whitney U tests.

Work-related variables	U	Sig.	r
No perturbations – Maintenance	5716	.01*	-.14
No perturbations – Disturbances	3360.5	.00**	-.41
Maintenance – Disturbances	804.5	.00**	-.42
Railway line – Node	602.5	.00**	-.24
Railway line – Busy node	484	.00**	-.25
Railway line – Very busy node	204	.00**	-.37

$* = p < .0167$, $** = p < .01$. r = effect size, 0.1 = small, 0.3 = medium, 0.5 = large

IWS scores were significantly lower for shifts with no perturbations and were lower for shifts with maintenance compared to disturbances. IWS scores were significantly lower for railway lines compared to nodes, busy nodes and very busy nodes.

The number of train movements was correlated with IWS scores using Spearman's correlation coefficient. A significant correlation was found, $Rs = .16$, $p < .01$. However, a small but distinct group (N = 15) was identified with a much higher number of train movements (+/- 400 per hour) compared to the rest (max +/- 200). After removal of this group the analysis was repeated, but no significant correlation was found, $Rs = .08$, $p = .15$.

IWS scores and Personal variables
Age and experience in years did not show a significant correlation with IWS scores, $Rs = -.14$, $p = .11$ and $Rs = -.03$, $p = .58$, respectively. Also, IWS scores were not significantly affected by gender, $U = 9375$, $p = .66$ or familiarity with the workplace, $H(2) = 3.7$, $p < .16$, but Education did have a significant effect on IWS scores, $H(2) = 6.6$, $p < .05$. After inspection of the data two Mann-Whitney U tests were performed, because no significant difference in IWS scores between medium and high education was expected. A Bonferroni correction was applied and so all effects are reported at a .025 level of significance. IWS scores were significantly lower for train dispatchers with a low compared to a medium education, $U = 1054.5$, $p < .01$, $r = -.22$, but no significant difference in IWS scores was found between low and high educations, $U = 64$, $p = .17$, $r = -.22$.

IWS scores and Task-related variables
To test for correlations between tasks and IWS scores, median IWS scores and the average time spend for each task were correlated using Spearman's correlation coefficient. Results are presented in Table 2.

Table 2: Correlation between work-related activities and IWS scores.

Work related activities	Spearman's rho	R^2
Monitoring	.22***	.02
Communication by phone	.51***	.28
Communication with planner	-.01	.00
Communication with colleagues	.27***	.06
Reading and writing	.33***	.07
Computer interaction	.45***	.20
Other activities	-.58***	.32

* = $p < .05$; ** = $p < .01$, *** = $p < .001$.

Underload and Environmenal and Personal Variables

Due to the high relative frequency of underload (19%), effects of environmental and personal variables on the frequency of underload were examined. Chi-Square tests were conducted for each variable. A significant association between Perturbations and underload was found, $\chi^2(2) = 24.15$, $p < .001$. This association represents the fact that based on the odds ratios underload is 3.41 times more likely when no perturbations are present compared to maintenance and 16.78 times more likely compared to disturbances. Also, underload is 4.92 times more likely when there is maintenance compared to disturbances. Workstation type also showed a significant association with underload, $\chi^2(3) = 18.36$, $p < .001$. This association represents the fact that based on the odds ratios underload is 9.96 times more likely at a railway line compared to a node, 8.45 times more likely compared to a busy node and 14.66 times more likely compared to a very busy node. Odd ratios between nodes, busy nodes and very busy nodes were small and varied between .6 and .9.

Work-related variables and activities

The effects of environmental variables on the amount of time spend on activities was examined by conducting ANOVAs. A Bonferroni correction was applied and all effects are reported at a .0083 level of significance. Results showed that Perturbations and Workstation Type have significant effects on time spend on activities (see Table 3). Communication with the planner was not significantly affected by both variables. Other work-related and personal variables did not show significant effects on the time spend on activities.

Table 3: ANOVA test results with Perturbations and Workspace as independent and time spend on an activity as dependent variable.

Activities	Perturbations		Workstation Type	
	Welch's F	Sig.	Welch's F	Sig.
Monitoring	3.88	.023	103.19	.00***
Communication by phone	26.67	.00***	102.97	.00***
Communication with colleagues	18.8	.00***	12.38	.00***
Reading and Writing	21.38	.00***	47.29	.00***
Computer Interaction	13.71	.00***	337.32	.00***
Other activities	30.4	.00***	183.78	.00***

* = $p < .05$; ** = $p < .01$, *** = $p < .001$.

IWS and TaskWeighing[TM]

IWS and TaskWeighing[TM] scores were correlated using Spearman's Correlation. Results showed a non-significant correlation, $Rs = .27$, $p = .23$, $R^2 = .05$.

Discussion

In this study mental workload of train dispatchers working for the Dutch rail network manager ProRail was examined in a practical setting. Although the IWS proved its usefulness in simulators (Pickup et al., 2005) due to its ability to detect fluctuations in mental workload related to environmental and task related factors, it had not yet been implemented in a practical setting. Goals of this study was the validation of IWS in a practical setting and also, to compare results from the IWS study with results from TaskWeighing[TM] to investigate the level of agreement between these different tools.

Environmental factors and mental workload

Results showed that perturbations and workstation type had a significant effect on mental workload. Also, mental workload was lower at a workstation type characterized by relatively simple infrastructure, but did not differ between workstations with different levels of higher infrastructure complexity. Thus, based on these results subjective mental workload is affected by deviations from the usual situation and by infrastructure complexity, but only when complexity is very low. There were no significant differences in IWS scores during rush hours, off peak hours and night hours and no significant correlation between mental workload scores and the number of train movements was found. In the stimulations performed by Pickup and colleagues (2005) fluctuations in mental workload related to lower and higher levels of traffic were observed, although not tested statistically. These observations are not consistent with findings in our study in which traffic levels do not seem to affect mental workload. More research is required to provide an explanation for this discrepancy.

Personal factors and subjective mental workload

Young and Stanton (2001) defined mental workload as the level of attentional resources required to meet objective and subjective performance criteria that may be mediated by task demands, external support and past experience. With this definition in mind, it could be hypothesized to find correlations between age and experience and mental workload scores. Also, workload scores may be affected by familiarity with the workplace. However, no such correlations or effects were found. Experience, age and familiarity with the workplace may influence mental workload in situations characterized by serious disturbances and/or accidents. However, these events did not occur frequently enough during the IWS observations to be analyzed. More research, perhaps using simulations, is required to investigate the existence of such potential influences. The personal factor education did show an effect on mental workload with lower mental workload scores for train dispatchers with low compared to medium educations. Although no difference was found between low and high educations, the sample size of the high-educated group was small (N = 6) and we hypothesize to find significant differences when the sample size is increased. Based on this study it is difficult to provide an explanation for this result. Time spend on activities did not differ between the different levels of education. Perhaps factors such as differences in task performance, self-reporting and/or other factors between people with lower compared to higher levels of education played a role, but these factors were not measured and further research is thus required.

Activities and mental workload

Several activities correlated significantly with workload scores. Increased time spend on monitoring, communication by phone, communication with colleagues, reading and writing and computer interaction correlated positively with mental workload. Time spend on the category other activities correlated negatively with IWS scores. The strengths of the correlations differed and computer interaction and communication by phone showed the strongest positive correlations. This differentiation in the strength of the correlations between activities may be explained by the amount of resources required. Based on an attentional demand model of mental workload (e.g. Young and Stanton, 2001) mental workload is influenced by the amount of attentional resources required to carry out a task. Computer interaction and communication by phone are most likely to require more resources compared to more passive activities such as monitoring or reading and writing and thus show stronger positive correlations with mental workload scores. The category other activities consist of activities unrelated to work and may require less attentional resources and could even be considered relaxing. This may explain the relatively strong negative correlation with experienced mental workload.

Underload

Due to the relative high frequency of underload, defined by an IWS score of 1, and the relevance of investigating underload, because of its relation with decreased performance (Brookhuis, 1993; Hancock & Caird, 1993), associations between environmental and personal factors on underload were examined.

Perturbations and workstation type showed a significant association with mental underload, represented by the fact that underload is several times more frequent when no perturbations occurred and when infrastructure complexity is very low.

Interaction Environment, Personal and Task-related variables
The effect of environmental and personal factors on time spend on activities was examined in order to create a more integrated picture of mental workload. Perturbations and workstation type showed significant effects on time spend on several activities. In the absence of perturbations and in situations with low complexity infrastructure, significant less time is spend on work-related activities while more time is spend on the category other activities. It seems therefore plausible that the effects of perturbations and workstation type on mental workload can, at least partly, be explained by differences in time spend on activities.

IWS and TaskWeighing™
Another goal of this study was to correlate results of IWS and TaskWeighing™. Based on Megaw's (2005) use of Hill's (1987) framework, measures of mental workload can be divided in two categories, analytical and empirical. The IWS is an empirical method, because it is a subjective tool designed to measure the experienced mental workload. TaskWeighing™ is an analytical tool developed to evaluate and predict mental workload based on task demands. Due to these differences the IWS and TaskWeighing™ are sensitive to different aspects of mental workload. However, it could be hypothesized that workload based on task demands is related to the amount of experienced workload. In this study, no significant correlation was found. The absence of a significant correlation may be explained by considering the model of mental workload adopted in this study (Figure 1). In this model experienced mental workload is influenced by both task demands and work environment. Perhaps work environment can protect against high levels of experienced workload, even when task demands are relatively high. It is important to note that some methodological factors, such as a relative small sample size and little variation in IWS scores on a workstation level, may have played a role in the lack of finding a significant correlation. It can be concluded that more research is required and by no means does the result of this study indicate that empirical and analytical methods can't be in agreement with each other. However, the absence of a significant correlation does stress the fact that different tools for measuring mental workload can't substitute for one another and that one should be very cautious when comparing and interpreting results from different tools for measuring workload.

Conclusion

IWS has proved to be sensitive to both several environmental and task-related factors. However, no effects of traffic levels, represented by rush hours, off peak hours and night hours and the number of train movements, on IWS scores were found. An integrated picture emerged in which environmental and task-related factors and mental workload interact. IWS also proved to be sensitive to

underload and factors that affect underload. This is important, because underload may be more harmful than mental overload, since it may be more difficult to detect (Hancock & Parasuraman, 1992; Hancock & Verwey, 1997). It is concluded that the IWS is a useful tool in assessing mental workload in the real-world setting, because of its sensitivity to fluctuations of mental workload related to several environmental, task and personal variables and its sensitivity to underload.

References

Brookhuis, K. A. 1993, The use of physiological measures to validate driver monitoring. In A. M. Parkes and Franzen, S. (Eds.), Driving future vehicles (pp. 365-376), (Taylor and Francis, London).

Desmond, P. A. and Hoyes, T. W. 1996, Workload variation, intrinsic risk and utility in a simulated air traffic control task: evidence for compensatory effects, *Safety Science*, 22(1-3), 87-101.

Hancock, P. A. and Parasuraman, R. 1992, Human factors and safety in the design of Intelligent Vehicle-Highway Systems (IVHS*), Journal of Safety Research*, 23(4), 181-198.

Hancock, P. A. and Caird, J. K. 1993, Experimental evaluation of a model of mental workload, *Human Factors*, 35, 412-429.

Hancock, P. A. and Verwey, W. B. 1997, Fatigue, workload and adaptive driver systems, *Accident Analysis and Prevention*, 29(4), 495-506.

Hill, S.G. 1987, Analytic techniques for the assessment of operator workload. In: Proceedings of the 31st Annual Meeting of the Human Factors Society, HFS, Santa Monica, CA, pp. 368−372.

Lysaght, R. J., Hill, S. G., Dick, A. O., Plamondon, B. D., Linton, P. M., Wierwille, W. W., Zaklad, A. L., Bittner, A. C., Jr. and Wherry, R. J., Jr. 1989, Operator workload: Comprehensive review and evaluation of workload methodologies, Report No. 2075-3, US Army Research Institute for the Behavioural and Social Sciences, Willow Grove.

Megaw, E. D. 2005, The definition and measurement of mental workload. In: Wilson, J. R., Corlett, E. N. (Eds.), Evaluation of Human Work, Third Edition, (Taylor & Francis, London).

Pickup, L., Wilson, J. R., Norris, B. J., Mitchell, L., and Morrisroe, G. 2005, The Integrated Workload Scale (IWS): a new self report tool to assess railway signaller workload*, Applied Ergonomics*, 36(6), 681-694.

Young, M. S., & Stanton, N. A. 2001, Mental workload: theory, measurement, and application, in W. Karwowski (ed.), *International Encyclopedia of Ergonomics and Human Factors: Volume 1* (Taylor & Francis, London), 507-509.

Zeilstra, M., Bruijn, D., & Van der Weide, R. 2009, Development and implementation of a predictive tool for optimizing workload of train dispatchers, Intergo, Utrecht, confidential.