PREDICTION OF MENTAL WORKLOAD OF MONITORING TASKS

Colete Weeda, Msc. Eur.Erg., Melcher Zeilstra, Msc. Eur.Erg.

Intergo BV
PO Box 19218, NL-3501 DE Utrecht
The Netherlands

Existing objective instruments to investigate mental workload focus on performance or on physiological measures and are dedicated to specific and existing monitoring jobs and tasks. More generic and simply applicable instruments are subjective instruments. They are not applicable to predict mental workload. To overcome these disadvantages Intergo developed OWAT™ (Objective Workload Assessment Technique). OWAT™ is a calculating formula to assess and to predict mental workload of monitoring tasks in control rooms. A case study shows the way OWAT™ is used for this purpose. Using OWAT™ it is possible to give a clear view on the demanding tasks in the job and to start an objective discussion about the delicate issue of workload. Based on the results possible solutions to optimize the mental workload of the operators can be formulated.

Introduction

Existing instruments to investigate mental workload are dedicated to specific and existing monitoring jobs and tasks. Generally it takes months of preparation on analysing employee's activities before the assessment of the tasks can start. Objective instruments focus on performance or on physiological measures. More generic and simply applicable instruments are subjective instruments. These are not applicable to predict mental workload for not existing jobs or situations not existing yet. To overcome these disadvantages Intergo developed OWAT™ (Objective Workload Assessment Technique).

Development of OWAT™

As a result of earlier investigation (Breed, 2008) Intergo decided to choose the VACP method, developed by McCracken & Aldrich (1984), as a base for the new technique. VACP stands for the 4 modalities of human information processing: visual, auditory, cognitive and psycho-motoric. The VACP method is developed to assess jobs of the military air force. The calculation of mental workload cannot be generalized to other jobs or functions outside the military sector without any thought. To verify the VACP method is suitable as a generic instrument to assess and to predict mental workload, alternative calculation and scoring methods were proposed and investigated.

VACP varieties

McCracken & Aldrich (1984) and Bierbaum et al. (1989) developed the VACP method to calculate the workload of a task using these 4 modalities. Within each modality a number of descriptors are used to categorize a task (Table 1). A descriptor is a general term meant to easily categorize tasks.

Table 1: VACP descriptors

Visual	Auditory	Cognitive	Psycho-motoric
Detect an image	Detect a sound	Automatic	Speak
Read	Detect feedback	Recognize	Actuate one movement (e.g. push)
Scan Search Monitor	Listen (general)	Select alternative	Manipulate
Inspect Check	Interpret (speech)	Transform Calculate	Actuate complex movement (rotate)
Discriminate	Listen (selection)	Assess one element	Actuate continuous
Trace Follow	Discriminate	Code Decode	Actuate serial (data input)
Localize Point	Listen (patterns)	Assess more elements	Write

In Schuck (1996) the descriptor paraphrases of the study by Bierbaum et al. (1989) are adopted. The descriptor paraphrases are both generic as well as unambiguous and without problems applied in various studies (Schuck, 1996). In Schuck (1996) each descriptor is given a weighing score as an indication of the amount of workload. To determine this weighing score for each descriptor some thousands of tasks derived from a database were coupled to the proper descriptor. The most representative set of tasks per descriptor was chosen. These sets of tasks, each with its descriptor, were stated to a panel (e.g. the CP-140

pilots in Schuck, 1996) that is asked to assess the weight of the descriptors compared to each other. The result is transformed linear on a scale from 1.0 to 7.0. So the final weighing score of a descriptor is derived from a panel's judgment. It is highly questionable whether this descriptor weighing can be generalized to other occupational groups. Unfortunately no studies are known where a panel from outside the military aviation assesses the descriptors of Bierbaum et al. (1989).

To gain a clear understanding of the generalizability of the weighing 4 descriptor assessments by CP-pilots (Schuck, 1996), Kiowa observers (Schuck, 1996), Kiawa pilots (Schuck, 1996) and UH-60 crew (Bierbaum et al., 1989) are statistically judged on their intra-class correlation coefficients (ICC's). From the calculated ICC scores it can be concluded that the descriptor assessments by the panels are not fully consistent. They are well consistent in the auditory and cognitive modalities, but less in the psycho-motoric modality. Consistency in the visual modality is very low. For monitoring jobs the visual modality is the most important, together with the cognitive modality. If the descriptor weighing scores within one occupational group differ that much, it can be expected that the differences among various occupational groups will even be bigger. Therefor it did not seem recommendable to adopt the descriptor weighing from Schuck (1996) as a generic instrument to measure mental workload.

Alternative descriptor weighing

Sarno and Wickens (1992) note that omitting the descriptor weighing scores does not alter the validity of the VACP instrument. Inspired by this observation some alternative descriptor scoring methods are proposed and tested within various jobs. The aim of these different methods is to make the descriptors more suitable for a generic instrument and to simplify the procedure to achieve descriptor weighing scores. The following descriptor weighing methods are tested:

- Descriptor sum method (dSOM)

This is the original VACP method developed by McCracken & Aldrich (1984) with different weighings per descriptor. For each modality the sum of the descriptor scores is calculated.

- Descriptor maximum method (dMAX)

This method also is based on the original method from McCracken & Aldrich (1984) to score descriptors by the dSOM method. The adjustment is only to score the highest weighing descriptor within a modality.

- Modality score (MoSc)

Independent of the amount of descriptors, only the number of loaded modalities is counted. So the score always lies between 1 and 4. This method is very simplified in comparison with the dSOM method, which may not contribute to the validity of the method.

- Descriptor unweighted sum (dSOMon)

As noted by Sarno & Wickens (1992), replacing the descriptor weighing scores in the VACP method of McCracken & Aldrich (1984) by 1 when the descriptor is present and 0 when the descriptor is absent does not alter the results. By doing

so the subjective judgment of the descriptors is omitted and the method looks more suitable for a generic instrument.

- Descriptor interference unweighted sum (diSOMon)

This method is highly comparable with the dSOMon method, except that the diSOMon method takes per modality interfering descriptors into account. Scoring of the descriptors in the VACP method using diSOMon is based on the Multiple Resource Theory (MRT) of Wickens (2002). This theory is developed to explain which tasks can be performed simultaneously and which cannot, depending on the usage of different resources of human information processing. Hamilton & Broughton (2004) elaborate on this theory with their interference matrix.

- Descriptor correction dSOMon (dSOMonCor)

The dSOMonCor arises from the dSOMon after applying a descriptor correction rule. The rule implies that, if within the modalities V, A or C more than one descriptor is scored, a specific descriptor implicitly related with the other descriptor is extracted.

- Descriptor correction diSOMon (diSOMonCor)

The diSOMonCor arises after applying the correction rule mentioned above on the diSOMon.

Prins (2009) advises to use the dSOMon method in OWAT™. This is the most practical instrument to achieve descriptor scores and the results are equal to other methods.

Validation of the VACP method

The validity of the developed assessment technique is demonstrated by comparing mental workload scores calculated with different approved methods with the scores calculated with the proposed methods based on the VCAP method (Prins, 2009). Breed (2008) already compared the VACP method with Task Weighing™, a validated instrument developed by Intergo (Zeilstra, 2009) to investigate workload of train dispatchers, within time frames of 5 minutes of performing train dispatching tasks. But it is also important to investigate the correlation between the separate task scores of the different instruments. Therefor the VACP scores of the supplier of travel information's task, assessed by Breed (2009) were used. To validate these scores Prins (2009) also determined the workload of this job using IWS and the SWORD method, both validated methods to assess mental workload. Results of the assessment of a train dispatchers' task show a correlation of .69 to .88 between the proposed scoring methods and Task Weighing™. Results of the assessment of a supplier of travel information's task show a correlation of .75 to .88 with the SWORD method and a correlation of .74 to .93 with the IWS method.

Normative boundaries

People are able to deal with mentally highly demanding tasks during short periods of time. They use compensating strategies to execute the tasks and to prevent faults and misses (Wickens, 2002; Kahneman, 1973; Embrey, Blackett, Marsden & Peachey, 2006; Rueb, Vidulich & Hassoun, 1994). These

compensating strategies are in the long term associated with performance devaluation, and problems with health and motivation. So the question is: when will the mental workload become too high? Normative boundaries are developed to assess not existing jobs or tasks. These normative values were derived from earlier investigation by Intergo using Task Weighing™ (Zeilstra, 2009).

Case study with OWAT™

Task analysis

The job concerning is the national supplier of travel information. Especially in case of delays of trains and disturbances on the track with an impact at national level, the supplier formulates travel information consisting of cause, prognosis and travel advise. Together with the suppliers, tasks and actions were described (see

Table 2, first column). At the occurrence of a disturbance the supplier is informed by various information systems and by persons by telephone or speech communication (e.g. train dispatchers). The supplier fills information systems called IVIS and ARGOS with data concerning the disturbance (records). IVIS is a system providing internal users (e.g. train personnel) and external users (e.g. travellers with mobile phones) with train information. ARGOS is a system providing written announcements on stations for travellers. During the disturbance the records need several updates. After the disturbance is dissolved the records are closed and removed.

VACP scores

Each task is assessed by a human factors expert of Intergo for aspects related to workload in information processing activities; the descriptors of the VACP method. Using the diSOMon method each task is scored. Also the duration of the task is taken into account and the VACP score is recalculated per minute. This results in a score per task per minute (see

Table 2, third column).

SWORD

The assessment resulting in the VACP scores of the tasks were validated using the subjective instrument SWORD (Subjective Workload Dominance). The SWORD method, developed by Vidulich, Ward & Schueren (1991), compares the mental workload of different tasks within a job by presenting each possible pair of tasks in a questionnaire. With SWORD mental workload is assessed on a 17 points scale.

The result of this validation in terms of ranking is shown in

Table 2, fifth column. The task indicated with 1st is the most demanding task for mental workload, the task indicated with 10th is the least demanding task.

Table 2 also shows ranking of tasks using VACP scores.

Differences in ranking could be clarified for the most by taking the factor time into account: when the VACP scores per minute were multiplied with the estimated duration of task performance in minutes, ranking in terms of VACP scores was equal to ranking in terms of SWORD assessment.

Table 2: Tasks of the supplier of travel information

Task	Task description	VACP score per minute	Ranking VACP	Ranking SWORD
1	Consulting coordinator or colleague supplier of travel information	4-14	5th	6th
2	Enter first announcement in IVIS: situation, consequences, affected train stations, prognosis on duration	10	1st	3rd
3	Update existing announcement in IVIS: changes on situation, consequences and prognosis, measurement and travel advice	7	3rd	2nd
4	Cancel existing announcement in IVIS as soon as the disturbance or delay is over	6	4th	7th
5	Remove announcement in IVIS	4	6th	9th
6	Enter first announcement in ARGOS: route, cause and message	8	2nd	5th
7	Update existing announcement in ARGOS	2.5	8th	4th
8	Extensive update of existing announcement in ARGOS	10	1st	1st
9	Cancel existing announcement in ARGOS as soon as the disturbance or delay is over	3	7th	8th
10	Remove announcement in ARGOS	2	9th	10th

* All tasks include broadcasting

Calculating workload

Furthermore the suppliers of travel information described 2 imaginary but realistic cases of 2 hours work, one during rush hour (7AM till 9AM) and one during a disturbance (also during rush hour). The VACP scores for each task were projected on the cases to predict the workload in both situations. The scores per 5 minutes are added up and weighted against the normative boundaries.

Figure 1 shows the results of OWAT^{TM} calculations during rush hours without disturbances. In these situations suppliers of travel information have to announce small delays of trains. Calculations showed that workload is acceptable for these situations.

During a disturbance the suppliers have to update the information and formulate an advise for travellers about alternative travel routes. Calculations showed that at the beginning of a disturbance, workload would be unacceptable for one supplier of travel information, but two can do the job. Nevertheless after a while activities to be performed will also lead to unacceptable workload, even for two suppliers.

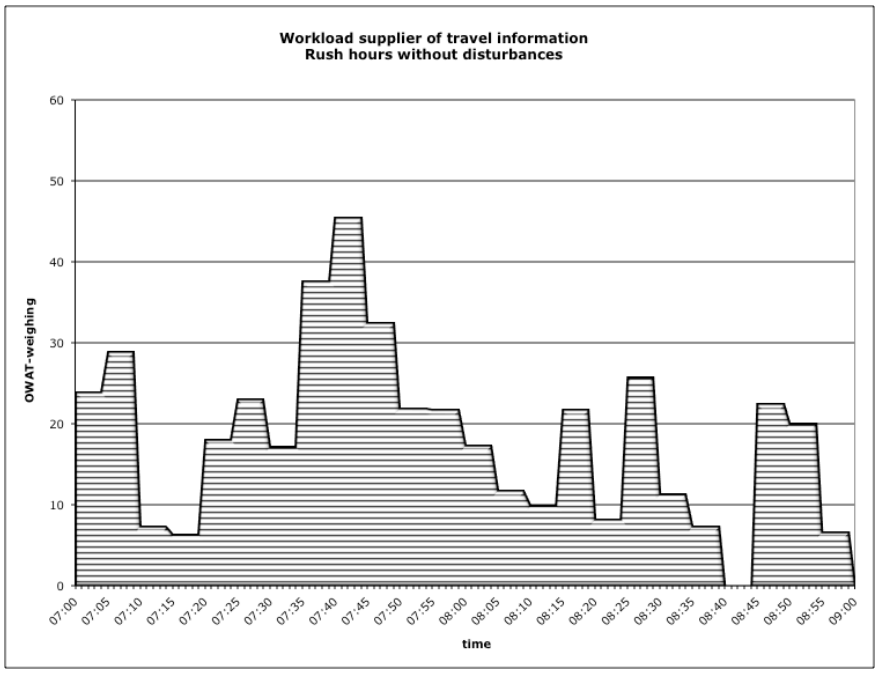

Figure 1: Workload for supplier of travel information during rush hours (no disturbances)

Conclusion and discussion

The suppliers of travel information agreed with the results of the OWAT^{TM} assessment. They confirmed the predictions for the mental workload in the descript cases. The supervisors were glad with the clear view on different situations. And the assessors were able to point the bottlenecks and relate them to certain circumstances. Also the assessors were able to suggest various solutions and, using the description of the cases and the figures, to discuss the influences on the mental workload with the supervisors and the suppliers.

References

Bierbaum, C.R., Szabo, S.M., & Aldrich, T.B. 1989, *Task Analysis of the UH-60 Mission and Decision Rules for Developing a UH-60 Workload Prediction Model*, (Army Research Institute for the Behavioral and Social Sciences, Fort Rucker U.S.)

Breed, I. 2008, *Mentale Werkbelasting voorspellen: het VACP-model*, (Intergo bv, Utrecht), confidential

Embrey, D., Blackett, C., Marsden, P., & Peachey, J. 2006, *Development of a Human Cognitive Workload Assessment Tool*, (Lancashire: Human Reliability Associates)

Hamilton, W.I., & Boughton, J. 2004, *A Theoretically Coherent Workload Prediction Technique*, (Bristol, Human Engineering Ltd)

Kahneman, D. 1973, *Attention and effort*, (New York: Prentice-Hall)

McCracken, J.H., & Aldrich, T.B. 1984, *Analysis of selected LHX mission functions implications for operator workload and system automation goals*, (Anacapa Sciences, INC. Military programs, Alabama)

Prins, E. 2009, *Ontwikkeling van een generiek meetinstrument voor mentale werkbelasting: OWAT*, (Intergo bv, Utrecht), confidential

Rueb, J.D., Vidulich, M.A., & Hassoun, J.A. 1994, Use of Workload Redlines: A KC-135 Crew-Redution Application, *International Journal of Aviation Psychology*, 4, 47-64

Sarno, K.J., & Wickens, C.D. 1992, Predictive Workload Models and Multiple-task Performance, *Proceedings of the human factors society 36th annual meeting*, 12-16

Schuck, M.M. 1996, Development of equal-interval task rating scales and task conflict matrices as predictors of attentional demand, *Ergonomics*, 39, 345-357

Vidulich, M.A., Ward, G.F., & Schueren, J. 1991, Using the Subjective Workload Dominance (SWORD) Technique for Projective Workload Assessment, *Human Factors*, (HFES), 677-691

Wickens, C.D. 2002, Multiple resources and performance prediction, *Theoretical Issues in Ergonomics Science*, 159-177

Zeilstra, M.P., de Bruijn, D.W., & van der Weide, R. 2009 *Development and implementation of a predictive tool for optimizing workload of train dispatchers*, proceedings of the European Conference on Rail Human Factors, Lille, March 2009.