# CHAPTER 4

# RESULTS

In this chapter the results of the verification study is discussed. Key results are referred to but the detail results are contained in the report attached as Appendix D. Furthermore a sensitivity analysis is performed to provide an indication of the relative contribution to the MWLI of each factors considered.

## 4.1 CORRELATING THE MENTAL WORKLOAD INDEX WITH PSYCHOPHYSIOLOGICAL PARAMETERS

A requirement was set that the mental workload assessment tool should be a valid assessment of mental workload. This requirement was set to ensure that a mistake made before by Spoornet was not repeated. It was previously decided to use a formula for workload assessment and through prolonged use it was assumed that this formula represented workload, only to discover much later that no proof could be found of any verification that the formula was indeed an indication of workload.

As mentioned earlier, it was initially decided to correlate the NASA-TLX with the MWLI in order to verify the MWLI as a valid assessment of mental workload. Permission to do this was obtained from the developers of the NASA-TLX. The practical circumstances, with varying shift patterns at the different train control centres, meant that it would intrude on the tasks of the TCOs and the notion was therefore rejected.

It was decided that physiological and other data would be collected over an eight-hour period, irrespective of the shift length, starting at 06:00 and finishing at 14:00. For some TCOs this would entail an entire shift but for those who worked a 12-hour shift, this would not be the case.

The pilot study (see Appendix C) showed promising results and, upon approval from the Ethics Committee, the Department of Physiology, University of Pretoria, assisted with the physiological measurements. The author was responsible for the design of the research parameters and approach, setting the requirements of the protocols to be used, as well as the criteria for testing and the desired outcomes.

There was considerable difficulty in finding a suitable instrument that would measure heart-rate variability under the conditions of electromagnetic interference, as was experienced in the pilot study. An instrument had to be imported from the USA, and only after a delay of several months could the field study commence.

Data collection occurred over a period of 9 weeks at 20 train control centres. An initial calculation of ideal sample size indicated that a minimum of 50 centers had to be included in the study to be clinically significant. A maximum of 36 RTO centers existed. All 36 centers were not included due to resource constraints and control over the sample size was therefore problematic. A sample of 20 was considered feasible considering the logistical and resource restrictions. Due to the small sample size, this was not a proper validation study, but rather a pilot verification study through which an approach was developed for a future validation study.

The results of the field study are contained in a report *Validation of the Spoornet Mental Workload Index against an Allostatic Load Index* (see Appendix D for the full report).

## 4.2 THE VERIFICATION STUDY

The MWL methodology was developed but it still remained to be proven that the index was a valid indicator of mental workload. The purpose of the verification study was to use an independent, objective measure of stress associated with workload, apply it to the different classifications of train control centres, and to determine whether there was supporting evidence for the original postulate.

The approach in the verification study was to examine the physiological changes in TCOs, if any, as indicators of workload stress at the selected train control centres.

The results of these physiological measures were then correlated with the Mental Workload Index to determine whether the calculated index correlated with measures of high and low stress.

### 4.2.1 Identification of the Experimental Group

Twenty train control centres across South Africa were selected as test sites for the field study.

Using the newly developed MWLI, the train control centres were clustered into high, medium, and low workload by the project team. Using a stratified selection approach, nine *high* and ten *low* workload centres were identified for inclusion in the field study. The scatter gram (Figure 4.1) gives an indication of the clusters that were formed. (A 20$^{th}$ train control centre was included to form part of the field study, although it does not have a pure RTO system – a hybrid system of RTO and track warrant is being used here. This centre was included at the request of the zone operational manager because a decision was pending on whether to split the section and allocate two TCOs instead of one to the section, and objective information was required to make the decision. This train control centre is considered to be a high workload centre.)

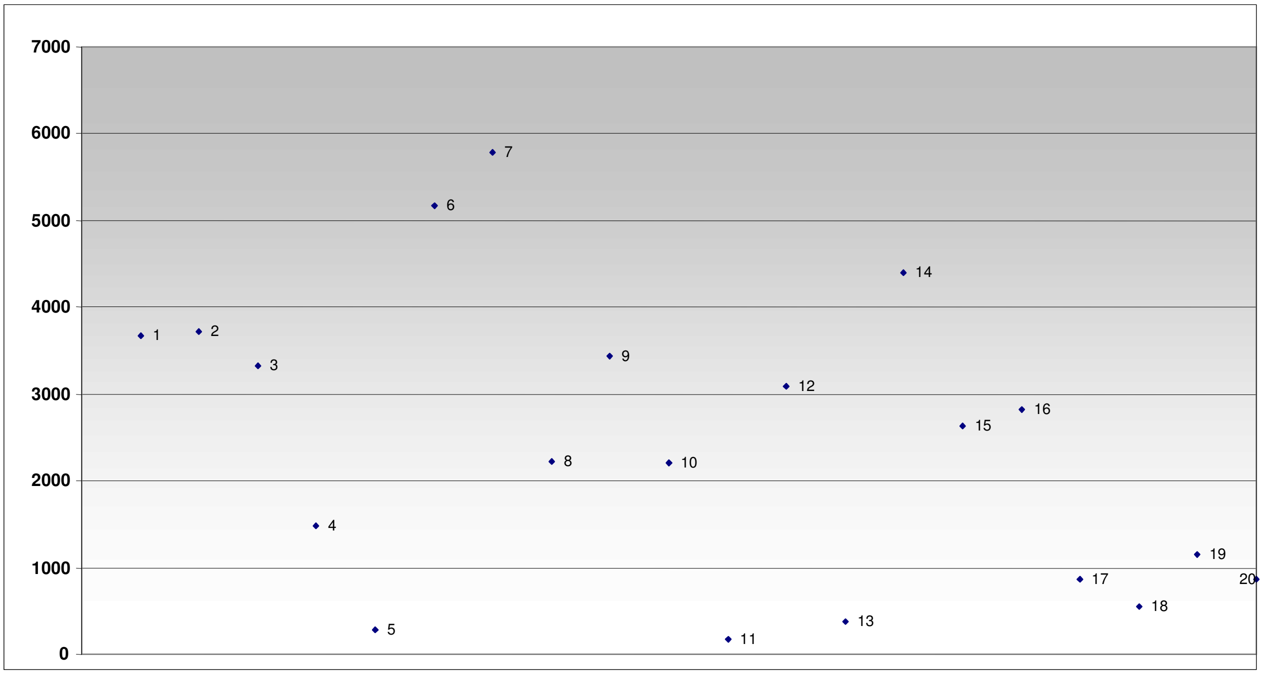


**Figure 4.1 Scatter gram of centres used to correlate with MWLI**

In the selection of the centres, it was decided to concentrate on distinctive high- and low-classified centres according to the calculated index. Borderline cases were excluded because it was felt that they might not assist in the verification study. Time and cost constraints were also limiting factors. The choice of centres was also influenced by geographical proximity to each other and ease of accessibility, i.e., their proximity to airports.

Table 4.1 contains the train control centres that were selected for inclusion in the verification study and their respective calculated mental workload indexes. To ensure confidentiality the names of the train control centres were omitted and identification numbers were allocated.

**Table 4.1: Train control centres included in the verification study**

| **ID** | **Name of centre/section** (Deleted to maintain confidentiality | **Calculated MWL Index** | **Rating** |
|---|---|---|---|
| 1 | | 2629 | H |
| 2 | | 553 | L |
| 3 | | 5789 | H |
| 4 | | 1155 | L |
| 5 | | 281 | L |
| 6 | | 3320 | H |
| 7 | | 1476 | L |
| 8 | | 2228 | L |
| 9 | | 3722 | H |
| 10 | | 3666 | H |
| 11 | | 5174 | H |
| 12 | | 2214 | L |
| 13 | | 171 | L |
| 14 | | 3443 | H |
| 15 | | 2828 | H |

| ID | Name of centre/section (Deleted to maintain confidentiality | Calculated MWL Index | Rating |
|---|---|---|---|
| 16 | | 4399 | H |
| 17 | | 374 | L |
| 18 | | 3086 | L |
| 19 | | 866 | L |
| 20 | | 863 | L |

## 4.3 MEASUREMENT TECHNIQUES AND METHODOLOGY

### 4.3.1 Measurement Techniques and Instruments

The following measurements were performed during the field study:

- **Cohen's Measure of Perceived Stress** – Subjects completed the stress questionnaire, with the assistance of the field worker, to assess their non-specific appraised stress during the last month. The questionnaire consists of 14 items of which seven are positively formulated and seven are negatively formulated. (See Appendix A.)
- **Blood pressure** – Diastolic (DBP) and systolic (SBP) blood pressure were measured with a digital electronic blood pressure meter (ALP K2, model DS-125D, Japan). Three consecutive measurements were taken, the first at the commencement of the shift and every two hours thereafter during the shift. The average of the three measurements is reported.
- **Heart Rate Recording and Analysis** – A Direct Wired POLAR ® heart rate belt (Mini Mitter Co. Inc. Bend, OR, USA) was attached to the subject. Heart rate data was transferred instantaneously via a direct wire to the Mini-Logger® Series 2000 (Mini Mitter Co. Inc.) and stored until download. Heart rate data, recorded continuously during the eight-hour exposure, was then downloaded to a laptop computer using the Mini-Log™ 2000W for Windows® (Mini Mitter Co. Inc. Bend, OR, USA).
- **Cortisol Collection and Data Analysis** – Cortisol is found in the blood or saliva. Free salivary cortisol was determined with a Salivary Cortisol ELISA kit (DX-SLV-2930, AEC Amersham Pty, Ltd., South Africa). Saliva was collected in a clean collection test tube at

the start of the shift and every two hours thereafter until 14:00 and then stored on ice to centrifuge later.

- **BMI** was calculated for all subjects.
- **TLA** – A real-time TLA was recorded during the same shift that the physiological measurements were performed. (See Appendix B for TLA template.)

The same observer was responsible for the TLA and the physiological measurements of all the subjects tested at all 20 train control centres. This was done to ensure objectivity and to eliminate any contamination of data that may have resulted from the different approaches of different individuals.

### 4.3.2 Measurement Methodology

Testing commenced at the start of the shift at 06:00. In four of the cases the TCOs and area management had an agreement that the shift would commence only at 07:00. In these instances testing commenced at 07:00. The reasons for deciding on this time period were:

- Physiological measures had to be taken during the day shift because of the body's physiological changes, including changes in cortisol levels, during the night due to the circadian rhythm.
- A period of eight hours was chosen due to the fact that in most of the train control centres eight-hour shifts were worked. An eight-hour period was therefore used consistently even if a 12-hour shift was worked.

All operations managers were informed of the study a month prior to commencement of the field study and were requested to inform TCOs of the study and its aims. Employee representative organisations were involved throughout the development process and ensured communication with and cooperation from their members.

Upon commencing their shift, the subjects performed their handing over duties and were then informed by the observer about the exact nature of the physiological measurements. The experimental procedures and aims were explained, consent to participate was obtained where after testing commenced. Participation in the study was voluntary and TCOs could choose not to be tested. There were no instances of refusal to participate. Although the field workers knew that the workload at the different train control centres differed, i.e., nine were high and ten were low workload centres, they did not have insight into which centres were high workload and which were low workload centres until the completion of the statistical analyses of the results.

### 4.3.3 Measurement Administration

The test protocol was submitted to and approved by the Ethics Committee, Faculty of Medicine, University of Pretoria and Pretoria Academic Hospitals (Number IPA-41 - 150/2000). Before testing commenced, all participants in the validation study were informed of the measurements by their managers. On the day of the study itself they were again informed of these by the field worker.

The tests were conducted by a postgraduate student in Physiology from the University of Pretoria under the supervision of a senior member of that department.

## 4.4 DATA ANALYSIS APPROACH

The statistics used in the analysis of the results can be summarised as follows:

All values are represented as mean ± standard deviation. A Two Sample T-test was used to determine whether differences existed in terms of allostatic load variables between the identified and high stress centres respectively. A two-way repeated measures Analysis of Variance (AOV) was used to evaluate between-subject differences (low and high stress centres) and within-subject differences (responses over an eight-hour shift). Group*Time interaction was also evaluated to determine whether both groups responded in a similar fashion over the eight-hour shift. If between- and within-subject analyses did not indicate any significant differences, but Group*Time interaction did, it would have meant that the two groups responded differently to the stressor over time.

All data and statistical analyses are contained in a separate report (Appendix D). The Statistix Ver. 8 program was used for the statistical analyses. The overall approach that was followed, i.e. the collection of the raw data, measurements, data and statistical analyses, conformed to the set requirements and provided the necessary objective information.

## 4.5 RESULTS OF THE VERIFICATION STUDY

Twenty train control officers at various train control centres were selected to take part in the physiological verification of stress levels manifested as a result of mental load. Selection was performed on the basis of an estimation of the stressor impact at the different centres as calculated by the MWLI. The brief for the verification study was to investigate whether the MWLI was supported by the values of physiological stress indicators.

All raw data, results of statistical analyses, findings, and discussion concerning these are contained in the comprehensive report attached as Appendix D.

The mean physiological values over the shifts for sub-groupings according to the MWLI were compared to see whether physiological differences could be found between individuals that had been grouped according to high and low MWLI values. The parameters that were included in this comparison comprised factors that could either reflect the stress reactivity in individuals, or the wear and tear as a result of chronic exposure to stressors, or factors that could influence the aforementioned two types of indicators. Comparisons were made for age, mass, height, body mass index (BMI), surface area (SA), systolic blood pressure, diastolic blood pressure, mean arterial pressure, pulse pressure, heart-rate variability variables, smoking, length of previous shift, length of test shift, number of years experience at a particular station, shift preferences and timeline analyses of the shifts when the physiological recordings were made.

Very few parameters showed any significant difference between the high- and low-stress groups. Only the TLA, which is built into the MWLI, showed, as could be expected, statistical differences between the high- and low-workload groups throughout. In short, the values of the physiological stress parameters did not mirror the MWLI. To see whether exposure to high workloads caused higher anticipation stress in workers, the arrival values (before the workload could have had any effect) between the two groups were compared for those parameters that were measured on arrival. Again, no significant difference could be found between high- and low-stress groups.

Statistical power analysis was not conducted due to the small sample size.

In order to ensure a sharp distinction between high and low workloads, as calculated by the MWLI, the values of only those individuals at the six highest and six lowest ranked train control centres were compared. The statistical analyses of this data once again showed that physiological stress values did not mirror the activity intensity as described by the MWLI. Very significant correlations were again seen between timeline analyses and experimental values. These correlations between the TLA factors and the early-morning values were probably only a reflection of the fact that timeline analyses were built into the MWLI. This nevertheless confirmed the fact that the MWLI is a good reflection of activities at the specific centres.

The next step was to look at the reactivity over the duration of shifts. When the high and low MWLI groups were compared in terms of the way they physiologically reacted to the workload over the total shifts, hardly any correlations were seen between the workload and reactivity. Many factors probably contributed to the fact that the results of the MWLI were not mirrored by the stress levels of the workers. The most likely contributing factors were probably individual differences such as age, health status, years of experience, and gender. The populations at the various stations differed significantly with regard to such aspects, and with the small experimental group size of this project, such differences could have nullified any significant differences. An additional confounding factor was the fact that three of the twenty workers were female. There are indications that females are perhaps more stress-responsive than males. This, however, is contradictory as there are certain factors that protect females against the negative effects of high-stress system activation. In this study the females had lower stress levels but were also younger than the average TCO – a fact that could very well explain their low allostatic loads.

There are several other reasons why the differing workloads of stations might not necessarily have been reflected in the physiological values of workers and in the changes in response to workload at the specific stations. The role of the adrenal medulla, an extension of the sympathetic nervous system, needs to be mentioned here. The secretory response of the adrenal medulla to stressors is dependent on the previous stress history of the individual. For instance, with exposure to the same stressor each day over an extended period of time, subsequent exposure to the same stressor may induce a lower adrenal medullary response. This response is known as habituation. However, if the same individual who has been stressed by a particular stressor over an extended period of time is suddenly, in addition to the previous stressor, exposed to yet another, novel, stressor the adrenal medulla secretory response may be significantly larger. This response is known as sensitisation.

Habituation to known stressors may occur as a result of repeated exposure to the same stressor, and TCOs may therefore adjust to high workloads if they are subjected to them over extended periods of time. It is only when novel stressors are introduced into the work environment, such as dramatic changes to the normal routine, that sensitisation and the effects of high allostatic loads, with their concomitant repercussions for performance and mental and physical health, might be noticed. It is self-explanatory that changes in Spoornet infrastructure and policies may result in such novel stressors.

It would seem that acute stress superimposed on chronic stress is dependent on the individual's familiarity with the type of stressor. It appears, then, that the response to an acute stressor in the chronically stressed individual would be less than expected if the stressor is homotypic (a repetition of what caused the chronic stress response). In contrast, if a heterotypic stressor were applied to a chronically stressed individual, the response would be greater. It is possible, however, that cross-tolerance to stressors may develop, especially with those stressors that involve the same neurological pathways. In theory, this can be extrapolated to the working situation where high cortisol levels might perhaps not increase significantly as a result of increased levels of homotypic stressors, such as increases in the amount of work to which the individual is accustomed.

In contrast, cortisol levels may increase when atypical stressors are encountered. The implication that can be deduced from this (i.e., that the amplitude of the stressor-induced increase in cortisol levels and the circadian amplitude difference to additional homotypic stressors may be lower during chronic high stress) is that individuals with chronically high cortisol levels will not show the expected increase from basal values when stress levels increase and that this could be ascribed to the values being higher than normal. To extrapolate this to TCOs, one could expect that habituation to their respective workloads could rule out the development of significant stress differences between the different workload groups. The magnitude of the stress reactions from baseline to stimulated responses could be negated by above normal baseline values.

The next step was to identify high and low stress groups on the grounds of physiological variables. A comparison was made between the values of the MWLI at the various centres on the one hand, and the physiological indicators of stress in the workers at the different stations, on the other. The possibility of separating the workers at the different centres into high- and low-stress groups on the basis of the values of their physiological parameters was investigated. The analyses, as previously mentioned, were performed in terms of an adapted allostatic load measurement where the data used in the calculation of the numerical value for each individual factor or parameter of the allostatic load were derived from the mean of the values obtained over the work shift. The reason for this was that baseline values in this group could not really be seen as baseline because all of them appeared to arrive with values higher than those recorded during the experimental procedure. Many factors could have contributed to this observation, most noticeably the fact that their arrival time coincided with higher cortisol levels relating to the circadian rhythm. Subsequent measurements were performed at times when the circadian values were already lower. The anticipation of stress before commencing a shift could have been a further contributing factor. In addition, it is known that individuals in high-stress jobs seldom recover to baseline values during their time off.

There can be no doubt that the stressors in the case of TCOs should be considered uncontrollable as the number of trains and other activities are predetermined by the job at hand and are not under their control. The fact that the stressor, in this case, is not a once off stressor but that TCOs live with these stressors for the major part of every day supported the decision to use the mean value over the shift worked.

As previously mentioned, in the process of subdividing TCO workers into high- and low-stress groups based on physiological values, the values of the following parameters were included: salivary cortisol and BMI as indices of hypothalamo-pituitary-adrenocortical axis activity; systolic and diastolic pressure as a measure of cardiovascular activity which largely reflects sympatho-adrenomedullary axis activation; heart rate as an indicator of sympatho-adrenomedullary activation; and heart-rate variability as indicators of autonomic activity. The values for each individual for each of the five indicators were classified according to the 50$^{th}$ percentile. Allostatic load was calculated by summing up the number of parameters for which the subjects fell into the highest risk, i.e., above the 50$^{th}$ percentile. Subjects were subsequently ranked according to their total hits.

Three models, Model A (including the values of cortisol, systolic blood pressure, diastolic blood pressure, heart rate, low frequency and BMI), Model B (cortisol, systolic, diastolic, heart rate and BMI) and Model C (cortisol, systolic, diastolic and heart rate) were developed and tested, and the TCOs were subdivided into high- or low-stress groups. When the TCOs were subdivided into high- and low-stress groups and the differences for all physiological parameters (not only those included in the three models) were tested for significant differences between the groups of individuals subdivided into the high- and low–stress groups according to Models A, B and C, it was seen that Model B was superior to Models A and C. This was supported by the fact that for Model A, 10 out of 17, for Model C, 8 out of 17, and for Model B, 12 out of 17 physiological variables differed significantly ($p < 0.05$) between the high-stress groups and the low-stress groups (two sample T-tests). The significant differences were also generally higher for Model B than for the other two.

In the final analysis, subdivisions of workstations on the basis of the MWLI were compared to subdivisions of the individuals at those stations and to perceptions of the observer regarding the stressor value of the workstation. A 90% agreement was found between the MWLI and observer perception (this was a subjective perception and not based on the sum of the activity as determined by timeline analyses) and a 60% agreement was found between the subdivisions into high- and low-stress stations by MWLI, on the one hand, and subdivisions based on Model B, on the other.

TLA is built into the MWLI and one would, therefore, expect to find some kind of correlation between TLA and MWLI. In an attempt to investigate the strength of the correlation and to see whether the MWLI has a significant advantage over simple TLA, correlations were tested between, on the one hand, the MWLI and, on the other hand, the individual factors in TLA as well as two combinations of TLA factors. The correlations between the MWLI and the two combinations of factors were $r = 0.7986$; $p = 0.0001$ (trains and authorisations) and $r = 0.9110$; $p = 0.0001$ (radio, telephone, schedules), and all individual factors also showed significant correlations. It would appear, therefore, that the TLA gives very much the same information as that derived from calculating the model. TLA as a stand-alone technique would, however, not be acceptable as a valid and reliable tool unless it had been validated through a scientific study. Furthermore, TLA on its own does not provide insight into the reason(s) for possible mental overload.

## 4.6 SENSITIVITY ANALYSIS AND OPTIMISATION OF THE MWLI

### 4.6.1 Sensitivity Analysis

Sensitivity Analysis is used to determine how sensitive the results of a study or systematic review are to changes in the manner in which a study was conducted. Sensitivity analyses are used to assess how robust the results are in the face of uncertain decisions or assumptions about the data and the methods that were used (Saltelli, Chan, and Scott, 2000).

Sensitivity analysis is also used to determine how ‘sensitive’ a model is to changes in the value of the parameters of the model and to changes in the structure of the model. Sensitivity analysis helps to build confidence in the model by studying the uncertainties that are often associated with parameters in models. Many parameters in system dynamics models represent quantities that, in the real world, are very difficult, or even impossible to measure with a great deal of accuracy. Therefore, when building a system dynamics model, modellers are usually slightly uncertain about the parameter values they choose and must therefore use estimates. Sensitivity analysis allows them to determine what level of accuracy is necessary for a parameter to make the model sufficiently useful and valid. If the tests reveal that the model is insensitive, then it may be possible to use an estimate rather than a value and achieve greater precision. Sensitivity analysis can also indicate which parameter values are reasonable to use in the model (Breierova and Choudhari, 1996).

The objective of sensitivity analysis here is to provide an indication of the relative contribution to the final MWLI of each of the factors considered. The sensitivity of the MWLI to changes in the *task factor* and *moderating factor* weights was determined by a first-order sensitivity analysis, using the measured data and changing only one parameter at a time while keeping all the other parameters at their nominal values. Sensitivity in this context is expressed as a ratio of the difference between the two factors. The nominal weights, relative to which the changes were considered, were taken as the weights determined by the consultative process described earlier. Table 4.2 shows the nominal weights of each of the task factors and moderating factors.

**Table 4.2: Nominal weights of task and moderating factors**

| **Task Factors** | **Symbol** | **Nominal weight** |
|---|---|---|
| 1. Data transactions | D | 1 |
| 2. Communications | C | 5 |
| 3. Authorisations | A | 15 |
| | | |
| **Moderating Factors** | | **Nominal weight %** |
| 1. Shift type | M1 | 12% |
| 2. Experience as a TCO on the particular system | M2 | 18% |
| 3. Planning complexity | | Total 33% |
| 3.1 Interface complexity | M3 | 5% |
| 3.2 Running times between crossing places | M4 | 8% |
| 3.3 Types of crossing places | M5 | 6% |
| 3.4 Locations of platforms | M6 | 3% |
| 3.5 Number of authorisations per shift vs. number of crossing | M7 | 11% |

| | | |
|---|---|---|
| places | | |
| 4. Inherent difficulty of the section | | Total 37% |
| 4.1 Type/mix of trains | M8 | 9% |
| 4.2 Presence of locomotive depots | M9 | 10% |
| 4.3 Presence of shunting yards/activities | M10 | 14% |
| 4.4 Topography | M11 | 4% |

Sensitivity of the MWLI was defined as

$$S_{PQ} = \frac{\Delta P / P_0}{\Delta Q / Q_0}, \qquad \text{(Generic Formula)},$$

with $\Delta P / P_0$ indicating the change in the MWLI in a particular group of train centres, relative to the nominal value $P_0$ of the MWLI for the same group, and with $\Delta Q / Q_0$ indicating the change in the weight of a particular parameter, relative to the nominal weight $Q_0$ of the same factor. The output variable P can be P=H, M or L representing the mean of the MWLI for the high (H), medium (M) and low (L) groups of train control centres, while the input variable Q can be index Q=A, C or D representing the task factors A=Authorisations, C=Communications, D=Data transactions or Q=M1…….M11 representing the moderating factors.

The train control centres were grouped as follows according to the calculated MWLI for the nominal values of the task and moderating factors. (See Figure 4.2 for H, M, L groupings):

High: MWLI $>$ 3000

Medium: 1000 $<$ MWLI $\leq$ 3000

Low: MWLI $<$ 1000

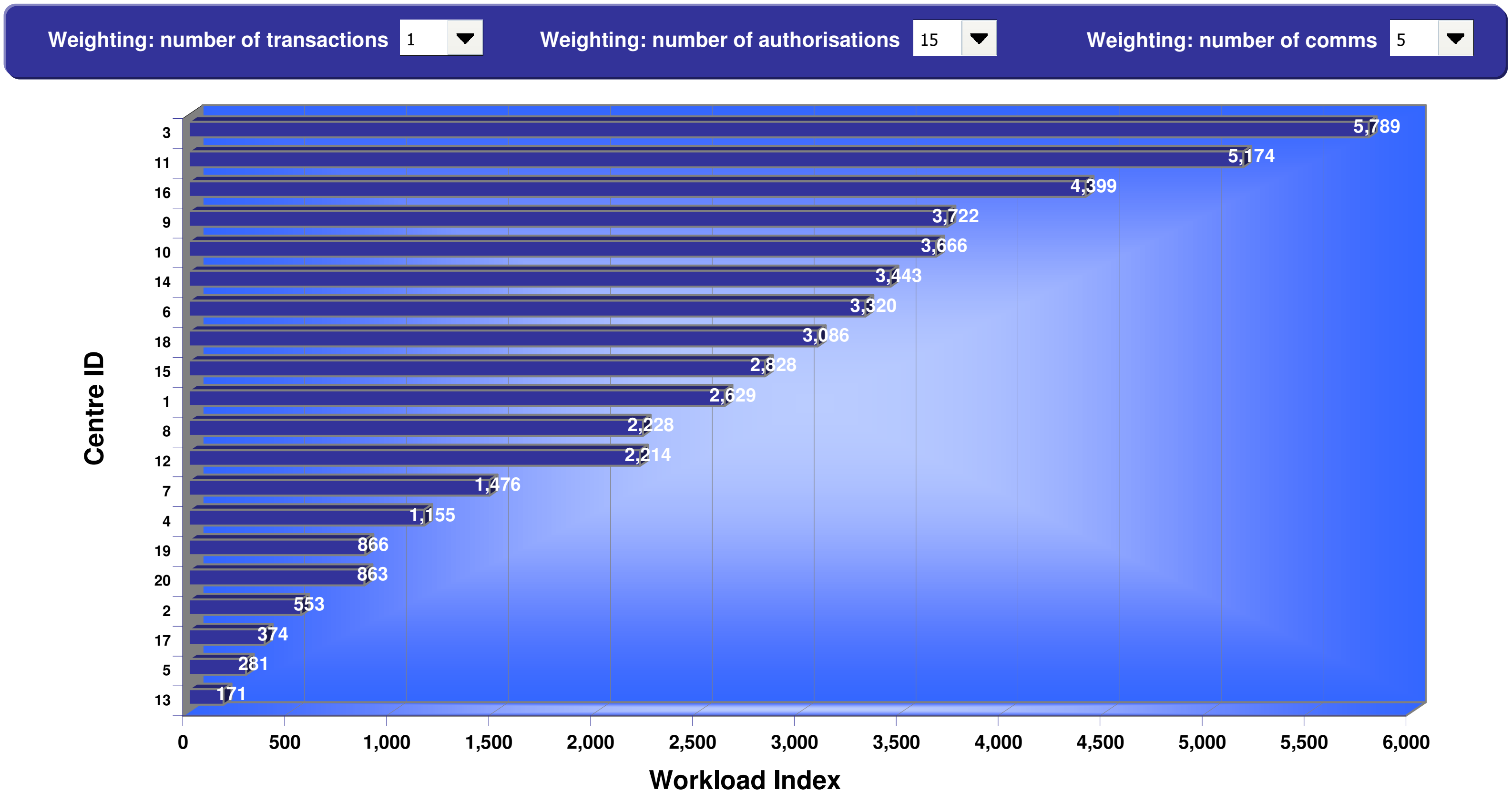


**Figure 4.2: High, medium and low groupings for calculated MWLI**

The mean values of the MWLI for each of the groups (H, M, L) for the nominal values are given in Table 4.3.

**Table 4.3: Mean value of MWLI for high, medium and low groups**

| Group | Mean MWLI | Centre ID |
|---|---|---|
| High | 4075 | 18, 6, 4, 10, 9, 16, 11, 3 |
| Medium | 2088 | 4, 7, 12, 8, 1, 15 |
| Low | 518 | 13, 5, 17, 2, 20, 19 |

The results of the sensitivity analysis for the above procedure are given in Table 4.4.

**Table 4.4: Results of sensitivity analysis of task and moderating factors for H, M, L groups**

| | | | | Group | | |
|---|---|---|---|---|---|---|
| Task Factor | Nominal value | Δ-Value | | High | Medium | Low |
| A | 15 | 16 | | 0.31 | 0.39 | 0.34 |
| C | 5 | 6 | | 0.69 | 0.60 | 0.65 |
| D | 1 | 2 | | 0.50 | 0.36 | 0.93 |
| Moderating Factor | Nominal value | Δ-Value | | High | Medium | Low |
| M1 | 12% | 13% | | 0.07 | 0.07 | 0.07 |
| M2 | 18% | 19% | | 0.07 | 0.02 | 0.04 |
| M3 | 5% | 6% | | 0.04 | 0.03 | 0.02 |
| M4 | 8% | 9% | | 0.06 | 0.06 | 0.06 |
| M5 | 6% | 7% | | 0.02 | 0.01 | 0 |
| M6 | 3% | 4% | | 0.02 | 0.02 | 0.01 |
| M7 | 11% | 12% | | 0.10 | 0.08 | 0.02 |
| M8 | 9% | 10% | | 0.04 | 0.03 | 0.04 |
| M9 | 10% | 11% | | 0.06 | 0.05 | 0.03 |
| M10 | 14% | 15% | | 0.10 | 0.11 | 0.07 |
| M11 | 4% | 5% | | 0.02 | 0.02 | 0.02 |

The sensitivities are expressed as fractions of 1 with S=0 indicating no change in MWLI for a change in the selected weight, S=1 indicating a 1:1 change in the MWLI for a change in the selected weight, and S=0.5 indicating that the MWLI changes by 50% from the nominal value for a 100% change in the selected weight.

In each case the weight was varied by one unit while all the other weights were kept at their nominal values. It should be noted that each of the calculated sensitivities is to some extent dependent on the nominal values of the other parameters. The results of a multivariate sensitivity analysis, in which more than one parameter is changed at a time, is more difficult to interpret and is considered beyond the scope of this study.

The moderating factors weights could not be analysed in the same way as the task factor weights, since there was a unique set of 11 nominal values for each train control centre. In the case of moderating factors the maximum contribution of each factor to the MWLI was considered as the nominal weight value.

The first-order sensitivity analysis revealed the following trends and their possible interpretations:

Task Factors:

- The sensitivity of the MWLI index in terms of the task factor with the highest weighted contribution, i.e. authorisations, with a nominal weight of 15%, has the lowest sensitivity with respect to all three the task factors. Further, the sensitivity is about the same (~0.3) for all three groups, high, medium and low. This indicates that the MWLI is least affected by a change in the number of authorisations, despite the fact that in the opinions of the expert advisory group involved in the consensus-driven design of the MWLI this task factor was considered to make the largest contribution to the MWLI. It appears, based on the results of the sensitivity analysis, that the perceptions, on which the relative contributions of the task factors to the MWLI were based, might have been incorrect.
- The MWLI in all three groups has a sensitivity to changes in the communications task factor of 0.6, which is roughly double that of the authorisations task factor, matching the intuitive observation that an incremental increase in the number of communications has

a greater effect on the workload of TCOs in all groups than an increase in the number of authorisations

- The sensitivity of the MWLI to changes in the number of data transactions shows the greatest variation (0.3 – 0.9). It is interesting to note that this factor shows the highest sensitivity in the low MWLI group (0.9). The interpretation of this could be that in a low workload train control centre a change in only one unit in the data transactions task factor results in a change of more than 90% in the MWLI. This fits the intuitive observation that an increase in the data transactions significantly increases the workload of TCOs in low workload centres.

Moderating Factors:

- The moderating factors with the lowest percentage contribution to MWLI (<8%) showed consistency in that the MWLI had the least sensitivity (<0.04) across all three groups. This was to be expected as these factors were considered to be the least important contributors to MWLI.
- The sensitivity of the MWLI to the moderating factor *running time between crossing places* (M4), which had a nominal 8% contribution, was higher than it was to the moderating factor *type/mix of trains* (M8) which had a nominal 9% contribution. The variation in the sensitivity was about 0.02 but there was consistency across all three groups.
- The moderating factors with the greatest variation in sensitivity across the three groups were also the factors with a higher percentage contribution (M2=18%; M7=11%; M9=10%; M10=14%). These factors also showed a higher MWLI sensitivity for the high workload group, except for M10, (*Presence of shunting yards/activities*) which showed a higher MWLI sensitivity for the medium workload group.
- The only moderating factor that presented an exception to the above trend was factor M1 (*shift type*) with a weight of 12%, which showed a lower comparative MWLI sensitivity (0.07) and no variation across the three groups.
- Except for shift type, it therefore seems that the moderating factors with the highest contribution to MWLI also showed the highest sensitivity. This can be taken to confirm the nominal contributions of the moderating factors being allocated consistently by the expert advisory group.

### 4.6.2 Optimisation of the MWLI

The MWLI is a multiparameter index that has no well defined minimum or maximum. Hence the optimum cannot be mathematically and quantitatively determined from its definition. The optimisation of the MWLI can only be addressed in a qualitative sense in terms of it being most appropriate for any given set of circumstances. The optimisation of the MWLI is beyond the scope of this study, inasmuch as it needs the long-term extensive use of the MWLI and an evaluation of its appropriateness needs to be made in terms of indirect and subjective indicators of mental workload – such as operator experience, turnover of staff, accident frequency, long-term health effects, sick leave and absenteeism – being correlated with the high values of the MWLI as determined by the proposed index.

## 4.7 SUMMARY OF RESULTS

The workloads at the various stations, as predicted by the MWLI, were not reflected by either the adapted allostatic load (including all measured parameters of the individuals working at those stations) or by changes in stress levels over shifts. However, in developing the three models, consisting of different combinations of allostatic load indicators, there was a 60% correspondence between the workstations subdivided into low and high workloads, according to the MWLI, and the subdivision of TCOs at the corresponding train control centres into high and low stress according to Model B. Model B was based on the mean of the values taken over the duration of the shift for cortisol, systolic pressure, diastolic pressure, heart rate and BMI. It can thus be said that the combination of physiological parameters used in Model B supports the potential usefulness of the MWLI to differentiate between high and low stress due to mental workload.

# CHAPTER 5

# DISCUSSION

In this chapter the hypothesis and research questions are critically reviewed in terms of the results and the literature. In closing, possible applications of the Mental Workload Index are discussed.

## 5.1 HYPOTHESIS AND RESEARCH QUESTIONS

The limiting factors such as sample size, heterogeneity of the sample and the habituation phenomenon, discussed in the previous chapter, compromised the significance of the results considerably. Despite this promising results were obtained and it can be concluded that this study has, with limited success, proved that it is possible to develop a preliminary objective and non-intrusive method to assess mental workload in a specific environment.

The scientific hypothesis that was tested in this research is:
*The objective mental workload index as developed in this thesis can differentiate between high and low imposed workload train control centres.*
The hypothesis can be conditionally accepted namely for this particular study with this particular sample, but it has not been proven correct for other circumstances.

The following research questions were posed:

- Which operator tasks could potentially be associated with mental workload?
- Which moderating factors, which would either increase or decrease mental workload should be considered and how can their respective moderating effects be determined?
- Which parameters should be considered to develop a measure of a mental workload that is completely objective and requires no estimation of experienced load by the operator?
- Can the proposed index differentiate between workload levels that could potentially be experienced at different train control centres?
- Which physiological measurements provide an objective assessment of mental workload?

The research questions are critically evaluated in order to determine whether they have been adequately addressed.

- Which operator tasks could potentially be associated with mental workload?
    - This question was answered adequately.
    - Tasks were identified by observing TCOs at work and then developing a TLA template containing these tasks. (See Chapter 3 – 3.7 and 3.8).
    - The tasks were verified by experienced TCOs as a true reflection of their daily activities.

- Which moderating factors, which would either increase or decrease mental workload should be considered and how can their respective moderating effects be determined?
    - This question was addressed satisfactory, but the contribution of the moderating factors to overall mental workload does not seem significant.
    - Through a process of user participation and after several iterations, moderating factors were identified. (See Chapter 3 – 3.7 and 3.8).
    - The contribution of each of these factors to mental workload was then determined using a modified nominal group technique. (See Chapter 3 – 3.9).
    - These values were normalised and scale points allocated to the descriptors. (See Chapter 3 – 3.9).

- Which parameters should be considered to develop a measure of a mental workload that is completely objective and requires no estimation of experienced load by the operator?
    - The answer to this question is not a definite affirmative or negative.
    - A qualified affirmative is probably the best answer based on the results obtained from this sample.
    - The shortcomings have been mentioned and better correlations are possible with proper sampling of subjects. (See Chapter 4 – 4.5).

- Can the proposed index differentiate between workload levels that could potentially be experienced at different train control centres?
    - This question was answered to an extent.

- With a refining of the groupings of parameters a correlation of 60% was obtained between these parameters and a high and low stress group of TCOs respectively. (See Chapter 4 - 4.5).
- Proper sampling would probably improve the correlations. This is addressed in Chapter 6 – 6.2.

- Which physiological measurements provide an objective assessment of mental workload?
  - This question could not be answered.
  - Based on the literature it was considered to be a good means of validation Physiological variables change with mental activity. It is therefore seen to be well suited as indicators of mental workload demand (Rasmussen, 1979).
  - Furthermore, because the changes that occur are involuntary and subjects are not requested to execute any extra overt behaviour it seems a suitable means of validation. (See Chapter 2 – 2.8.4).
  - Physiological measurements of stress associated with mental workload is however hard to isolate and because of other operator states that also impact on these parameters will always have a risk of being contaminated.
  - In this study the TCOs are all shiftworkers and therefore are subject to fatigue due to their schedules.
  - They perform safety-critical tasks and one would assume that the levels of stress would be elevated.
  - As explained in Chapter 4 (4.5) the relationship between stress and mental workload is a complex one and in retrospect this relationship is potentially not strong enough or weakened by related variables to such an extent that it should not be used as a basis for validation

It is concluded that there was little evidence of high levels of stress in TCOs, either from subjective stress ratings or from salivary cortisol levels or any of the other physiological parameters. These levels were well within normal levels for adults. These results should not be interpreted as an indication that workplace stress does not exist. It is more probable that data was collected too infrequently in this rapidly changing environment and may not have captured the changing workload and related stress.

It is significant that Popkin et al. (2001) documented results that are comparable to the results found here: 95% of cortisol samples in their study fell below the upper limits of normal cortisol levels. It is possible that these levels were suppressed due to either habituation or burn-out of the cortisol producing system within the brain. Either of these conditions could result if individuals are placed under high levels of chronic stress.

The physiological and self-rating data did not indicate that the dispatchers were under a high stress load. This is surprising given the number of items that were listed as contributing to their at-work stress. Anecdotal evidence, as well as the safety critical nature of the job, would also suggest a high-stress environment. It may be that dispatchers are a self-selected group that have either a higher tolerance to stressors or better coping mechanisms (Popkin et al., 2001).

Significant correlations between TLA and MWLI were found, which was to be expected. This means that the MWLI gives good reflection of the activities at the specific centres. Given these correlations it is possible to utilise only that part of the MWLI that calculates task factors, to determine the potential task load a train control centre.

## 5.2 CRITICAL EVALUATION OF THE CONTRIBUTION OF THE MENTAL WORKLOAD INDEX

Operator states such as fatigue and stress that are related to mental workload rarely present themselves as a single challenge in research. They are all relevant in environments of operational research (Desmond and Hancock, 2001).

The MWLI is an attempt to obtain a measure of 'pure' mental workload factors, but by its nature mental workload is always mediated by factors internal and external to the operator. An attempt was made to address mediating/ moderating factors in the MWLI, but a shortcoming of the index is that no other operator states were added as moderating factors. There was good reason for this namely the aim to develop a tool that operational personnel and others could use to objectively, reliably and quickly obtain taskload data. The effect of this shortcoming however is that the MWLI could potentially only measure operator stress and not mental workload.

The experience of mental workload depends on the individual and is the result of the interaction between the operator and task. Consequently, the same task demands do not result in an equal level of workload for all individuals (De Waard, 1996). A shortcoming of the MWLI is that it potentially underestimates the interactive nature of mental workload and a too simplistic view was enforced due to its emphasis on task and other objective factors. This simplistic view was acknowledged and possible criticism was anticipated (Chapter 2 – 2.2.3). Retrospectively, a more viable approach would have been to develop an index that would provide a calculation of only the task load at a specific train control centre rather than a mental workload index which implies operator states and other human factors.

It is necessary to ask whether either the development of a MWLI or the use of the physiological parameters of the workers at the various centres gave significantly better estimates of the stress levels at the stations than the use of only the TLA. At this stage the answer seems to be negative. Although the MWLI is a good reflection of the activities at the stations, and Model B (in the verification study) supports the use of the index, neither would appear to have an advantage over the TLA, except for the fact that the TLA is considered a subjective technique and therefore does not comply with the criteria of objectivity and scientific validation.

Another question, in view of the high correspondence between the MWLI and the subjective perceptions of an observer, is whether the workload could not be estimated simply through observation, without the use of the TLA, physiological measurements or calculation of the MWLI. Once again, the answer is likely to be negative as estimations of workload can differ, as they may be dependent on the ability, commitment, and objectivity of the observer.

It would seem from the literature and from the results that despite its shortcomings, the MWLI does offer some benefits over existing mental workload measuring techniques that potentially make it a useful index, provided that further research is done on its validity and application in other contexts.

Secondary measures of spare mental capacity offer the benefit of high face validity and a level of standardisation. Some of these measures can be applied to different primary tasks and the workload measure will be given in the same units. It however suffers on the intrusiveness criterion.

The MWLI also has high face validity due to the inclusion of task factors. It satisfies the intrusiveness criterion and information is collected unobtrusively and it can even be used as a desktop exercise.

Subjective rating techniques offer the advantages of not disrupting the task and ease of application. In comparison, the MWLI does not require any input from the operator, is completely non-intrusive and does not have the disadvantage of subjective techniques, namely that it is based on operators' perceptions and possibly does not diagnostically reflect mental workload.

Specific criticism of well-known subjective techniques such as NASA-TLX and SWAT, relate to the fact that these have been simulated in the military environment and is not considered relevant for TCOs as it does not capture all the relevant aspects of their tasks. The language and application is also not considered appropriate for use in real-time in the field with civilian populations. These instruments also assess the workload of a single task whereas TCOs complete multiple tasks concurrently (Pickup et. al., 2005).

The MWLI has distinct advantages over these shortcomings as it addresses the multiple concurrent tasks, is easy to apply and does not utilise subjective information.

The advantages of the MWLI over physiological measurements are evident. Ease of use, cost and the necessity of specialised field personnel make the use of these measurements virtually impossible.

In conclusion, the MWLI can potentially be used as a method of choice, provided that existing shortcomings, especially around its validation, are addressed in further research.

The sensitivity analysis in Chapter 4 was done to determine how sensitive the index is to changes in the value of parameters. It provides an indication of the relative contribution to the MWLI of each of the factors.

The first-order sensitivity analysis provided an independent confirmation of the assumptions about the data and the methods that were used in the design of the MWLI.
The conclusion is that the MWLI, as used in this study, could reliably discriminate between high and low mental workload centres based on traffic volumes. It is therefore a useful indicator and predictor of the stress associated with workload. The calculated MWLI should not be interpreted as an absolute value but rather as an indicator that should prompt managers of the system to take preventative action in order to ensure safe train operations.

A further aim of the study was to develop a methodology that would stand up to the tests of validity and reliability. This entailed that the criteria set by other researchers for mental workload assessment techniques should be met. Based on the results obtained with the MWLI, it is argued that this methodology meets the following criteria:

1) <u>Sensitivity</u>: The MWLI shows sensitivity to changes in task difficulty or resource demand, i.e., an increase in demand on the task elements has a direct and visible impact on the overall workload.

2) <u>Selectivity</u>: The index should be selectively sensitive only to differences in resource demand and not to changes in factors such as physical load or emotional stress, which may be unrelated to mental workload or information-processing ability. At the outset of the study it was made clear that the methodology should exclude personal factors and that these should be identified and addressed through other processes since they might impact on safe operations.

3) <u>Intrusiveness</u>: The MWLI does not interfere with, contaminate or disrupt performance of the primary task. The MWLI can be performed off-site once the required information has been captured. It does not require the direct involvement of the operator.

4) Operator acceptance: It was imperative that the TCOs supported and accepted the methodology and that they believed that the MWLI measures what it claims to measure. The MWLI was developed with the inputs from users of the system and, therefore, has gained credibility and acceptance.
5) Face validity: Because the MWLI uses task-related information and information that relates to operator capability as well as moderating factors that could add to or lessen operator stress and task difficulty, it has face validity.
6) Affordability: Costs were incurred in the development of the MWLI. There is no further cost in the implementation or use of the MWLI.
7) Implementation requirements: The MWLI can be calculated from information obtained from records and does not require site visits. It can be used to create various scenarios at a train control centre and determine the workload for different sets of variables.

Based on the process followed in this study, the MWLI could be adapted for other control and monitoring environments.

## 5.3 APPLICATION AND IMPLEMENTATION OF THE MENTAL WORKLOAD INDEX

The phenomenon that resulted in weaker correlations between the MWLI and specific physiological measurements, i.e., habituation to homotypic stressors, is very relevant in the train control environment. Usually, normal operating practices prevail. In other words, following the prescribed procedures is what TCOs do most of the time. It is only under abnormal conditions, such as the failure of radio communication equipment or a dramatic increase in train volumes (as is the case when maize is exported to neighbouring countries), that TCOs are stretched in terms of their capacity to deal with different ways of train control. This is called *abnormal working* in railway terms. Abnormal working forms part of the training of TCOs but is rarely used and only tested when the TCO undergoes refresher training every two years. TCOs therefore become used to, or habituated to, the normal operating procedures. It becomes a habit because it is part of normal daily behaviour.

This study did not focus on abnormal working conditions (or heterotypic stressors), but there is ample evidence that under abnormal conditions (or when *fall back* or abnormal working procedures have to be utilised) the workload becomes dangerously high and errors may result. The work group, all of whom are experienced TCOs, reported that, from personal experience, abnormal working conditions are considered to be very high risk. Everyone in the group reported having had an experience where abnormal working procedures had on occasion led to a near-miss situation in which an accident was averted at the last minute, or an accident was caused due to an error committed by the TCO. This is, however, seldom uncovered when a formal incident inquiry is made.

The reason for the high-risk nature of abnormal working conditions is that these pose a heterotypic stress situation, which adds to the mental workload already experienced by the TCO. Homotypic stressors do not refer to low stress levels but rather to the fact that individuals become used to the type of stressor and therefore do not show physiological responses to the stressor because they do not label the situation as stressful. Under abnormal working it is the novelty of the situation that makes it stressful and therefore compounds the existing mental workload experienced. This increased experience of workload can lead to errors of judgement or incorrect actions.

In recent years several accidents were attributed to abnormal working conditions in the train control environment. None of these occurred in the RTO environment, but all occurred in the colour light signalling environment due to copper cable theft which resulted in the failure of signals. TCOs then have to revert to authorising trains through radio communication (as is done in RTO) but it is not the normal way of working within the colour light environment. Two of the accidents resulted in loss of life and due to the sensitive nature of these accidents and ongoing claims against Spoornet, details may not be provided here.

It is therefore imperative that, in light of this study, cognisance is taken of those train control centres with a high workload in terms of the MWLI. Should there be abnormal conditions, especially in these centres, it should automatically follow that special measures are taken to avert errors that may occur as a result of overload.

## 5.4 APPLICATION OF A CRITICAL MENTAL WORKLOAD LEVEL

The unavailability of documented research that deals with the practical application of mental workload measurements in terms of a specific cut-off number for safe operations is problematic because very few benchmarks exist. A concept that could be an appropriate application in this context is the *workload redline* concept referred to by De Waard (1996).

In recent years, the question, “How much workload is too much?”, has received increased attention. In an applied setting such as traffic research, the workload redline could prove to be a very useful concept as the consequences of too much workload within the driving environment can be very serious. A proposed model of mental workload, task performance and demands provides the rationale for understanding the workload redline (De Waard, 1996).

A relationship between task demand and task performance has been described by Meister (1976). Meister defined three regions, regions A, B and C. Region A is characterised by low operator workload with high performance. An increase in demands does not lead to performance decrements. In region B the level of performance declines with increased task demands. So, region B is the region where performance decreases with increases in demand and workload. In region C extreme levels of load have diminished performance to a minimum level, and performance remains at this minimum level with further increases in demand (see figure 5.1).

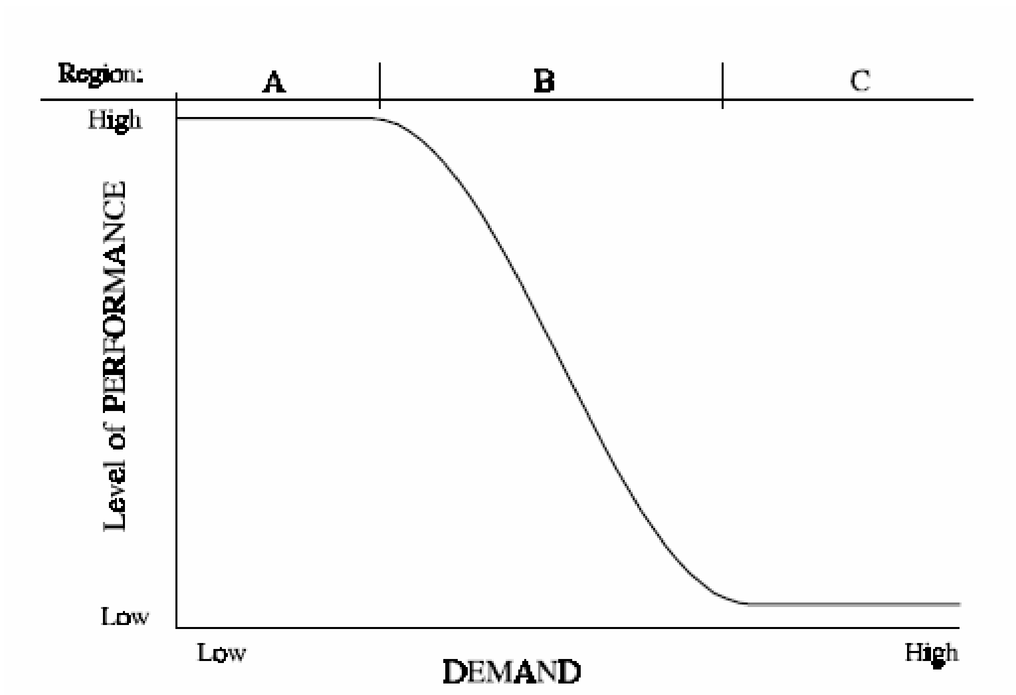


**Figure 5.1: Hypothetical relationship between demand and performance (based on Meister,1976)**

According to this model, also referred to as the region model, a primary-task workload measure, i.e. a measure of performance, will only be sensitive to variations in levels of workload in region B. In region A performance remains stable and is independent of variations in demand, while in region C performance remains at a minimum level, independent of demand. Other measures, such as self-report (or subjective) measures of workload, may be sensitive in region B and may clearly reveal overload in the C-region, while they need not be sensitive in region A. While extreme levels of load that result in overload can be situated in the C region, it is not clear where the domain of underload is situated (De Waard, 1996).

This model could be completed with the addition of a deactivation or a D region at the far left end. The effects of monotonous tasks, for example, are situated in the D region. These are low-demand tasks that can result in increases in task difficulty and workload as a result of a reduction in capacity. In case of boredom, for example, a reduction in capacity requires that a larger proportion of the capacity be used for performance of the same task, thus increasing mental workload (Meijman and O'Hanlon, 1984). By means of the addition of the D region the complete inverted-U is split into four regions, the D, A, B and C regions (see Figure 5.2).

The issue of too much workload is usually referred to as the determination of a *workload redline* (Reid and Colle, 1988, Wierwille and Eggemeier, 1993). When trying to tackle the determination of a redline there is a need to first decide upon the context of what constitutes 'too much'. Degraded performance may indicate too much workload, but the effects of personal wellbeing may prove equally valid. Preliminary work on workload redline puts this line at the transition from region A to B (see Figure 5.2).

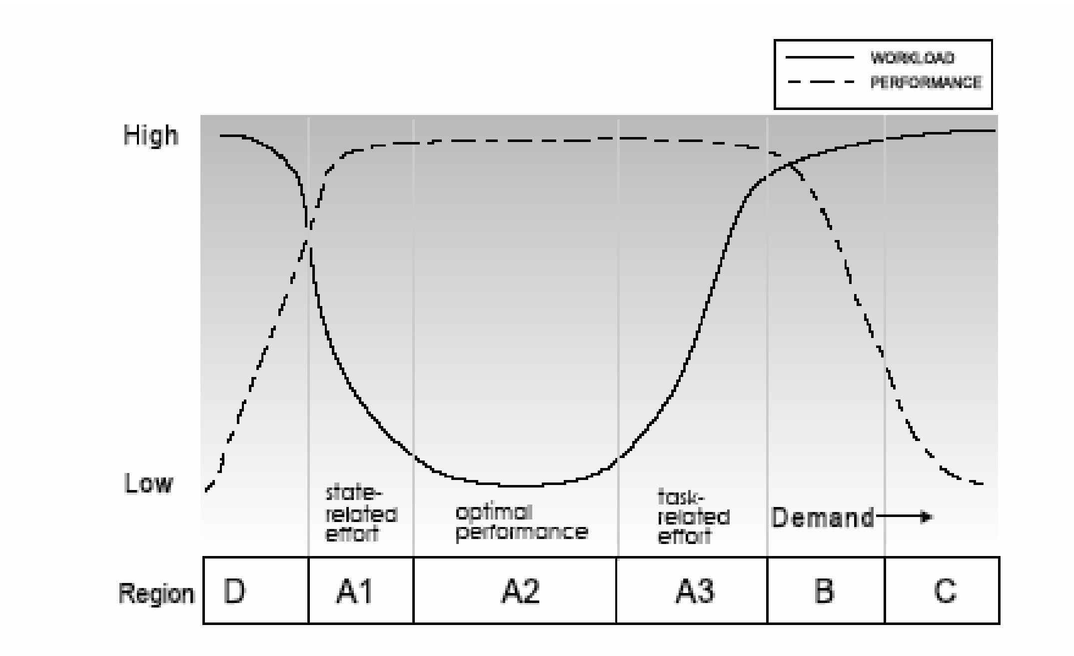


**Figure 5.2: Workload and performance in 6 regions (De Waard, 1996)**

Reid and Colle (1988) related just-detectable performance decrements to self-report ratings, and this workload rating was used to designate the absolute workload redline. The point of a just-detectable performance decrement occurs at the transition from region A to B. While it is clear that performance measures themselves have defined the A region, it may be useful to split the A-region into three parts. In the middle part, region A2, operators can easily cope with task demands and performance remains at a stable level with increases in demand without increased effort. In the A3 region, however, performance measures still do not show a decline, but operators are only able to maintain the level of performance with increased effort. Temporary compensation by the exertion of effort in region A3 is one of the advantages of human flexibility and is not critical. If, however, continuous effort is required to maintain performance, or if peak loads occur frequently, this can lead to stress, and an unhealthy situation that has to be avoided. This is particularly true if operators have no control over the situation. It may therefore be more useful to put a workload redline at the transition from region A2 to A3 instead of at the transition from region A (A3) to B. In this way, the term *workload redline* remains related to *workload* instead of to primary-task performance breakdown.

A similar situation exists at the region that is to the right of the D-region, region A1. Here, for instance, monotony starts to affect the operator's state, but by 'trying harder', i.e. by the investment of effort, the primary-task performance level is not yet affected. A second workload redline then arises at the transition from region A2 to A1, where operators are effectively counteracting a reduced operator state. When effort investment is no longer effective, the D-region is entered where performance is affected. When demand increases, starting from the optimal operator state in region A2, the operator's capability of (effort) compensation will be exceeded at a certain moment and a transition from the A3 to the B region takes place. In the B region performance is affected and at the moment that it deteriorates to a minimum level the C region is entered (De Waard, 1996).

In region D (*D* for *deactivation*) the operator's state is affected. In region A2 performance is optimal. The operator can easily cope with the task requirements and reach an adequate, self-set level of performance. In the regions A1 and A3 performance remains unaffected but the operator has to exert effort to preserve an undisturbed performance level. In region B this is no longer possible and performance declines, while in region C performance is at a minimum level: the operator is overloaded. (De Waard, 1996)

If the workload redline is not determined by the point at which performance measures start to deteriorate, but is determined by the point at which region A2 is departed, then performance (primary task) measures alone are by definition not sufficient to determine whether the load is unacceptable. Nevertheless, performance measures remain indispensable in redline research to determine whether workload is in the A region. Again, this is an argument in favour of the use of multiple measurements.

One of the aspects of workload measures that is emphasised in workload redline is the use of absolute versus relative measures. Traditionally, relative measures have been used. With relative measures, task performance, self-reports and physiological measures during baseline performance are compared with the same measures during performance of the task or system that is being evaluated. Some authors claim that absolute measures are required for workload redline (e.g., Wierwille and Eggemeier, 1993). So far, critical values on the SWAT rating-scale have been proposed only by Reid and Colle (1988). However, their mention of the critical SWAT value of 40 refers to the point at which performance begins to be affected (the transition from region A to B). Such a workload redline is a *primary-task workload margin*. This margin is defined as a critical level at which the (primary) task has to be performed. Beyond that point, primary-task performance is affected. Although performance margins can be successfully determined, an absolute criterion for workload itself, i.e. the critical value of a measure denoting that region A2 has been left is, according to De Waard (1996), not feasible. The reason for this is that workload is a relative measure; it is the proportion of the capacity that is allocated for task performance. The amount of resources allocated does not depend only upon task demands, but depends also on capability or willingness of the operators to handle the demands.

The conceptual problems of a workload redline become very prominent in applied settings. In traffic, for instance, the capabilities of the population who drive vehicles vary to a great extent. Novice drivers have to allocate more resources for task performance than experienced drivers. Similar differences in capability exist between young and elderly drivers. Consequently, for the same tasks individuals have their own workload redline. In spite of the problems associated with redline definition, an approach that includes primary-task performance margins as they relate to the cost of maintaining performance is useful in any applied field of workload assessment. Self-report scales and performance measures (for the A to B region shift) are probably the most promising measure groups for this. Physiological indices that are opposed to baseline measurements can be very useful in assessing operator effort and the cost of performance maintenance (De Waard, 1996).

It is proposed that, typically, operating in the C region would be dangerous and that a redline standard should be set at a lower level: at the transition from Region A to Region

B. In most circumstances aviation systems should be operated in Region A (Colle and Reid, 2005).

De Waard (1996) the correctness of putting the redline at the point at which performance is affected and suggested as an alternative the point at which effortful, compensatory processes are initiated. For this, the combination of performance measures with physiology and/or self-report measures could provide a clearer picture of mental workload. Critical levels of measures of mental workload are, however, not attainable because mental workload itself is a relative measure. The resources that operators are willing or capable of allocating to task performance differ between individuals. This makes a *redline* that can be defined according to a critical level in a measure of mental workload an impossibility.

The *workload redline* is the only attempt at determining a quantifiable cut-off, or critical, level for mental workload that could be found in the literature. From the above discussion it can be concluded that this is indeed not a simple, if at all possible, task. This approach however provides a useful framework for conducting further research of this concept, which would add value to applied mental workload settings such as traffic control (air and train), driving tasks and other monitoring-task settings.

Practical application of MWLI in terms of a specific cut-off number for safe operations needs further investigation.

# CHAPTER 6

# CONCLUSIONS

In closing, the aims of the study as well as the results obtained are integrated and a balanced view of the utility of the index is provided. The limitations of the study are discussed and recommendations for further research are offered.

## 6.1 SUMMARY AND CONCLUDING REMARKS

The primary aim of this research project was to develop an objective measure of mental workload for classifying train control centres. In order to fulfil this aim, the definition and concept *mental workload* was studied. The various definitions provided by researchers were put forward and the different components of workload were discussed and contextualised by the model developed by Meshkati (1988).

The study of the existing body of knowledge often proved contradictory, if not confusing. The clear and consistent message, however, is that there is no empirical technique which can be used to accept or reject a single, simplified definition of workload. Workload is therefore acknowledged as a multidimensional construct and it is recognised that different definitions contribute to the perspective of workload.

As far as mental workload measurement techniques are concerned, the categories of workload measurement were identified and the various techniques representative of the categories were discussed. The review of the research indicated that the workload imposed on operators led to changes that manifested in different ways. These were grouped as changes in performance, changes in subjective experience and physiological changes.

The biggest limitation in the reported literature has been identified as the fact that there are no specific guidelines for the application of workload measurement techniques. In addition to this, no specific guidelines exist with respect to the practical implementation of the techniques and the potential value for the train control environment and rail safety.

The Mental Workload Index (MWLI) did show promising results in discriminating between the *low* and *high* workload categories of train control centres and, if used correctly, could be applied for such purposes. It does not appear to have discriminating abilities in the medium range of mental workload. Validation studies need to be done to determine the point or number on the index that will activate warnings that the workload in a particular centre is dangerously high. This methodology for measuring workload provides a more valid technique than the previous index that Spoornet has been applying. Even without cut-off levels for safe operations, the MWLI could be used to identify high workload centres that require special attention, such as closer supervision, the deployment of more TCOs and the deployment of suitably qualified TCOs. These actions would be appropriate to mitigate the potential or existing hazards in these centres.

Translating the results of the research into norms for safe operations is not a simple exercise, simply because of the variable nature of human operators within the system. There is no clear-cut evidence in the research results that points to a specific number or figure for a physiological parameter that would represent a dangerously high mental workload level. It can be conclusively deduced that most of the TCOs who participated in this study carry a high allostatic load, which could predict health problems in the long term.

Finally, it can be concluded that there is not one single method or procedure that is ideal for the measurement of mental workload. There are, however, sufficient methodologies available that can be applied, and if properly validated, will give a scientifically defendable and practical index of mental workload.

The MWLI developed in this study is, as far as could be ascertained from the existing body of knowledge, a unique contribution to the field of human mental workload measurement, since such an application model, which utilises only objectively gathered and observed information, does not exist. Interest in this model, while it was still in the process of being developed, has already been shown by the nuclear industry in South Africa.

## 6.2 LIMITATIONS OF THE STUDY

A limitation of this study, which negatively affected the statistical significance of the results, was the small sample population available to the researcher. The reason for this was that the specific train control system (radio train order), which was the focus of the study, was only used in 36 train control centres. The data was therefore limited to a sample of these centres.

A further limitation was the high cost associated with this kind of research, due to the specialised nature of equipment and the knowledge required to analyse and interpret the data. The cost of research is proportional to the sample size. The project had to be executed within severe cost constraints and the sample for the validation study had to be limited in terms of numbers as well as in terms of geographical spread, in an attempt to limit travelling expenses. The verification field study was therefore limited to a small sample of 20 train control centres. Because the data was collected from a small, non-randomly selected sample of TCOs, it may not be representative of all Spoornet TCOs.

The small sample size has resulted in another limitation. Due to the heterogeneity of the sample it was not possible to randomly select both a control and an experimental group. It was also necessary for the given group to be studied within specified geographical boundaries.

## 6.3 RECOMMENDATIONS

It is suggested that data should be collected more frequently in order to capture the variation and short-term fluctuations in workload that are inherent to a TCO’s job. Data collected every two hours may be insensitive to these variations. Increasing the frequency of data collection would probably not be acceptable to either management or TCOs. These individuals are conducting safety-critical work, so there is also an increased risk of distracting them from their work if more is asked of them. A medium- to high-fidelity dispatching simulator would be more suitable to this type of research. A simulator would enable researchers to control workload and other conditions of the work environment in order to see their effects on the TCO, but as discussed, several obstacles were encountered when attempting to create a high-fidelity simulator specifically for RTO. This possibility is worth exploring if the research could be undertaken in a collaborative effort with an academic or research institution.

It is further suggested that a validation study of the MWLI be undertaken on a larger sample in order to allow for more statistically significant analyses. Once this has been done, the development of a *workload redline* for TCOs should be considered.

As it would entail experimental conditions, the workload-redline development should be performed under controlled conditions in a simulated environment. An *in situ* approach would be invasive and could potentially pose safety risks for the TCOs. It is suggested that the workload redline is approached from two perspectives, namely, from a performance as well as a health perspective. The following approach is suggested:

- The MWLI factors should be 'loaded' in order to recreate conditions that would result in a high calculated MWLI. Under these conditions the physiological parameters used in the field study should be measured to determine the physiological load placed on the TCO.
- Simultaneously, a measure of performance should be performed. Under simulated circumstances it is possible to develop performance indicators for the primary tasks of TCOs. Alternatively, a secondary-task measure of spare mental capacity could be considered.
- The same group of TCOs could then be subjected to conditions resulting in a low calculated MWLI and the concurrent physiological parameters and performance measure could be measured.
- Incremental increases of the MWLI with its associated physiological and performance measure for the same group could be performed.
- With the assistance of physiology experts a 'healthy' threshold of the physiological parameters could be determined and correlated with the MWLI value.
- Similarly, a corresponding MWLI value for deterioration in performance could be established.
- These MWLI values could then be correlated and a redline MWLI value determined.

Future research could also focus on the validation of the MWLI in other safety-sensitive and control or monitoring environments.

The ultimate value of the MWLI would be determined if a longitudinal study of train control incidents (in the RTO environment) could prove a downward trend.