CHAPTER 2

LITERATURE STUDY

The elements of mental workload that are addressed in this chapter relates to the definition of the concept, its association with other related operator states and the variables that influence mental workload. An operational definition of mental workload is developed, based on the existing literature and in the context of this definition, the various mental workload assessment techniques are discussed. In closing, relevant railway related research is referred to in order to create the context for the study.

2.1 THE DEFINITION AND DESCRIPTION OF MENTAL WORKLOAD

2.1.1 The concept *mental workload*

The term *workload* covers a broad spectrum of human activity. *Mental workload* focuses on activities that are primarily mental or cognitive in nature and may involve physical coordination, but excludes activities resulting only in muscular fatigue.

The practical importance of the concept of *human mental workload* was established several decades ago during the investigation of human-machine systems such as ground transportation, air traffic control and process control. The theoretical development of the field can be traced back to a NATO conference and the subsequent text, Mental Workload (Moray, 1979). Since that seminal volume, many studies have been conducted on the theoretical underpinnings, assessment techniques and real-world implications of mental workload in a variety of work domains (Parasuraman and Hancock, 2001).

Mental workload should not be treated as a pure science – it is an applied science or technology. It should be viewed as a theoretical construct that is defined differently by engineers and psychologists. Engineers would typically put emphasis on operational definitions based on time available to perform a task, while psychologists would tend to emphasise the information processing aspects and define it in terms of measures related to channel capacity and residual attention. Physiologists may rather emphasise considerations of operator stress and arousal (Wierwille, 1988).

The intention of this study is to integrate these views into an operational definition in order to assess the relevance and effectiveness of techniques that objectively measure mental workload. From a review of the literature it is clear that the multifaceted nature of human mental workload has been emphasised by various researchers. The literature presents the reader with a maze through which careful navigation is necessary in order to attain clarity and avoid further confusion. In the search for a definition of workload, it is clear that different perspectives have led researchers to define workload in various ways. A number of definitions are reviewed here in order to develop a working definition for the purposes of this study.

Not only is there no single, commonly accepted definition of workload but there are many conflicting concepts of what workload comprises, and often the term is used without any definition at all. There is also the difficulty of differentiating workload from stress, which has many features in common with workload. Often the terms workload and stress are used interchangeably (Meister, 1985).

When the concepts of fatigue and situational awareness are added to those of workload and stress, one realises that the territory is complex and problematic, especially for newcomers to the research field. Researchers are rarely confronted with only one of these issues. All of these issues present themselves in the environments in which operational research is performed. These environments are always safety-critical in nature. One only has to think of the environments where mental workload research originated, namely air traffic control, aircraft handling and aerospace operations to realise that these operators are all exposed to stress and fatigue (most of these are 24-hour operations), they have to be situationally aware and they could potentially become mentally overloaded.

This section therefore aims to derive an operational definition of mental workload that is grounded in the research of the past three decades. The overview of the research on mental workload does not claim to be exhaustive but covers the most important conceptual work in the field.

In an attempt to create some order among the multitude of mental workload definitions, they have been grouped together based on the premise from which the definition was developed. As many of the definitions were based on more than one premise, this proved fairly challenging.

2.1.2 Mental workload as a function of capacity

Mental workload is often conceptualised as the interaction between task demands and the *capacity* of the operator. Thus the ratio between demand and capacity determines the level of workload. *Capacity* is determined by the operator's skills and training but may also be influenced by stressors such as fatigue, noise etc. Task demands are determined by the number of tasks to be performed, the amount of attention needed and the time available.

Human operators could be viewed as communication channels with limited and fixed *capacity*. It has been theorised that as long as total capacity limits are not exceeded, capacity can be divided at any time among concurrent processes and activities. Another view is that the capacity needed for a particular activity is expressed as the time taken by that activity, with a more complex activity taking more time. In this case, mental load then equals input load, which can be largely defined in terms of task variables and can be directly measured according to the time taken to perform activities. A considerable amount of capacity may remain unused in a single reaction to task demands because only a limited number of mechanisms may be involved in that reaction. It can be argued that the more complex the task, the more mechanisms are involved and consequently the more capacity is used (Senders, 1979).

Capacity does not remain constant however. Available capacity increases with the task demands as a result of physiological mechanisms related to arousal. This reflects the physiological components and it can be argued that mental load equals the effort needed to transform the actual task demand into an adequate work result. Factors such as 'willingness to spend capacity' may moderate the amount of effort operators are prepared to spend on a specific task (Moray, 1979).

Workload measurement is the specification of the amount of capacity used. In this definition workload is not solely task-centred. Mental workload depends on the demands in relation to the amount of resources operators are willing or able to allocate and is therefore a relative concept (O'Donnell and Eggemeier, 1986).

Related to the concept of capacity but referring to operators as adaptive agents, Parasuraman and Hancock (2001) propose that mediating factors should be considered rather than assuming that mental workload is an intervening construct that reflects the relation between the environmental demands imposed on operators and the *capabilities* (or capacity) of operators to meet those demands. Workload may be driven by the taskload imposed on human operators from external environmental sources but not deterministically so. This is because workload is also mediated by individual responses to the workload and varying skills levels, task management strategies and other personal characteristics. Although such definitions immediately suggest the notion of adaptation by human operators, most studies have failed to explicitly examine this aspect of workload. The majority of studies have been concerned with empirical evaluations of the effects of various taskload factors on various measures of mental workload – whether based on performance outcome, physiological response or subjective report.

2.1.3 Mental workload defined in terms of *experienced load*

A simplistic definition of workload is that it is a demand placed on humans. This definition attributes workload exclusively to an external source. However, workload can be better defined in terms of experienced load. With *experienced load*, workload is not only task-specific but also person-specific. Besides individual capabilities, the motivation to perform a task, the strategies applied in task performance, mood and operator state also affect experienced load. Therefore, workload depends on the individual, and as a result of the interaction between operator and task structure, the same task demands do not result in an equal level of workload for all individuals (De Waard, 1996).

Workload has been described as multidimensional because it can be viewed either as an input or as an output or consequence. As an input, workload is represented by stimuli that load the operators in the sense that they are caused to bear a burden. As an output, workload affects not only performance but, when operators are part of a larger system, it impacts (usually negatively) on the system itself. As an intervening variable, workload is the operators' internal experience of difficulty and discomfort, their recognition that they are experiencing a load and the strategies they employ to overcome this. It can be said that as an intervening variable, workload alters the human information-processing system by allowing the operator to adopt strategies to cope with the experienced workload. Workload can thus be viewed as a feature of the system (even when it is the operator's own incapacity) that forces operators to work harder. It can also be viewed as their feeling of being stressed and of having to work harder. The effect of these factors may cause them to make errors (Meister, 1985).

Mental workload also represents the cost incurred by a human operator to achieve a particular level of performance. An operator's subjective *experience of workload* is the product of many factors, in addition to the objective demands imposed by the task. Thus, workload is not an inherent property but rather the product that emerges as a result of the interaction between the requirements of a task, the circumstances under which it is performed, and the skills, behaviours and perceptions of the operators themselves (Hart and Staveland, 1988).

This internal experience of difficulty and discomfort is the basis of the stress created by mental workload. This stress is in most cases, especially for TCOs, a time-related factor.

2.1.4 Mental workload defined in terms of *time load*

The definition below relates mental workload to time but then considers, as an additional factor, limited capacity which results in competition for attention:

i) Tasks must be performed within a certain length of *time*. The degree of workload is the percentage of that time which the operator actually has to perform those tasks. This concept focuses on time availability. A high degree of workload is experienced when the task requires most or all of the available time.
ii) The operator has limited capacity, usually conceptualised in terms of attention. When

the operator must perform multiple tasks within the same *time period*, there is competition for the operator's attention. That competition 'loads' the operator (Meister, 1971).

In a similar vein, Reid and Nygren (1988) propose that mental workload should be viewed as a multidimensional construct that can be explained largely by three component factors: *time load, mental effort load* (which relates to capacity) and *psychological stress load* (which seems to be more related to the consequences of the previous two). Again, as with Meister's definition, workload is defined in terms of two factors: time and capacity.

Time load involves both the time available for a task and task overlap. If the time required to perform a task exceeds the time available, the operator encounters a time-load problem. If the tasks that the operator has to complete start to compete for time resources, the operator will be forced to evaluate the tasks for priority and allow the performance of some tasks to deteriorate and/or their completion to be delayed.

The channel capacity concept deals with task factors such as difficulty, complexity and required effort. This dimension assumes that the human operator has limited capacity. Performance of one task may consume a certain amount of an operator's resources, while another task may consume other resources. The implication is that resources that are not expended in task performance are either held in reserve to be used for other tasks or used for extra effort in accomplishing the current task. This dimension therefore refers to *mental effort load*.

The third component factor deals with the concept of *psychological stress* and encompasses a number of operator variables such as motivation, training, fatigue, health and emotional state. This dimension is broadly defined as anything that contributes to an operator's confusion, frustration or anxiety and is referred to as psychological stress load.

2.1.5 Mental workload as a function of *demand and supply*

Workload can be defined in terms of the relationship between *resource supply and task demand*. It is argued that operator workload is directly related to the extent to which the tasks performed by the operator utilises the limited resources (Wickens, 1992).

Figure 2.1, reproduced from Wickens and Hollands (2000), shows the supply-demand relationship between the important variables in the workload model. The resources demanded by a task are shown on the horizontal axis. The resources supplied are shown on the vertical axis, along with the level of performance. If adequate performance of a task demands more resources from the operator than are available, performance will break down, as shown on the right of the figure. If, however, the available supply exceeds demand, as shown on the left, then the excess expresses the amount of reserve capacity. According to this model, changes in workload may result either from fluctuations of operator capacity or from changes in task resource demands.

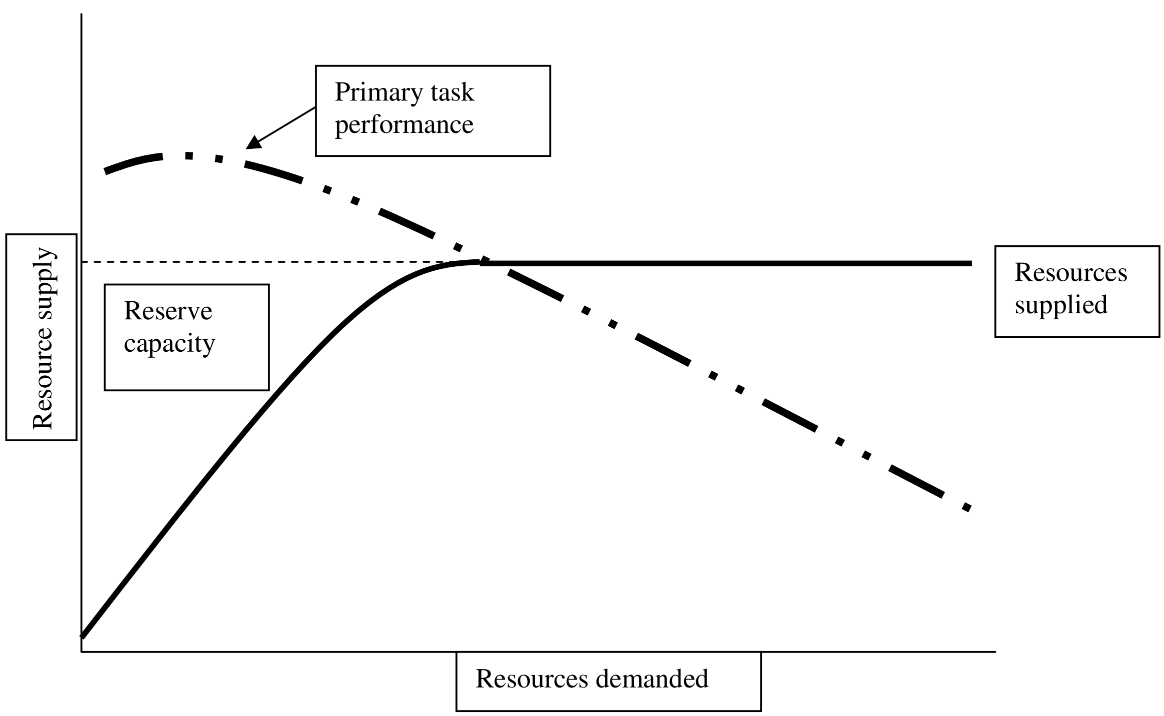


Figure 2.1: Schematic relationship among primary-task resource demand, resources supplied and performance (Wickens and Hollands, 2000).

2.1.6 Mental workload versus taskload: an argument for a continuum

The term *taskload* is often encountered in the literature on mental workload. In some instances the term is used to refer to the task demands placed on the operator but it is often used as a synonym for mental workload. It would seem more likely that the distinction between the two is a continuum rather than an absolute distinction.

Simply stated taskload is what the operator is confronted with in the execution of duties i.e., the number of tasks that have to be carried out. Mental workload also takes into consideration what the operator does to cope with the work. Workload is often defined as the interaction between the demands of the task and the ability of the operator to meet those demands (Reinach, 2001).

Hilburn and Jorna (2001) distinguish between the two concepts as follows: Taskload is the demand imposed by the task and mental workload is the operator's subjective experience of that demand.

A number of studies have attempted to identify taskload indexes for Air Traffic Controllers (ATC). Out of many prospective taskload indexes, the number of aircraft under control (i.e., traffic load) has shown the clearest predictive relationship to workload measures. However, traffic load alone does not accurately capture the total load imposed by what is happening in the airspace. Factors such as skill, training, experience, fatigue and other stressors all mediate the relationship between task demands and workload experienced by the controller.

The distinction between taskload and workload in the ATC context can be depicted graphically as follows:

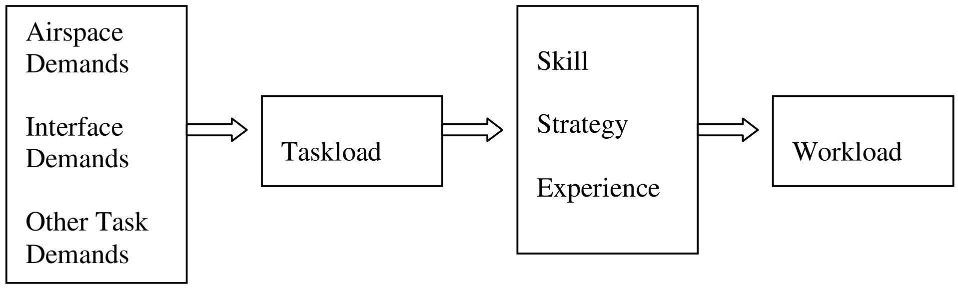


Figure 2.2: Taskload and workload determinants in ATC (Hilburn and Jorna, 2001).

The ratio between demands and capacity therefore determines the level of workload. Capacity is determined by the skills and training of the operator but may also be influenced by stressors such as fatigue, noise etc. Task demands are determined by the number of tasks to be performed, the amount of attention needed and the time available (Hancock and Meshkati, 1988).

The above distinction between mental workload and taskload is very similar to that of Hilburn and Jorna (2001) and shows that something interacts with the task demands (capacity) in order for taskload to become mental workload.

To place taskload and mental workload along a continuum may seem mechanistic but this means of definition would appear to add both clarity and understanding to the construct of mental workload.

2.2 MENTAL WORKLOAD AS RELATED TO OTHER OPERATOR STATES

Although the states of workload, stress and fatigue may occur simultaneously and are sometimes difficult to distinguish from one another, they should not be regarded as synonymous and their differences should be clearly understood. A proper distinction between these concepts is not only important for theory building but also for the restructuring of the work environment (Gaillard, 2001).

The confusion between these concepts originates in their poor definitions. As discussed in the previous paragraphs, mental workload may refer either to the objective demands imposed by the task (e.g., complexity or pacing) or to the subjective judgement of the operator with regard to the task demands. In most theories, workload refers to the limitations in the information processing capacity of the operator. It may, however, also encompass feelings of work pressure, which have a more emotional connotation.

2.2.1 Workload as related to stress

The concept of stress has a large variety of meanings:

- An *input* variable referring to either work demands (difficulty and time pressure), emotional threat (accidents) or adverse environments (noise and sleep loss).
- An *output* variable referring to a pattern of behavioural, subjective and physiological responses, often labelled *strain*.
- A *state* in which we feel strained, pressurised and threatened on the basis of a subjective evaluation of the situation.
- A *process* that gradually results in a dysfunctional state, degrading the work capacity and the potential to recover from work.
  (Gaillard, 2001).

Load is defined as the demand placed on the operating resources of a system. Stress and strain are indications of the effects of workload on individuals. Mental workload does change an individual's mental shape in terms of the resources they make available. Physiological variables change with mental activity and therefore seem to be well suited as indicators of changes in mental workload demand (Rasmussen, 1979).

Workload should not be confused with stress. Some of the negative effects of stress on performance include narrowing of the attention span, forgetting the sequence of actions, incorrectly evaluating situations, slow decision making and failing to carry out decisions (Meshkati and Loewenthal, 1988).

Hart and Bortolussi (1984) draw the following distinction between workload and stress in their assessment of sources of workload. Workload was evaluated as the change in effort required when dealing with tasks, whereas stress was regarded as an experience that is the consequence of changes in the tasks. However, the results of their study indicated closely correlated estimations of workload and stress.

2.2.2 Workload as related to fatigue

Fatigue is interpreted as a response of mind and body to a reduction in resources resulting from the execution of a mental task. It is also considered to be a warning of the increasing risk of performance failure. Fatigue may also refer to a subjective complaint that encompasses a general feeling of lack of energy that is not necessarily related to the amount of work (Gaillard, 2001).

Although two types of fatigue are referred to, only passive fatigue is relevant in the train control context:
Active fatigue results from continuous and prolonged, task-related, perceptual-motor adjustment.
Passive fatigue relates to system monitoring with either rare or no overt perceptual-motor response requirements. Closely related to vigilance, this form of fatigue develops over a number of hours during which individuals appear to be nothing at all (Desmond and Hancock, 2001).

Fatigue represents one form of stress. Fatigue is considered as the product of some aspect of environmental input on the individual, such as heat, noise, poor ergonomic design of the work environment and number of hours worked. Fatigue is evident in performance decrements of operators.

This is one of the reasons why *shift worked* has been included as a mediating factor in the MWLI. This information can be collected objectively and it represents an input stressor on the individual.

2.2.3 Workload as related to situational awareness

Situational awareness has not received as much attention as stress and fatigue, but it is worth mentioning in the mental workload context.

Situational awareness is defined as the perception of elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future. It is a uniquely cognitive phenomenon which supports action but is not part of the action itself. It applies to a dynamic, changing environment, and the nature of the situation of which awareness is maintained, needs to be specified.

There is an interaction between workload and situational awareness. Maintaining a high and accurate level of situational awareness is a resource-intensive, cognitive process. Thus, on one hand, one cannot gain accurate situational awareness without expending resources, and this may in turn compete with other concurrent cognitive tasks. On the other hand, heavy concurrent task demands may divert resources from the maintenance of situational awareness.

This relationship between situational awareness and mental workload is mediated, to a certain extent, by the degree of operator expertise or skill. Skilled operators can generally preserve situational awareness with lower resource cost (Wickens, 2001).

The terms mental load, fatigue and arousal are often viewed as synonymous with stress. Despite attempts to give unique definitions to these terms, it is difficult to see what their distinguishing features are.

The interrelatedness of the concepts of workload, stress and fatigue is clearly complex, often confusing and a potential pitfall for researchers. From the outset of this study it was decided to simplify the approach to workload as far as practicably possible but not to lose sight of the other related factors. By adopting this approach, this study could be criticised for simplifying the concept of workload, a calculated risk that had to be taken in order to provide a user-friendly, fit-for-purpose solution to an operational challenge.

2.3 AN OPERATIONAL DEFINITION OF MENTAL WORKLOAD

2.3.1 Comments

Considering the wealth of information about mental workload that has been generated over the past three decades and the variety of definitions from different perspectives, it is disappointing that consensus has not been reached among researchers as to how mental workload should be defined. This makes it difficult for new-generation researchers. Moray's 1979 foundation, as the starting point, has remained relatively unelaborated upon. Although this is an excellent foundation, there is an expectation that advances in terms of reaching an agreement on the definition of mental workload would have been made.

As far back as 1988, Wierwille (1988) expressed the opinion that mental workload research, having made initial strides, appeared on the verge of becoming mired in its own details. It is important that momentum in this research is not lost. It is doubtful whether, in terms of finding a consensual definition of mental workload, there has been any progress since this statement was made.

As a final comment on the definition of mental workload, it is important to continue questioning the very concept of mental workload. After many years and many hundreds of empirical investigations, there is still no satisfactory, consensual definition of workload. A deeper understanding of mental workload is essential for the following reasons:

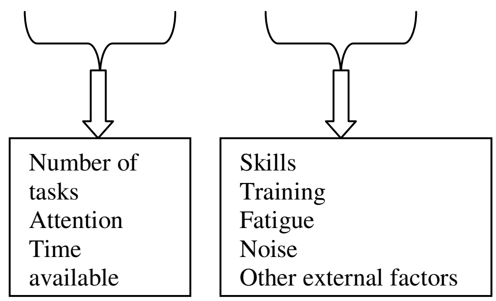
- Theoretical motivations – we seek to understand and relate the phenomenon of mental workload to related cognitive constructs such as attention, effort and resources.
- Practical motivations – the nature of work in the developed world has changed from being physical to primarily cognitive. The 21st century is an era of intelligent automation and there is a need to assess the load of such information-based work. It is important to know how much mental load is too much and may prove hazardous or stressful. There is also a need to know how little is too little, so that individuals are sufficiently challenged to sustain useful levels of output.
- Technological motivations – mental workload must be assessed as an indicator for use in complex technical systems, to dynamically change the nature or the demands of the work at hand. As almost all forms of work now involve human-computer interaction,

this strategy becomes progressively more feasible, and indeed crucial, for complex systems.
(Parasuraman and Hancock, 2001).

2.3.2 Synthesis of definitions

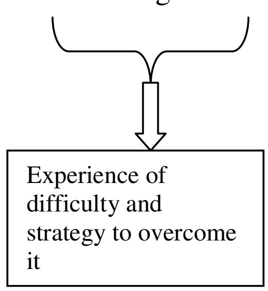
When the four categories of definitions are considered, the following key elements are evident:

- Mental Workload is the interaction between: *task demands* and *capacity*



- *Experienced load* – is an intervening variable

Input ⟹ Intervening variable ⟹ Output/Performance



- *Time load* – relates to capacity and stress

- *Demand and supply* - relate to task demand and resource supply

These elements fit neatly into the very comprehensive and cohesive model proposed by Meshkati (1988) (see Figure 2.5). This model maps the major interrelationships of the interacting variables of mental workload. Meshkati's model is the most comprehensive single model of workload found in the reviewed literature and it ties together most definitions of mental workload.

Meshkati's model consists of two sections – Causal Factors and Effect Factors – each with two primary component groups.

- The two primary groups of Causal Factors are the following:
    - task and environmental variables and
    - operators' characteristics and moderating variables.
- The two primary groups of Effect Factors are the following:
    - difficulty, response and performance variables and
    - mental workload measures.

This model has been found to be very helpful in integrating the various concepts relating to mental workload but is somewhat complex to use when it is necessary to explain the concept of mental workload to an uninformed audience. The model proposed by Jahns (as cited in Johannsen, 1979) appears to be a good foundation and one that is both sufficiently comprehensive (with a few additions) and easy to use.

Jahns (as cited in Johannsen, 1979) has proposed a concept for the assessment of workload that divides the broad area of human operator workload into three functionally relatable attributes:

Input load ⟹ Operator Effort ⟹ Performance

In terms of this model the following can be said:

- Major sources of input load may be separated into three categories: environmental, design-induced or situational and procedural. Environmental variables are noise, vibration, temperature, etc. Design variables are, for example, the characteristics of displays and control devices, crew station layouts and vehicle dynamics. Procedural variables include briefing and instructions, task sequencing and task duration.
- Operator effort depends on a number of factors including the input load and the performance requirements of a given task, as well as internal goals and motivation. The operator state depends on many factors. Relatively stable factors would be

psychophysical characteristics, general background and personality. Fluctuating factors would be experience, motivation and attentiveness.

- Internal performance criteria are maintained by human operators and influence their tolerated error levels. These criteria depend on factors of the operator state (e.g., motivation), performance requirements and instructions.

Workload is the umbrella concept which includes input load and operator effort.

Using the elements in Jahns' model as the foundation and relating it to the definitions of mental workload referred to earlier in this chapter, the following working definition of workload is proposed as a basis for the development of the MWLI.

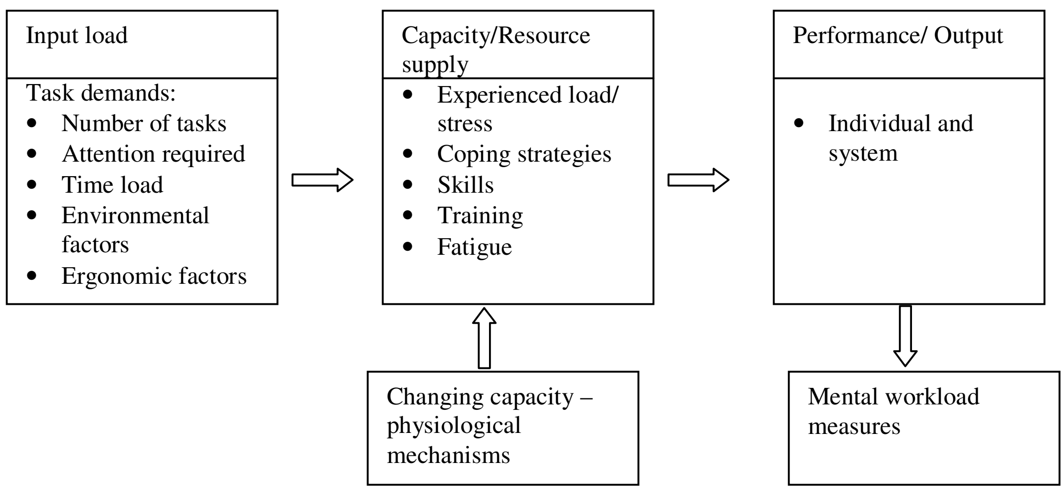


Figure 2.3: Flow diagram of an operational definition of mental workload

2.4 VARIABLES THAT INFLUENCE MENTAL WORKLOAD

Mental workload, like fatigue and occupational stress, is one of the factors in the human-machine/system interface that needs to be managed in order to sustain safety and productivity. Now that an operational definition for mental workload has been established, it is necessary to briefly refer to those variables that could influence mental workload.

There are two categories of variables that will be discussed:

- individual factors
- task factors

The Ergosystem provides a framework for relating the factors that influence mental workload. For any system to function effectively, all the elements must be integrated in a functional manner. The nature of any system is such that any change in a sub-system will result in disequilibrium of the system.

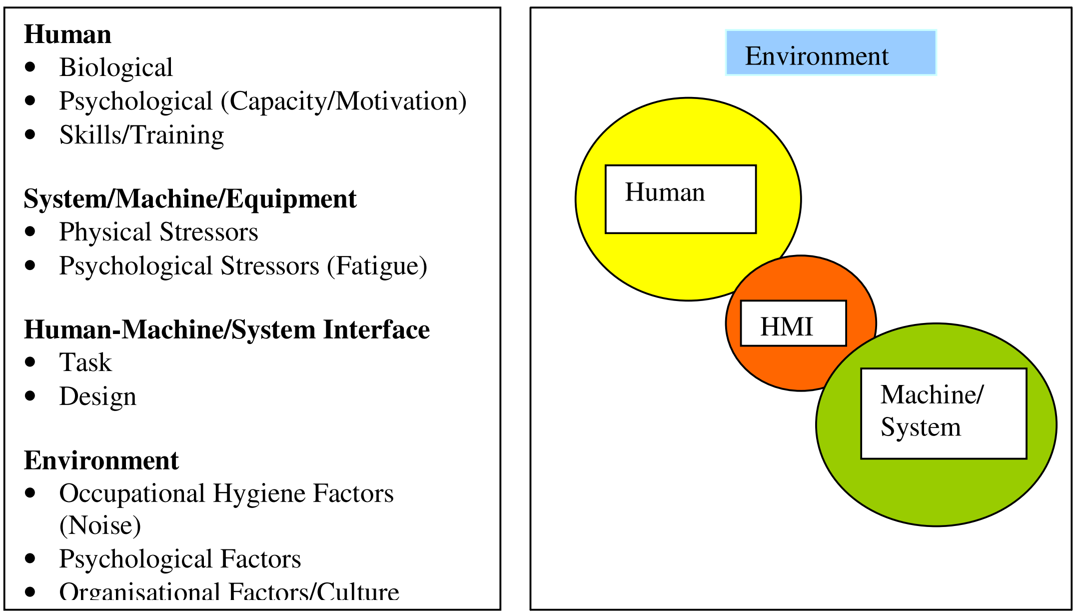


Figure 2.4: The Ergosystem (Adapted from Van Tonder, 1999)

2.4.1 Individual factors affecting workload

A common feature mentioned in the assessment of mental workload is the differential sensitivity of measurement methods to the factor individual differences or personality traits. This has been acknowledged as an influential variable. It manifests itself in the subjective assessments and physiological responses of operators (Meshkati, 1988).

In this regard, it is worth noting the comments of a few experts. According to Firth (1973), in the real-life working environment individual differences in operators' characteristics greatly influence an individual's information-processing capacity. These differences arise from a combination of past experience, skill, emotional state, motivation and the estimation of risk and cost that may be involved in performing a task.

Meshkati and Loewenthal (1988), studied the role of personality in performance decrement and attention. Their results indicated that individuals who scored high on a distractibility scale (i.e., extroverts) found it difficult to maintain a uniform mode of performance. This group of subjects exhibited increasing lapses in attention, while the introverted group failed to show any evidence of a decline in attention.

Damos (1988) has done a comprehensive study of the research available on individual differences in the subjective estimates of workload. She has identified the following variables as significant:

(a) Personality traits and behavioural patterns

Little research has been conducted that examines the relationship between mental workload and personality traits.

To date, the only behavioural pattern investigated is the Type A coronary-prone behaviour (Friedman and Rosenman, 1974). Individuals who have high scores on tests measuring coronary-prone (Type A) behaviour are characterised by an extreme sense of urgency. As a result, they prefer a more rapid work pace and tend to perform better on tasks that do not have a deadline than individuals who have low scores on tests of coronary-prone behaviour (Type B individuals).

It appears that, under dual-task conditions, Type A subjects generally experience less effort (or mental workload) than Type B subjects and the same or more effort under single-task conditions.

Meshkati and Loewenthal (1988) emphasised that cognitive style, i.e., learned thinking habits, may have an influence on behaviour, subjectively experienced workload and perceived task difficulty.

(b) Motivation

The effects of motivation on the differences between subjective workload assessments and performance have been reviewed by Vidulich (1988). It was predicted that although increased motivation would lead to better performance, it would do so by encouraging subjects to expend greater effort, thereby increasing the experienced workload.

(c) Mental Capacity

Mulder (1979) reviews the difference between the mental capacity expended by the main task and the total available capacity. It is assumed that the difference between the capacity of the operator and the load imposed by a specific task is called reserve capacity. The total available capacity is not constant but varies with the operator's state of arousal. The more effort (i.e., processing capacity) an operator expends on the task, the less capacity remains available for other tasks or circumstances that may demand attention.

(d) Gender and time of day

Hancock (1988) reviewed the effect of two operator-specific factors, namely gender and time of day, on subjective estimates of mental workload. Significant differences were found between males and females. The study indicated a higher intolerance among female subjects for the repetitive and boring task examined. Should this represent a gender difference in tolerance to work under conditions of underload, this study may present an important finding, specifically from a selection perspective in safety-sensitive operations, such as train control or monitoring. No significant effects were found for time of day.

(e) Other Factors

Additional factors emphasised in the study of Meshkati (1988) include the importance of the individual's state of arousal, sensory capabilities and the level of training and experience. All of these factors may have a considerable influence on perceived mental load, which is reflected in workload measures.

In research it is imperative to state in the design phase how individual differences in mental workload assessment will be dealt with, rather than ignoring their potential effects. In this particular study a conscious decision to eliminate personal factors from the MWLI was made. This was done in order to obtain a predictive value per train control centre based on the 'average' operator.

2.4.2 Task variables that affect workload

An interesting concept that could only be found in one literature source is the notion that workload researchers have paid little attention to the sources of variance in the independent variables and, instead, have focused their attention and experimental conclusions on the dependent variables, i.e., measured indications of workload. The independent variables referred to are those task variables that affect mental workload (Cilliers, 1992).

The variables quoted here refer to the military aviation sector, and those that are potentially relevant to the train control environment will be discussed briefly.

(a) Task criticality

Task criticality in the military context can be viewed as a function of the task content and the effect of execution or non-execution on the mission or task as a whole. The criticality of a task increases with the unpredictability and the severity of the effect.

Analogous to the military aviation scenario, the various tasks or phases in the train control environment that constitute task criticality are the following:

- Planning of shift activities – by studying the daily train plan, TCOs can get an overview of which trains are in their section, which trains have to depart, at what time and from where and which other activities (maintenance and shunting) are planned for the day. These trains are then plotted on train schedule sheets and mapped according to the predetermined running times between sections. Crossing places are predetermined and mapped on the schedule.
- Communicating by radio with train drivers to obtain their positions and provide authorisations to proceed up to a predetermined kilometre point.
- Communicating with and monitoring conversations of track maintenance teams.
- Answering telephone calls from other staff (maintenance and commercial) regarding train movements.
- Writing of reports on activities/incidents such as train failures and shunting activities.
- Capturing of incidents such as derailments, infrastructure problems in RIMAS (Risk Information Management System).
- Termination of trains in their section (when a train enters the next train controller's section).
- Handing over tasks on completion of the shift.

(b) Environmental factors

According to Cilliers (1992), environmental factors also influence mental workload.

In the train control environment the following environmental factors are relevant:

- Design and layout of train control centres – seat, desk, ambient noise and temperature. Due to the multitude of possible variations of this factor and to the task team not considering this to be an important contributing factor, this was not included in the MWLI.
- Topography – the nature of the terrain and landscape through which the section runs (see description in Chapter 6, Table 6.4). This was included in the MWLI as variations in this factor impact on the ability of train drivers to keep trains running on time which could impact on the TCO's scheduling and planning.

(c) Amount of information and complexity

The amount of information to be processed by the pilot can be defined as a direct result of the tasks imposed by the mission as well as those tasks imposed by emergency conditions and the tactical environment. A portion of the information to be processed by the pilot is predictable and another portion is highly unpredictable.

In the train control environment a similar situation exists. The tasks of TCOs are highly structured and predictable when normal operations prevail. It is when an abnormal situation arises (i.e., an accident, derailment or communication failure), which requires quick problem solving and the use of discretion and prompt response, that the information to be processed becomes excessive and complex, which could lead to mental overload and increase the risk of errors.

(d) Task structure (High versus Low)

According to Cilliers (1992), task structure can be described as a function of design of the man-machine interface that influences the normal and emergency procedures, and the tactical environment which elicits the task response.

In both the aviation and train control environments, the task structure is high. This means that work is done in a proceduralised way. For instance, the issuing of an authorisation occurs according to a set step-by-step procedure. Train planning (which train has priority when crossing another train, for example) occurs according to a set of rules. It is only when equipment such as radio communication fails and abnormal working procedures are activated, that tasks become less structured, time pressure increases and the knowledge and discretion of TCOs are stretched. This is often due to the fact that TCOs have not been properly trained and/or the training has not been regularly refreshed when it comes to fall back or emergency procedures. When certainty and predictability decrease, workload increases.

(e) Time response

In terms of time availability for task execution, a distinction should be made between routine and emergency conditions. Under emergency conditions, time available will be a function of the number of emergencies (or out-of-the-ordinary tasks) to which TCOs must attend, as well as a function of the criticality of the emergency condition.

(f) Equipment and design

The radio train control environment is significantly less sophisticated than that of a cockpit but the effect of the environment on the workload of TCOs is no less important. The full spectrum of ergonomic factors (i.e., the adjustability of desks and chairs, ambient noise, temperature, illumination and the reliability of equipment) was considered when analysing the tasks of TCOs. Initially this factor was included in the MWLI but was later discarded by the task team due to the variability of train control centres and the potential for subjective interpretation.

(g) Task Novelty

It has been speculated that a reduction in task novelty will result in a reduction in the degree of workload experienced (Cilliers, 1992). This factor partially explains why more incidents occur when TCOs have to use fall back procedures, as may be necessary when equipment fails and the normal procedures cannot be followed. This provides a good rationale for regular refresher training for TCOs in abnormal working or fall back procedures, which will reduce the novelty of the task. The importance of this factor is emphasised by the inclusion of *number of years experience* as a moderating factor in the MWLI.

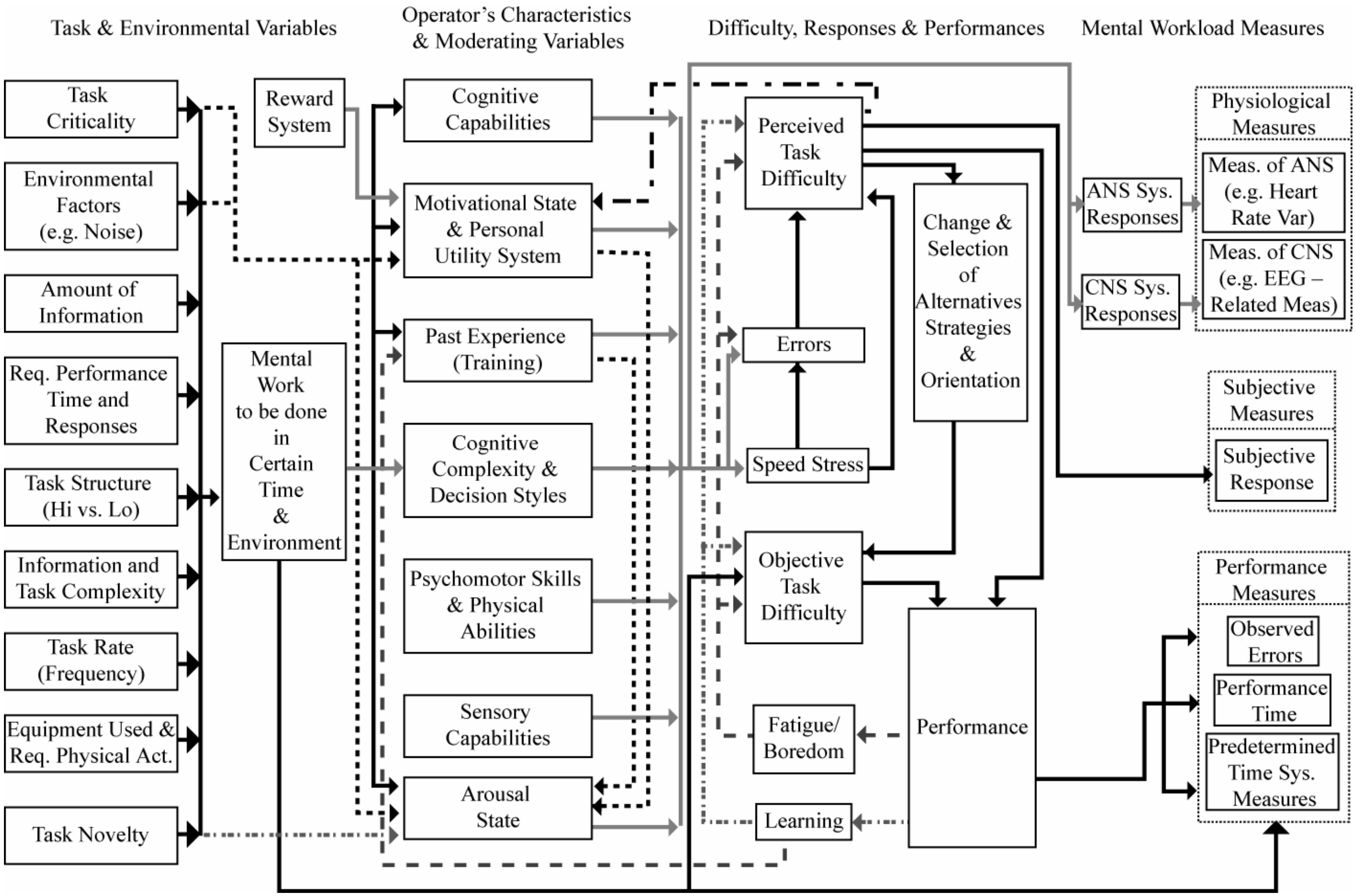


Figure 2.5: Major Components of Comprehensive Mental Workload Model and Related Assessment Variables (Meshkati, 1988)

2.5 APPLICATION OF WORKLOAD THEORY

As the previous sections indicate, the concept of mental workload is considered to be highly complex and cannot be simplified or reduced to a single dimension. This has lead to a situation where there appears to be an inability among the researchers in the field of workload to reach agreement on workload definitions, research procedures and the degree to which this information can be generalised from one context to another.

It has become clear that few researchers provide clear explanations as to the areas of application of their workload research. A contributing factor may be the fact that most of the studies were conducted for US and other military institutions, with military security classification, which has denied South African researchers access to this information. It appears however that much of the published research focuses on laboratory experimentation or simulated situations, and that the procedures used by researchers cannot be made applicable in a similar fashion in applied environments.

The study of Cilliers (1992) is the only South African study on mental workload that could be found. A search of the literature for the application of mental workload research principles or measurements rendered no returns. Railway companies that commission research, normally classify the information and it is either not available for public access or access is limited to a few published articles. Apart from a number of comprehensive Federal Railroad Administration (FRA) research reports, information on how other railway companies address the problem of mental overload in the train control or other operational environments is therefore limited to personal communications with individuals involved in this work. In addition, access to research is often dependent on membership of professional research bodies.

2.5.1 Application in safety operations

It would appear that individuals are most reliable under moderate levels of workload that do not change suddenly and unpredictably. Extreme levels of workload increase the likelihood of human error as a result of the operator’s inability to cope.

When workload is too low, the operator may become bored and lose concentration. This may also lead to increased errors.

2.5.2 Application in automation

Tsang and Johnson (1987) maintain that there are a number of reasons for incorporating higher degrees of automation in complex aircraft systems (or any monitoring and control environment). The most important reasons being the increase in the level of safety and the reduction of pilot (or operator) workload. Another important consideration is that it would free the pilot involved in the system from repetitive processes which could best be performed by automated devices. This would allow the pilot to concentrate on activities such as planning and decision making.

The closest that the Spoornet train control environment has come to the employment of automated devices is the use of computer-based systems such as TWS, which is programmed to disallow a planned train movement that is in conflict with a previously planned movement.

The following is concluded from the existing body of knowledge of mental workload:

- A variety of different definitions of workload are being used by researchers.
- Different perspectives (e.g., engineering, psychology or physiology) may lead to different definitions of workload. The importance of the context of workload measurement has been emphasised.
- Changes in the definition of workload have occurred over time. Initial emphasis was on the time dimension. As a result of increased research by psychologists, the information-processing aspects received more attention, and issues related to channel capacity and residual attention were highlighted. Psychologists have also investigated the concepts of arousal and stress as these relate to workload.
- Any attempt to decide which of the workload definitions is most accurate would appear to be unproductive. There is no empirical technique for providing a single definition of a multidimensional construct such as workload. Each definition is useful and contributes to the understanding of workload.

2.6 AN OVERVIEW OF MENTAL WORKLOAD MEASUREMENT TECHNIQUES

In this section a general overview of the existing mental workload measurement techniques is given. The criteria for the selection of mental workload assessment techniques are also discussed and a set of criteria that is applicable for this study is provided.

In the literature, the measurement techniques are divided into the following groups:

- measures of primary task performance
- secondary measures of spare mental capacity
- subjective rating techniques
- physiological indicators of mental workload

These same groups are used in the overview. All the techniques that could be found in sources available in the public domain will be referred to, rather than singling out a select few. When this study was initiated this was considered to be the most effective approach. Before the decision was made to develop the MWLI, the existing techniques were critically evaluated in order to find an existing tool that would address the requirements of this particular study. Each of the techniques is commented on in terms of its applicability and appropriateness for this study.

From the literature on mental workload assessment techniques, it is evident that a great deal of information has accumulated in the period between the mid 1970s and late 1980s. During this period, significant changes were occurring in the theoretical conceptualisation of workload itself. It became increasingly clear that workload was a multidimensional rather than a unitary construct. Workload became conceptualised as that portion of the operator’s effort which was actually required to complete a task. This view clarified a very important fact regarding any attempt to measure the elusive construct workload. No single measure would ever suffice as the holy grail of workload measurements. The multidimensional nature of workload demanded that multiple measures be used to deal with the construct adequately. Thus while a single metric may assess a specific causal factor, it would be a mistake to demand, as a criterion of acceptance, that any such measure be used to generalise over all kinds of workload or even different levels of the same kind of workload (Wilson and O'Donnell, 1988).

During this period groundbreaking research was performed, which resulted in the development of some of the most widely used workload measurement tools, such as the Subjective Workload Assessment Technique (SWAT) and the National Aeronautical and Space Administration Task Load Index (NASA-TLX). Today these measurement tools are considered to be best practice or gold standards for subjective mental workload measurement. It is also evident from the literature that most of the development work of this period was funded and undertaken or commissioned by military, aviation and aeronautical enterprises.

As mentioned earlier, Wierwille (1988) suggested that mental workload research made initial strides and then appeared to have become bogged down in details. This could be possible, but it may also be possible that researchers have moved on to new (and greener?) research pastures, such as fatigue management, stress and situational awareness. Twenty-eight years after the peak period in mental workload research, there is still no consensual definition of mental workload among researchers. This could have an affect on the interest and motivation of prospective researchers in choosing mental workload as an area of research.

Between 1990 and 2003 very little original work was produced and certainly no new tools for the measurement of mental workload saw the light. Publications from this period focus mainly on the investigation and refinement of physiological measurements of mental workload.

There has however been a resurgence of interest in the relevance of human factors in the railway environment, especially in terms of the mental workload of train drivers and train controllers or dispatchers. Several papers have been published since 2003 and experimental work with new mental workload assessment tools has been conducted. This research is reported on in paragraph 2.9.

A comprehensive listing of papers and presentations on the investigation of human mental workload has been compiled by Hancock, Mihaly, Rahimi and Meshkati (1988).

The literature on workload measurement between 1970 and 1990 reflects a trend towards mental workload quantification, which has resulted in the development of several measures, test instruments and analytical procedures, collectively called *workload estimation techniques* or *workload measurement methodologies*.

Designers and operators realise that performance is not all that matters in the design of a good system. It is just as important to consider what demands a task imposes on operators' limited resources. More specifically, the importance of research on mental workload may be viewed in three different contexts: workload prediction, the assessment of workload imposed by equipment, and the assessment of workload experienced by operators. The difference between the second and the third context exists in their implications for action. When the workload of systems is assessed or compared, the purpose of such a comparison is to optimise the system. When the workload experienced by operators is assessed, it is for the purpose of choosing between operators or providing operators with further training (Wickens and Hollands, 2000).

The assessment of workload, whether psychological or physical, commonly relies on the resource construct, meaning that there is a given (measurable) quantity of capability and attitude (or capacity) available, of which a certain percentage is demanded by the task at hand. If less capacity than is available is required, a reserve exists. Accordingly, workload is often defined as the portion of resource (i.e., of the maximal performance capacity) expended in performing a given task. Following on from this concept, it is obvious that any situation in which more is demanded from operators than can be given should be avoided. This is because the task performance will not be optimal and operators are likely to suffer, physically or psychologically, from the overload (Kroemer, Kroemer and Kroemer-Elbert, 1994).

An increasing proportion of work taxes the information processing capabilities of operators, rather than their physical capacity. This is due to progressively increased automation. It is the load placed upon such cognitive capabilities that mental workload assessment is designed to measure (Meshkati et al., 1995).

The multidimensional nature of workload reflects the interaction of the multiple elements of task demands and operator variables as they relate to workload. When attempting to measure mental workload the challenge is to assess which of the behavioural, physiological or psycho-physiological dimensions will change as a result of mental workload, what form these changes will take, and how to measure them.

The model provided by Meshkati (1988) (Figure 2.5) indicates that the effects of mental workload will be reflected in directly or indirectly observable behaviour, which can be measured by means of various techniques. His model makes provision for three groups of measurement techniques, namely physiological measures, subjective measures and performance measures.

In terms of the operational definition of mental workload, a measurement technique would have to reflect elements of the input load and the capacity of operators. Furthermore, it should enable users of the system to predict the workload at a specific train control centre.

2.7 CRITERIA FOR THE SELECTION OF MENTAL WORKLOAD ASSESSMENT TECHNIQUES

In order to determine the suitability of mental workload measurement techniques for application in research and practical environments, several researchers have developed criteria against which the methods could be measured. The newly developed Mental Workload Index (MWLI) will also be measured against some of these criteria. These criteria will eventually determine the acceptability and utility of the index.

The measures that can be used for the assessment of mental workload have different properties and these properties range from very general to very specific aspects. A general aspect is, for instance, the amount of equipment that is needed. A more specific, and from a scientific perspective, more important aspect is the validity of a measure. Does the measure reflect the concept of mental workload as intended, or does it reflect other concepts, e.g., physical workload?

The following criteria are referred to in the literature (O'Donnell and Eggemeier, 1986; Gawron, Schflett and Miller, 1989):

1. Sensitivity: Is the technique able to detect changes in task difficulty or demands? The index should be sensitive to changes in task difficulty or resource demand.
2. Diagnosticity: Does the technique have the ability to discern the type or cause of workload, or the ability to attribute it to an aspect or aspects of the operators' tasks? An index should not only indicate when workload varies but also the cause of such variation.
3. Selectivity/Validity: The index should be sensitive only to differences in cognitive demands and not to changes in factors such as physical load or emotional stress, which may be unrelated to mental workload or information-processing ability. Types of validity include face, construct, content and predictive (validity).
4. Intrusiveness: The degree to which a technique degrades ordinary or primary task performance is called intrusiveness. The disruption in ongoing task performance as a result of the application of the measurement technique is an undesirable property and should be minimised. It can pose serious problems in the application of a workload measurement technique.
5. Reliability: Reliability refers to the workload estimate that has to be reliable both within and across tests – it must consistently reflect the mental workload. Measures that have been developed in a laboratory setting do not have to indicate workload equally well in the field. Between applications much will depend upon the region of task performance. A measure sensitive only to low levels of workload will not be able to discriminate between levels within high demand situations.
6. Implementation requirements: This criterion refers to the practical constraints associated with the complexity of the measurement procedure and apparatus, such as the requirement of specific equipment or operator training. In field studies in particular, implementation requirements can become important. A workload measurement should be easily transferable from the laboratory environment to the field situation. Factors that contribute to making a technique cumbersome include instrumentation, analyst and operator training, and data recording and analysis.

7. Operator and user acceptance: The degree of approval of the technique by the operators or the perception of the validity and usefulness of the procedure is referred to as operator acceptance. Operators' opinions about a measurement technique, especially the use of self-reports, can affect the correctness and accuracy of the measure. In general, acceptance is higher if the technique is less intrusive or artificial, while the face validity of specific measurements may enhance operator acceptance.
8. Affordability: The affordability of the application, administration and analysis of the technique remains an important criterion in the research process. The balance between affordability and reliable results is an important consideration, but such a balance is not easily obtained.

All the criteria mentioned should be considered and if possible, satisfied in the choice or development of a mental workload measurement technique. For the purposes of this study, the stakeholder group, namely the users of the new measurement technique, operators (TCOs) and employee representative organisations, were consulted on which criteria should be included. The following criteria were considered especially important:

Sensitivity

Selectivity/Validity

Intrusiveness

Implementation requirements

Operator acceptance

Affordability

2.8 DETAILED DISCUSSION OF MENTAL WORKLOAD ASSESSMENT TECHNIQUES

Workload measurement methods can be regarded as a practical intervention that attempts to translate the theoretical concept of *workload* into practical methods for the measurement and evaluation of workload and its effects on system performance. The terms *workload estimation techniques* or *workload measurement methodologies* refer to the numerous measurements, test instruments and analytical procedures that attempt to measure workload.

Mental workload per se cannot be directly observed and therefore its presence and its severity must be inferred from changes in overt behaviour and/or measurable physiological and psycho-physiological functioning (Casali and Wierwille, 1984).

The measurement techniques or methodologies in the literature therefore assume some link between workload and specific behavioural changes in the form of psycho-physiological or physiological change or change in performance. The basic paradigm is therefore that the extent and the nature of these changes occur as a direct consequence of changes in workload.

As far back as 1979, authors have developed different classifications for workload assessment methodologies, some more comprehensive than others (Williges and Wierwille; 1979; Johannsen, 1979; Jex, 1988; Eggemeier, 1988). Essentially these classifications are similar and only semantic differences are evident. For the purposes of this study, the different approaches to the assessment of mental workload are discussed under the following headings:

- Measures of primary task performance
- Secondary measures of spare mental capacity
- Subjective rating techniques
- Physiological indicators of mental workload

Johannsen (1979) added a measurement to his categorisation that was used effectively in the current study, namely Timeline Analysis (TLA). This method is discussed in Chapter 3.

Each of the approaches comprise several techniques, which are discussed under each heading (Kroemer et al., 1994; Meshkati et al., 1995; Wickens and Hollands, 2000).

Some of these assessment techniques meet many of the criteria proposed in the literature, but none satisfy all of the criteria. Some of the criteria may trade off with one another, and until now no technique has been found that satisfies all of the criteria.

2.8.1 Measures of Primary-Task Performance

Performance-based procedures, which include primary- and secondary-task measures, are based on operators' performance levels. Measures of primary-task performance use the adequacy of performance on the task to characterise capacity expenditure. Two types of performance measure have been identified: single measures of the primary task and multiple measures of the primary task. These measures attribute changes in the primary-task performance that can be measured to changes in the workload imposed by the task.

There are several methodological approaches to the measurement of performance, otherwise known as system-output measures (Meshkati et al., 1995).

The analytical approach looks in detail at the actual performance of the task to be assessed, examining not only overall achievement but also the way in which it was attained. From a practical perspective the analytical approach appears most appropriate. The advantage of this method is that the various decisions and other processes that constitute performance are considered in the context in which they normally occur, so that the complexities of any interaction between different elements in the task can be observed.

This approach presents two difficulties. Firstly, the detailed scores required may be difficult to obtain for tasks such as process monitoring, in which most of the decisions made do not result in any overt action. Secondly, even where there is sufficient observable action, recording may have to be elaborate, and analysis of the results can prove laborious.

Synthetic methods provide another approach to performance measurement as a mental workload assessment technique. Specific performance demands that may be placed on operators are identified through task analysis. Performance times and operators' reliability are assigned to the distinct tasks. Information on performance time is then accumulated for a given phase, and the total is compared with the predicted duration of the phase. This comparison of required time with available time can be employed as an index of workload.

Benefits and Shortcomings of Primary-Task Measures

Primary-task measures are probably the most obvious method of mental workload assessment, and one would assume that this method would satisfy a number of the criteria for the selection of techniques, such as ease of implementation and user acceptance. However, the shortcomings far outweigh the benefits of this approach.

There are four reasons why the primary-task performance approach may prove inadequate in clearly revealing the measures of the primary task:

- Firstly, two primary tasks may lie in the *underload* region of the supply-demand space (see Figure 2.1). Since both have sufficient reserve to reach perfect performance, this measure cannot discriminate between them.
- Secondly, there may be differences in how two primary tasks that are to be compared are measured, and what those measures mean.
- Thirdly, sometimes it is simply impossible to obtain good measures of primary-task performance (as with measures of decision-making or vigilance tasks).
- Finally, two primary tasks may differ in their performance, not by the resources demanded to achieve that performance, but by differences in data limits.
  (Wickens and Hollands, 2000).

The lack of sensitivity of performance measures to changes in mental workload levels is one of the major problems of these methods. The level of mental workload may increase while performance is unchanged, so that performance may not be a valid measure of workload. The generalised application of this method to different task situations poses another set of problems, since for each experimental situation, a unique measure must be developed. This problem is not shared by the other methods of mental workload assessment (Meshkati et al., 1995).

Numerous studies support the same conclusion, namely that no substantial change occurs in the primary task as a function of workload (Williges and Wierwille, 1979).

Further, unless one is willing to assume that humans always operate to capacity and that all humans have the same capacity, the performance-based workload scales would reflect only the states of the particular individuals from whom the data were collected. In other words, inter-individual comparisons may not be valid for this measure (Meshkati et al., 1995).

Other problems with primary-task measures are intrusiveness and the interpretation of results. Primary-task data may reflect a wide variety of influences, such as motivation and learning effects (Meshkati et al., 1995).

Williges and Wierwille (1979) concluded that only high workload situations (those that approach operator overload) are discernable by primary-task performance measures, while low workload conditions may not be because at these low levels operators ordinarily adapt in an effort to maintain output variables at an acceptable level.

For these reasons, system designers have often turned to the three other workload assessment techniques, which may assess more directly either the effort invested in primary-task performance or the level of residual capacity available during that performance.

2.8.2 Secondary Measures of Spare Mental Capacity

The largest body of research that deals with operator mental workload is focused on the evaluation of the concept of spare (residual or reserve) mental capacity. This concept is founded on the assumption of a limited-channel model of the human operator (Williges and Wierwille, 1979). Secondary-task measures are typically derived from the levels of performance on a concurrent or secondary task (Eggemeier, 1988).

According to the above-mentioned authors, spare mental capacity is the difference between the total workload capacity of operators and the capacity needed to perform the task. As spare mental capacity decreases, operators' workload increases until the point of overload is reached. At this point, the information processing demands of the task exceed the workload capacity of the operators. This approach assumes that an upper limit exists on the ability of human operators to gather and process information.

A study of the research literature rendered three approaches in the measurement of spare mental capacity: task analytic methods, secondary-task measures, and occlusion procedures.

(a) Task Analytic Methods

Task analytic methods are based on mathematical methods derived from systems engineering. This approach assumes that all task components, when performed serially, require specific lengths of time to complete, and as long as the time available for overall completion exceeds the sum of theoretical time durations for performing the task components, spare mental capacity exists. When, however, the actual time available is insufficient, stress and task loading occur (Wierwille and Gutmann, 1978).

The methods used under the Task Analytic approach are:

- *Task Component/Task Summation*: System engineering principles and task analysis procedures have been widely used to develop computer-based models of aircrew and cockpit simulation techniques with which to obtain descriptions of aircrew performance. An essential input is the detailed task analysis that forms the basis of time assessments for each task component.
- *Information Theoretic*: This approach attempts to quantify workload in terms of an information-transmission metric that is specified in bits/second. This approach has been applied in the evaluation of visual monitoring but has not found wide application and support (Wierwille and Gutmann, 1978).

(b) Secondary-Task Methods

Secondary tasks have been used in most behavioural research approaches that seek to estimate spare mental capacity. A secondary task is a task that operators are asked to perform in addition to their primary task. The task is performed only once the primary task has been fully attended to. This method is therefore an indirect measure of operator workload. If operators are able to perform well on the secondary task, this indicates that the primary task is relatively easy; if they are unable to perform the secondary task and simultaneously maintain the primary task performance, this indicates that the primary task is more demanding. The difference between the performances obtained under the two conditions is then taken as a measure or index of the workload imposed by the primary task (Meshkati et al., 1995).

Measuring performance on a secondary concurrent task is intended to assess the spare capacity that remains after capacity resources have been allocated to the primary task. If subjects allocate some of the resources actually required for primary-task performance to the secondary task, the secondary task is intrusive (or invasive) on the primary task. An intrusive secondary task modifies the condition that is to be assessed.

A multitude of secondary tasks have been proposed and employed over the last two decades to assess the residual capacity available after the completion of primary tasks. Some of the most prominent examples are described here (Wickens and Hollands, 2000):

- Rhythmic tapping task: Operators must produce finger or foot taps at a constant rate. Tapping variability increases as primary workload increases.
- Random number generation: This requires that operators generate a series of random numbers. Normally the degree of randomness declines as workload increases, and more repetitive sequences are generated.
- Probe reaction time: These tasks are often used as a workload measurement technique, as it is assumed that greater primary-task workload will prolong the reaction time to a secondary-task stimulus.
- Other techniques: These include arithmetic addition, and critical tracking tasks.
- Handwriting analysis: Handwriting deteriorates as a result of distraction that can result from having to perform other tasks. This deterioration is a potentially effective measure of mental overload.

(c) Occlusion procedures

Occlusion is a time-sharing task that is forced rather than voluntary. Operators are given samples of the visual information that is required to perform the primary task and then information inputs, i.e. from the visual display, are suppressed or blocked. This can be achieved by operators having to wear a helmet fitted with an opaque visor that can be closed by external control. Alternatively, visual input is blocked by directly blanking the electronic displays of the system. This method is however considered intrusive and potentially unsafe (Meshkati et al., 1995).

(d) Benefits and Shortcomings of Secondary-Task Measures

There are two distinct benefits of secondary-task techniques. Firstly, they have a high degree of face validity. They are designed to predict the amount of residual attention operators will have available if an unexpected failure or environmental event occurs. Secondly, the same secondary task can be applied to two very different primary tasks and it will give workload measures in the same units (Wickens and Hollands, 2000).

However, the secondary task as a mental measurement technique has many shortcomings. Perhaps the most difficult aspect of secondary-task methodology for assessing workload is intrusiveness. When the secondary task is introduced, it normally interferes with the primary task and performance on the primary task is known to be modified and usually degraded. This technique therefore suffers on the intrusiveness criterion (Williges and Wierwille, 1979; Wierwille, Rahimi, and Casali, 1985).

One the one hand, this may be inconvenient or even dangerous if the primary task is one like flying, driving or controlling the movement of trains, and a diversion of resources to the secondary task at the wrong time could lead to an accident. On the other hand, disruption of the primary task could present problems of interpretation if the amount of disruption suffered by two primary tasks to be compared is not the same, that is, the measurement technique differentially disrupts that which is being measured (Wickens and Hollands, 2000).

2.8.3 Subjective Rating Techniques

Subjective assessment or self-reporting is probably the most common method used to evaluate mental workload. This method provides a valid assessment of the overall workload inflicted on the working memory of operators. The ease and speed of application and interpretation, as well as the high face validity of these measurements make them popular to use (Vidulich, 1988).

Although various individual assessment techniques have been developed, all subjective procedures use some report of experienced effort or capacity expenditure to characterise workload levels. Subjective methods include direct or indirect questioning of individuals for their opinions of the workload involved in tasks (Eggemeier, 1988; Meshkati et al., 1995).

Subjective measures represent the operators' conscious judgements regarding the difficulties encountered in the performance of the evaluated task. If a task imposes a high workload on operators, then it is expected that operators will *feel* loaded and will be able to report it. Care should however be taken with this approach. It is essential to understand which factors determine the relationship between subjective workload and objective performance. A subjective workload assessment should be viewed essentially as a verbal report on the level of information load experienced when performing a task (Vidulich, 1988).

Subjective assessments of the perceived workload rely on internal integration of the demands, but they may be unreliable, invalid, or inconsistent with other performance measures. If subjective measures are taken after the task has been completed, they are not real-time evaluations; if performed during the task, they may intrude (Nygren, 1991; Wickens and Yeh, 1983).

Some proponents of self-report measures consider them to be the best measures since they come nearest to tapping the essence of mental workload. Critics, on the other hand, say that the source of resource demands is hard to introspectively diagnose within a dimensional framework. Physical and mental workload is, according to the critics, hard to separate (O'Donnell and Eggemeier, 1986). For reasons that are discussed later in this chapter, self-report measures were not considered viable for the Spoornet study.

Pulat (1992) expresses his view of subjective ratings as follows: "The easiest but probably the most unreliable method is simply to ask operators their opinion on their workload on completion of a task or shift." He provides an example of a list of key words or definitions describing the different levels of workload that are typically used in rating scales:

Task difficulty	Low .... High
Time pressure	None .... Rushed
Effort	None .... Impossible
Frustration	Fulfilled .... Exasperated
Stress level	Relaxed .... Tense
Fatigue	Alert .... Exhausted

Another approach to the subjective rating of workload is to use interviews and questionnaires. The procedures used in this approach are not as structured as rating scales. They range from completely open-ended debriefing sessions to self-reporting logs of stressful activities (Meshkati et al., 1995).

Subjective measures of workload, such as the modified Cooper-Harper Scale (Wierwille et al., 1985), the NASA Task Load Index (TLX) (Hart and Staveland, 1988) and the Subjective Workload Assessment Technique (SWAT) (Reid and Nygren, 1988) have been widely used. According to a personal communication from a human factors expert in the transport industry in the USA, the SWAT and NASA-TLX are considered to be industry standards and therefore a good benchmark when developing a measurement tool (Reinach, 2006a). These instruments have, however, been criticised on both theoretical and technical grounds.

An individual's perceived difficulty (of a task) is influenced by long-term memory, which includes both memory of general experience and memory of similar tasks. Background factors, such as personality traits, habits and general attitudes as well as transient conditions (e.g. emotional state, fatigue, motivation and anticipated success or failure) play a role in operators' subjective ratings of their experience of the task (Meshkati et al., 1995).

Care must be taken when interpreting these scales since they contain words that may be interpreted differently under different task situations.

The rating scales and questionnaires identified in the research literature are discussed in more detail below.

(a) Variations of the Cooper Scale

Since the 1960s the mental load imposed by manual control tasks, specifically aircraft handling qualities, has been measured by the Cooper (C) Scale and modifications thereof, namely the Cooper-Harper (C-H) and Modified Cooper Harper (MC-H) scales (Moray, 1982).

The Cooper Scale (Moray, 1982) offers ten statements to the evaluator, who has to indicate which statement best approximates his opinion of the handling qualities of the aircraft or aspect under consideration. Various authors have commented on the deficiencies of this instrument. Criticism of the instrument relates to: (i) wording ambiguity and confusing nomenclature; (ii) the dual mission character of the scales and mixing of the tasks, e.g., normal and emergency conditions; and (iii) the lack of information about the quantitative character of the psychological scale continuum (Cilliers, 1992).

Table 2.1: The Original Cooper Rating Scale (reproduced in Cilliers, 1992.)

	Adjective Rating	Numerical Rating	Description	Primary mission accomplished	Can be landed
NORMAL OPERATION	Satisfactory	1	• Excellent, includes optimum	Yes	Yes
		2	• Good, pleasant to fly	Yes	Yes
		3	• Satisfactory, but with some mildly unpleasant characteristics	Yes	Yes
EMERGENCY OPERATION	Unsatisfactory	4	• Acceptable, but with unpleasant characteristics	Yes	Yes
		5	• Unacceptable for normal operation	Doubtful	Yes
		6	• Acceptable for emergency condition only	Doubtful	Yes
NO OPERATION	Unacceptable	7	• Unacceptable, even for emergency condition	No	Doubtful
		8	• Unacceptable – Dangerous	No	No
		9	• Unacceptable - Uncontrollable	No	No
	Unprintable	10	• "Motions possibly violent enough to prevent pilot escape"		

The Cooper-Harper (C-H) Scale is the most widely used rating scale for assessing the handling qualities of an aircraft. The descriptors of the scale pertain to the *flyability* of an aircraft, and although the scale contains some reference to workload, the descriptors would have to be modified for use in other workload applications (Cilliers, 1992).

The C-H Scale is one of the more validated scales for the subjective measurement of workload on aircraft handling qualities. It consists of a 10-point scale with a decision-tree format. It makes provision for performing the rating task sequentially, reaching a final rating in a deliberate and careful manner (Cilliers, 1992).

The main disadvantage of decision-tree rating scales is that they cannot provide information of an equal interval nature. At best they provide ordinal estimates of workload (Jex, 1988).

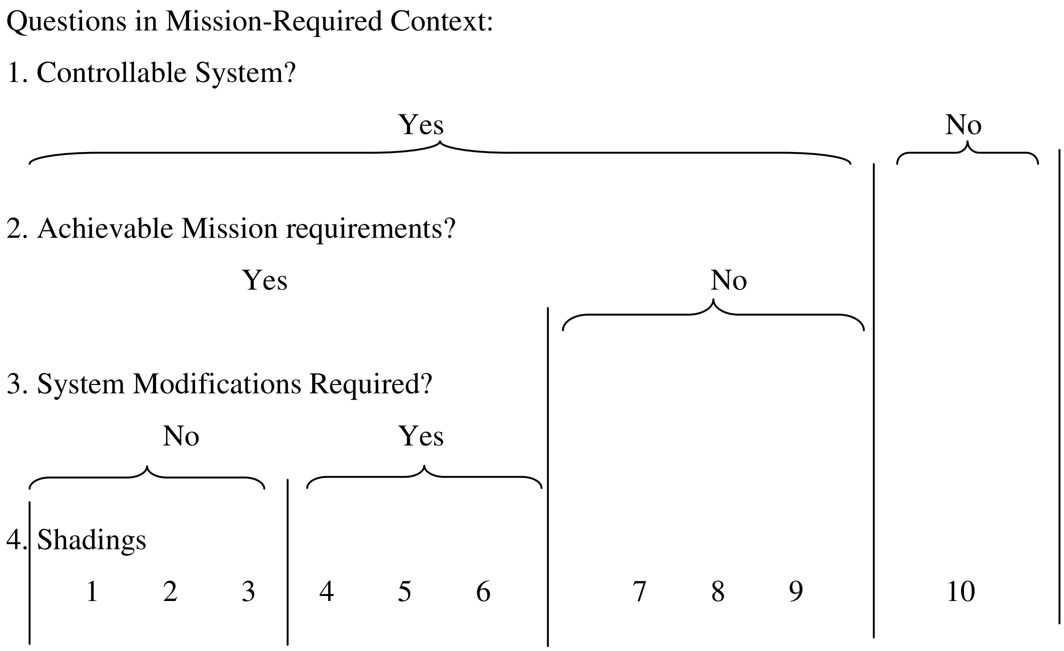


Figure 2.6: Cooper-Harper Rating Scale (reproduced in Cilliers, 1992)

The Modified Cooper-Harper (MC-H) Scale is considered a further development in the subjective measurement of mental workload. This scale is applicable to a wider variety of task workloads, especially for systems that may load perceptual, mediational and communication activities. The scale can be applied to obtain mental workload estimations regardless of the type of loading imposed by the task. The wording of the scale has been modified to enable the assessment of mental workload as distinct from the psychomotor workload, which the original scale was designed to measure (Rahimi and Wierwille, 1982; Wierwille and Casali, 1983a, Wierwille et al., 1985).

Figure 2.7 provides an excerpt from the Modified Cooper-Harper Scale.

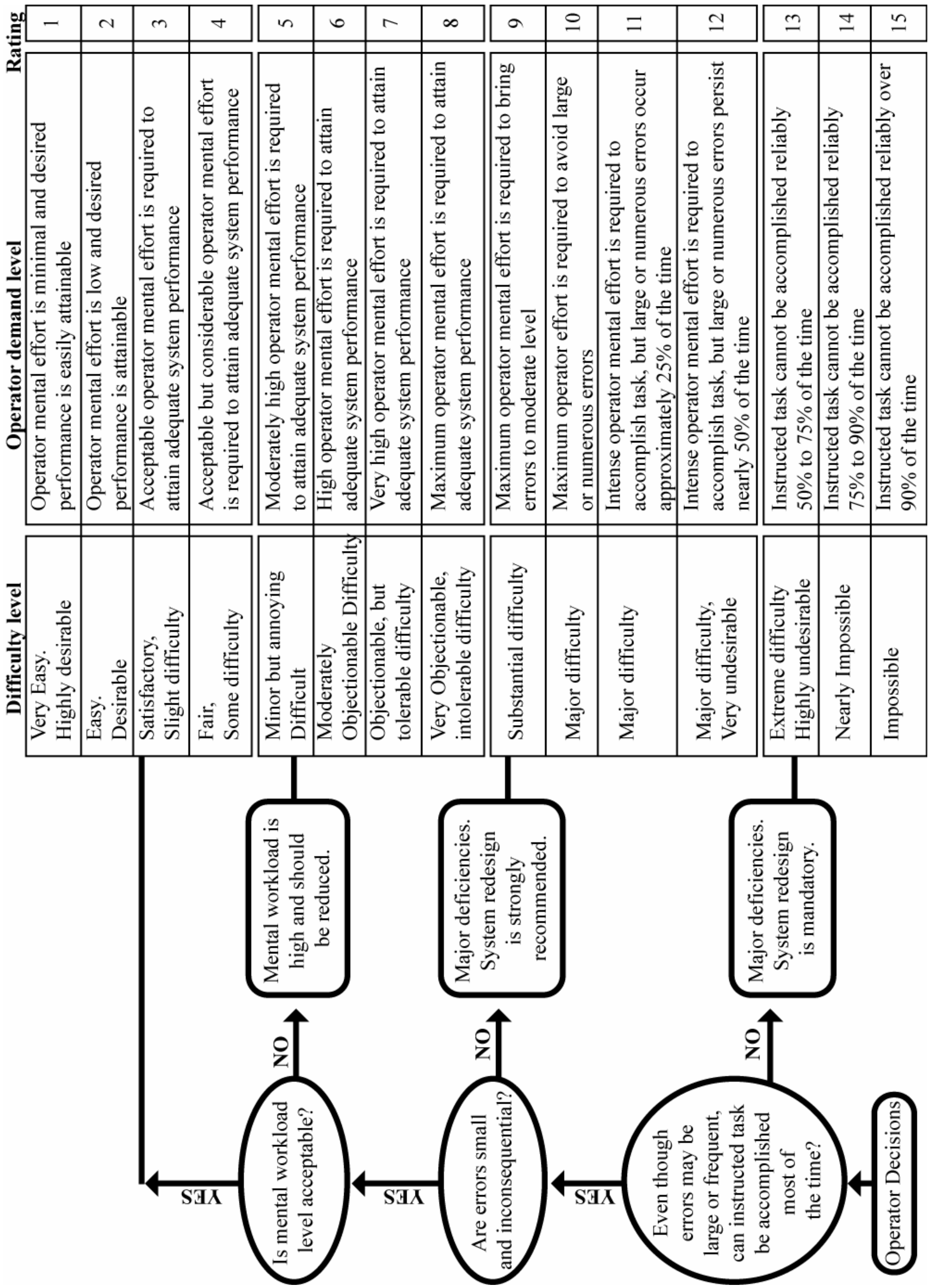


Figure 2.7: Modified Cooper-Harper Scale and rating scales (reproduced in Cilliers, 1992)

(b) The Bedford Scale

The Bedford Scale is described as similar to the Modified Cooper-Harper Scale, in which a single score is generated on a 10-point scale. A workload of 10 is considered to be extremely high with no spare capacity and a level of 1 is considered to be insignificant.

The underlying assumption of the Bedford Scale is that workload is best defined in terms of the subjective experience of *effort*. This definition takes into account moderating factors such as ability, experience and the individual's response to stress.

Critics of the Bedford Scale are of the opinion that the scale is not linear and lacks sensitivity at the lower end. The scale has been used in field studies but has not undergone laboratory validation (Cilliers, 1992).

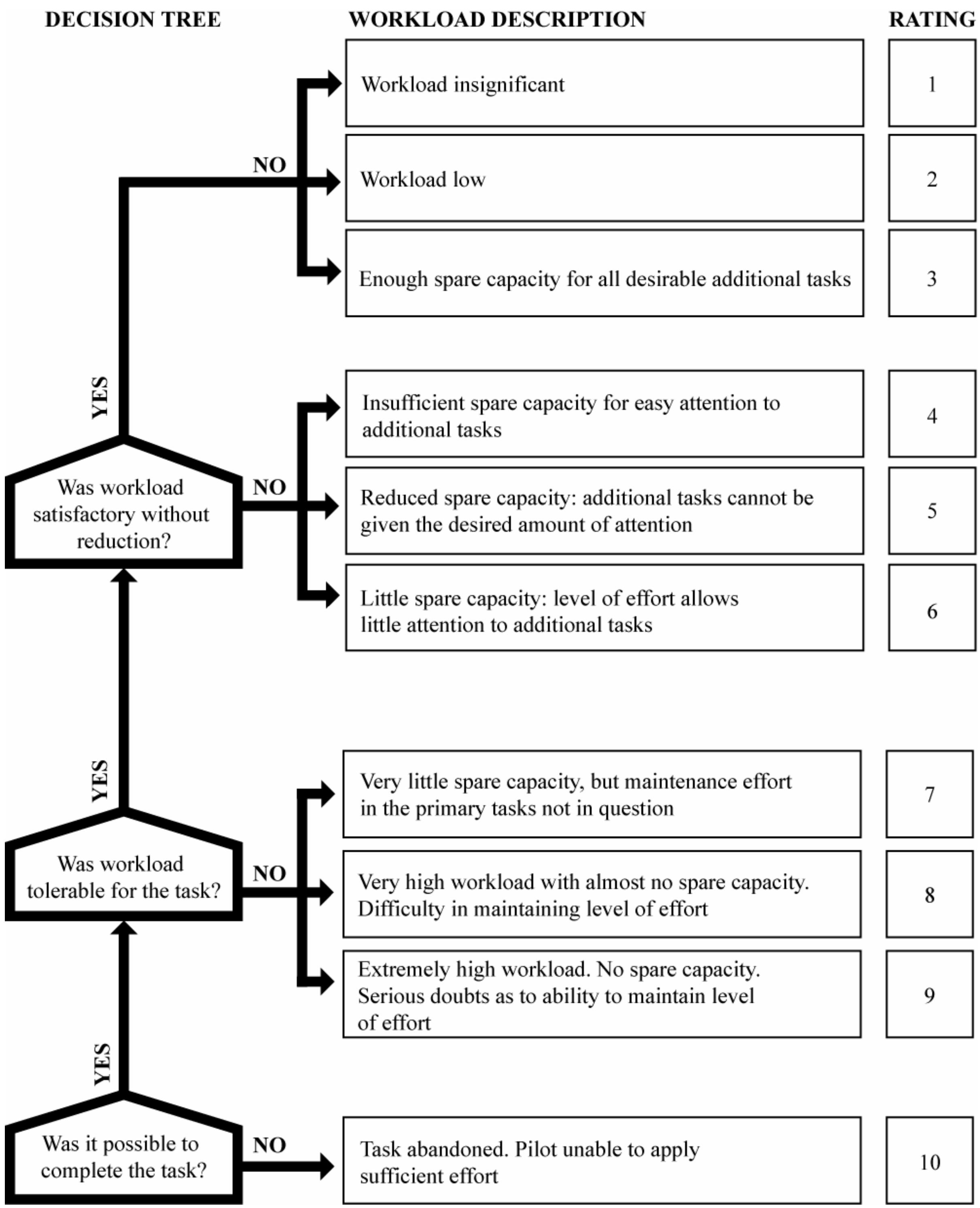


Figure 2.8: The Bedford Scale (reproduced in Cilliers, 1992)

(c) The Subjective Workload Assessment Technique (SWAT)

Another widely used subjective rating technique is the Subjective Workload Assessment Technique, which is a multidimensional rating scale, specifically designed to measure operator workload in a variety of systems, and for a number of tasks (Reid and Nygren, 1988, Meshkati et al., 1995).

The SWAT has been developed as a generalised procedure for scaling pilot mental workload. The procedure entails a two-step process wherein each subject completes a scale development phase and an event-scoring phase. During the scale development phase, data necessary to develop a workload scale is obtained from a group of subjects. During the event-scoring phase, subjects rate the workload associated with a particular task and/or mission segment. These two steps are considered to be distinct events which occur at different times (Reid and Nygren, 1988).

The principles that guided the development of SWAT were the following:

- To develop as precise a measure as possible while minimising the intrusiveness of the data collection procedure in the operational situation;
- To place minimal measurement constraints on the complexity of the judgement task that is required of the operators making workload evaluations;
- To provide a mechanism for testing the validity of the formal measurement model that is assumed by the underlying additive model in SWAT (Reid and Nygren, 1988).

The SWAT measures workload on three, three-point scales (See Table 2.2).

Table 2.2: The SWAT Scale (From: Wickens and Hollands, 2000)

Time Load	Mental Effort Load	Stress Load
1. Often have spare time. Interruptions or overlaps among activities occur infrequently or not at all.	1. Very little conscious mental effort or concentration required. Activity is almost automatic, requiring little or no attention.	1. Little confusion, risk, frustration, or anxiety exists and can be easily accommodated.
2. Occasionally have spare time. Interruptions or overlaps among activities occur frequently.	2. Moderate conscious mental effort or concentration required. Complexity of activity is moderately high due to uncertainty, unpredictability or unfamiliarity. Considerable attention required.	2. Moderate stress due to confusion, frustration, or anxiety noticeably adds to the workload. Significant compensation is required to maintain adequate performance.
3. Almost never have spare time. Interruptions or overlaps among activities are very frequent or occur all the time.	3. Extensive mental effort and concentration necessary. Very complex activity requiring total attention.	3. High to very intense stress due to confusion, frustration or anxiety. High to extreme determination and self-control required.

Although SWAT is one of the most sophisticated workload assessment techniques available, it is by no means free of problems. The following are unresolved practical and theoretical issues (Hart, 1986):

- The between-subject variance is high and standard deviations can be as high as 70-80% when compared with average SWAT scores.
- The scale development phase is time consuming.
- The assumption that people can accurately predict the workload of 27 combinations of abstract variables is probably optimistic and not justifiable.
- High inter-rater reliability may only reflect agreement on extremes of workload.

- Three factors alone are probably not sufficient to characterise workload experiences and definitions.

SWAT demonstrated sensitivity for workload assessment related to visual display monitoring (the scales were designed to heavily load perceptual input capacity); verbal and spatial short-term memory (the scales primarily loaded two major central processing coding dimensions); and unstable tracking (the scales exerted heavy demands on motor output capacity). SWAT scales (like the M C-H and NASA-TLX Scales) are capable of reflecting variations in effort expenditure across a variety of processing functions, but should be considered as global measures of workload rather than diagnostic in their sensitivity (Eggemeier, 1988).

(d) The NASA Task Load Index (TLX)

The work of Hart and Staveland culminated in the development of the NASA Task Load Index (TLX), an extensively researched and widely used subjective mental workload measurement tool that is considered one of the gold standards in the subjective assessment of mental workload.

The NASA-TLX was developed as part of a research program that was conducted over several years to identify the factors associated with variations in subjective workload within and between different types of tasks. The tasks included simple cognitive and manual control tasks, complex laboratory and supervisory control tasks and aircraft simulation (Hart and Staveland, 1988).

A multidimensional rating scale was proposed by Hart and Staveland, in which information about the magnitude and source of six workload-related factors were combined to derive a sensitive and reliable estimate of workload.

Table 2.3: The NASA Bipolar Rating Scale descriptions (Hart and Staveland, 1988)

RATING SCALE DESCRIPTIONS		
Title	Endpoints	Descriptions
OVERALL WORKLOAD	*Low, High*	The total workload associated with the task, considering all sources and components.
TASK DIFFICULTY	*Low, High*	Whether the task was easy or demanding, simple or complex, exacting or forgiving.
TIME PRESSURE	*None, Rushed*	The amount of pressure you felt due to the rate at which the task elements occurred. Was the task slow and leisurely or rapid and frantic?
PERFORMANCE	*Failure, Perfect*	How successful you think you were in doing what we asked you to do and how satisfied you were with what you accomplished.
MENTAL/ SENSORY EFFORT	*None, Impossible*	The amount of mental and/or perceptual activity that was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)?
PHYSICAL EFFORT	*None, Impossible*	The amount of physical activity that was required (e.g., pushing, pulling, turning, controlling, activating, etc.).
FRUSTRATION LEVEL	*Fulfilled, Exasperated*	How insecure, discouraged, irritated, and annoyed versus secure, gratified, content, and complacent you felt.
STRESS LEVEL	*Relaxed, Tense*	How anxious, worried, uptight, and harassed or calm, tranquil, placid and relaxed you felt.
FATIGUE	*Exhausted, Alert*	How tired, weary, worn out, and exhausted or fresh, vigorous, and energetic you felt.
ACTIVITY TYPE	*Skill-based, Rule-based, Knowledge-based*	The degree to which the task required mindless reaction to well-learned routines or required the application of known rules or required problem solving and decision making.

Each scale had bipolar descriptors at each end (high/low). Numerical values were not displayed, but values ranging from 1 to 100 were assigned to scale positions during data analysis. This set of scales was used to evaluate the experiences of subjects in 25 different studies.

One of the goals in the gathering of workload and workload ratings was to amass a database that would allow for the examination of the relationships among different task, behaviour and psychological factors in order to create a valid and sensitive rating technique for subjective workload assessment. The assumption was made that the technique would be multidimensional, but that the number of subscales should be less than the number used for research purposes.

In the comprehensive experimentation that followed, data was collected from various types of tasks: single cognitive tasks, single axis manual control tasks, dual task experiments and various simulated tasks. A comprehensive amount of data was collected and analysed to construct a workload rating scale.

From their experimentation Hart and Staveland (1988) reported a number of key points that emerged from the subjective experiences and evaluations of workload by the subjects:

- A phenomenon exists that can generally be termed *workload*, but its specific causes may differ from one task to the next.
- Ratings of component factors are more diagnostic than overall workload ratings.
- Subjects' workload definitions are different, which contributes to between-subject variability. However, the specific sources of loading imposed by a task are more potent determinants of workload experiences than such *a priori* biases.
- A weighted combination of the magnitudes of factors that contribute to subjects' workload experience during different tasks provides an integrated measure of overall workload that is relatively stable between raters.

The research resulted in the development of the NASA-TLX rating scale of which the following sub-scales form a part and which and are rated by the subjects:

Table 2.4: The NASA-TLX Rating Scale Definitions (Hart and Staveland, 1988).

NASA-TLX RATING SCALE DEFINITIONS		
Title	Endpoints	Descriptions
MENTAL DEMAND	*Low, High*	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
PHYSICAL DEMAND	*Low, High*	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
TEMPORAL DEMAND	*Low, High*	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
PERFORMANCE	*Good, Poor*	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
EFFORT	*Low, High*	How hard did you have to work (mentally and physically) to accomplish your level of performance?
FRUSTRATION LEVEL	*Low, High*	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

This scale has been subjected to extensive studies of reliability and validity (Hart and Staveland, 1988). However, none of these studies involved the physiological measurement of the stress parameters associated with mental workload.

(e) The Workload Profile (WP)

More recently the Workload Profile (WP), a multidimensional instrument for the assessment of subjective mental workload, was developed and introduced. This instrument is based on the multiple resource model of Wickens (1992). The WP attempts to combine the advantages of secondary-task performance-based procedures (high diagnostic ability) and subjective techniques (high subject acceptability and low implementation requirements and intrusiveness). The WP asks subjects to provide the proportion of attention resources that they had used after they had experienced all of the tasks.

In a recent study the characteristics of the WP were compared with NASA-TLX and SWAT. The WP bore the highest sensitivity (the tool's ability to detect changes in task difficulty demands). It also showed high diagnostic ability and low intrusiveness. Problems with implementation and acceptability were however experienced (Rubio, Diaz, Martin and Puente, 2004).

(f) Benefits and Shortcomings of Subjective Techniques

The benefits of subjective techniques are apparent. They do not disrupt primary-task performance, and they are relatively easy to apply and interpret. Their shortcomings relate to the uncertainty with which operators' verbal statements diagnostically reflect the investment of or demand for processing resources and are not influenced by other biases (e.g., dislike of or unfamiliarity with the task, or the rater's reluctance to report that tasks are difficult).

Among the shortcomings of subjective ratings is the problem posed by a task being primarily a function of the raters' perceptions. The perceived difficulty of a task might alter operators' attitudes to it. Subjective ratings of task difficulty could also be affected by the situation and the job as a whole rather than by task-induced or individual rater factors alone (Meshkati et al., 1995).

Furthermore, these methods are insensitive to demands outside the component of the human information-processing system that deals with working memory. Also, performing multiple tasks concurrently (as is the case with TCOs in this project) seems to render subjective assessments somewhat insensitive to changes in one or more of the tasks.

Air traffic controller (ATC) workload evaluations have tended to rely mainly on subjective measures. Although the use of subjective measures is attractive (they are inexpensive and easily collected), researchers should remain mindful of their limitations. This admonition seems especially relevant to the ATC domain, in which highly skilled operators are asked to give subjective reports. Such operators may be unable – or for reasons of job security, unwilling – to give accurate reports. Moreover, such reports may be subject to individual biases, preconceptions, or memory limitations (Hilburn and Jorna, 2001).

2.8.4 Physiological Indicators of Mental Workload

One solution to performance intrusiveness is to record, unobtrusively, the manifestations of workload through appropriately chosen physiological measures of autonomic or central nervous system activity. These techniques offer two distinct advantages over secondary-task and subjective measures: participants are not required to execute any extra overt behaviour, and psycho-physiological measures are capable of measuring fluctuations in mental workload over time.

The advantage of physiological responses measurements is that they do not require overt responses from operators, and most cognitive tasks do not require overt behaviour. Moreover, most of the measures can be collected continuously, and measurements have now become relatively unobtrusive due to the miniaturisation of instruments (De Waard, 1996).

Human mental workload requires the expenditure of physiological effort and resources. The group of workload techniques termed *physiological* can generally be discerned from other workload estimation techniques by the fact that the changes that cause measurable variations are largely involuntary. The autonomic (involuntary) nervous system is of particular interest in the context of workload measurement as it controls the heart, secretion glands and the involuntary muscles (Wilson and O'Donnell, 1988).

Physiological measures reflect processes such as the demand for increased energy, progressive degradation of the system or homeostatic functioning of mechanisms designed to restore system equilibrium that have become disturbed by cognitive requirements. It is, however, difficult to distinguish between activation that is specific to the perceptions of individual operators, and the activation that results from the actual workload imposed (Hancock, Meshkati, Robertson and Robertson, 1985). In any situation in which individuals are required to increase mental or physical effort or to deal with stimuli that evoke emotion, involuntary changes in the autonomic nervous system (ANS) are produced. Although it would be reasonable to assume that these changes would be reflected in physiological measurements, attempts at finding such a specific mental workload index have been unsuccessful.

2.8.4.1 Mental Workload and Stress Measurement

To understand how mental overload manifests itself physiologically and why physiological measurements are considered useful in measuring workload, a discussion on the functional relationship between workload and stress is necessary.

Physiological measures are one of the most widely used research methods for assessing operator mental workload. The physiological method generally involves measuring and processing data for one or more variables related to human physiological processes. The underlying concept in physiological monitoring is as follows: As operator workload increases or decreases, involuntary changes take place in the physiological processes of the human body (body chemistry, nervous system activity, circulatory or respiratory activity). Consequently, workload may be assessed by measuring and processing the appropriate physiological variables.

Among many researchers there is an underlying assumption that high workload levels are accompanied by increased emotional stress. This stress is then measured by physiological recordings and is related back to workload. Stress in this case is assumed to act as an intermediate variable, causing physiological changes. Among other researchers the underlying assumption involves changes in the state of arousal. *Arousal* may be considered to be a state of preparedness or level of activation of the human organism (Wierwille, 1979).

Through changes in the status of a number of physiological systems, individuals engaged in cognitive activities provide indirect indices of their level of effort. Ursin and Ursin (1979) recognised that these physiological methods do not measure the imposed load but rather give information as to how individuals respond to the load and, in particular, whether they are able to cope with it.

Successful physiological measurement of operator workload has, unfortunately, been much easier to conceptualise than to achieve. Workload would intuitively seem to require the expenditure of physiological effort and resources, and it is reasonable to assume that some central or peripheral measure could be found that would provide an index of this expenditure. Early attempts to find such an index were, however, remarkably unsuccessful. The frequent failure of specific physiological measures to correlate with imposed workload has been noted (O'Donnell and Eggemeier, 1986). Indeed such failures led to an early belief that physiological measures might have questionable value as valid, reliable measures of mental workload. However, theoretical and laboratory work continued to include and refine a variety of such techniques, such as heart rate, eye blink and EEG measures. These efforts met with mixed success, and while the inconclusive results tended to stimulate interest in physiological measures, the preponderance of data continued to be negative, ambivalent or contradictory (Wilson and O'Donnell, 1988).

It should be recognised that many aspects of operator behaviour other than mental workload may have an effect on physiological measures. A large number of researchers who did not obtain significant results in applying mental workload measurement techniques suggested that either the sample population must be homogenised, or that personality traits, individual differences and other related factors such as past experience, skill, and motivation should be incorporated in the model, since many of these factors have been shown to influence physiological measurements (Meshkati and Loewenthal, 1988).

Early indications of specificity for at least some physiological measures came from studies such as those by Lacey and Lacey (1958). These investigators demonstrated the remarkable ability of cardiac measures to differentiate between various types of task-related ECG deceleration patterns. Various aspects of the cardiac cycle were shown to be dependent on both the task activity and the mental and/or physical involvement of the individual.

It is recognised that the functional relationship between mental workload and stress is not a very firm one, due to the complexity of the concepts of mental workload and stress and the variety of other factors that influence the measurement of these concepts. Also, one single measure has proven insufficient to measure these concepts. Despite these difficulties, it is the opinion expressed in this paper that physiological measurements still provide a valuable objective measure of mental workload. The next section explains the relationship between mental load and the effect on the human nervous system.

2.8.4.2 The Organisation of the Human Nervous System

In order to understand how mental overload manifests itself physiologically and why physiological measurements are considered useful in the measurement of workload, an understanding of physiological stress manifestation mechanisms is necessary.

From an anatomical perspective, two fundamental nervous systems exist: the central nervous system (CNS) and the peripheral nervous system (PNS):

(a) CNS

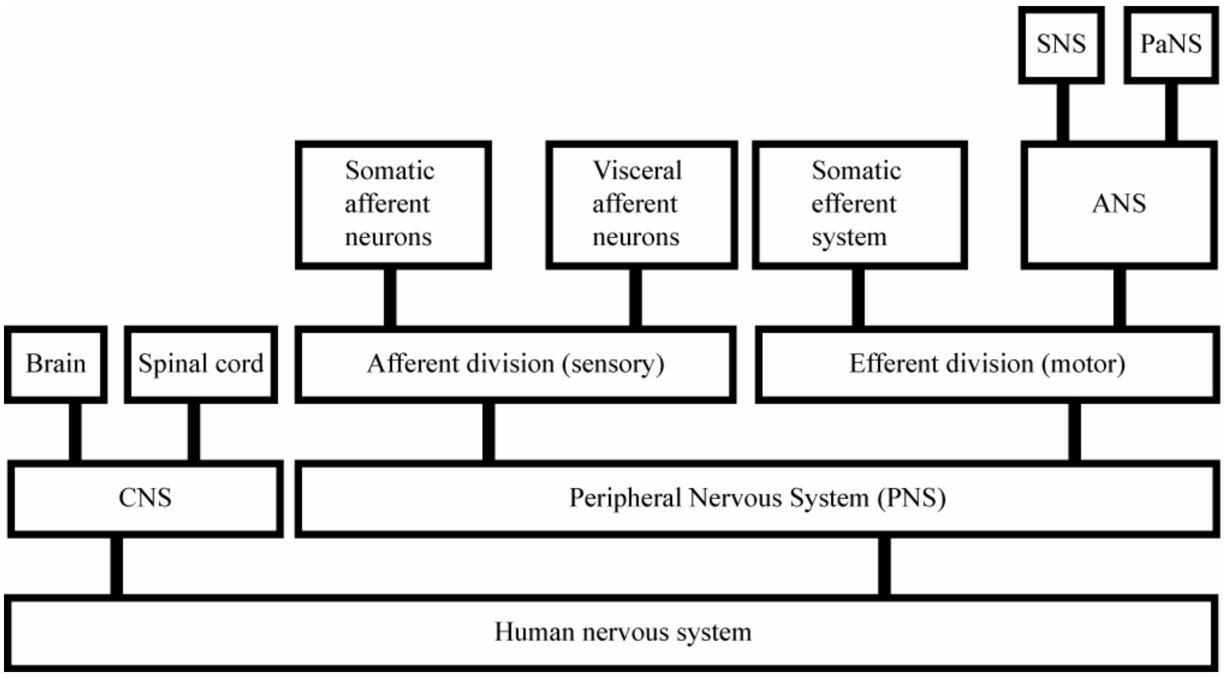
- The CNS consists of the brain and spinal cord. Together they function as the integrated control centre of the complete nervous system. Incoming sensory information is serviced by responses based on experience (memory), reflexes and current conditions (Gericke, 1994).
- The CNS is further divided into three functional levels (Everly and Rosenfeld, 1981):
    - The **neocortex** is the most sophisticated component of the human brain. Among other functions such as the decoding and interpretation of sensory signals, communication, and gross control of motor behaviour (musculoskeletal), the neocortex (primarily the frontal lobe) presides over imagination, logic, decision making, memory, problem solving, planning and apprehension.

- The **limbic system** (or limbic brain) represents the major component of the second level of the brain. The limbic brain is of interest in the discussion of stress because of its role as the emotional (affective) control centre for the human brain. The limbic system consists of numerous neural structures, e.g., the hypothalamus, hippocampus, septum, cingulated gyrus, amygdala and pituitary gland (the master endocrine gland).
- The **reticular formation** and the **brain stem** represent the third and most primitive level of the brain. The major function of this level is the maintenance of vegetative functions (heartbeat, respiration, vasomotor activity) and the conduction of impulses through the reticular formation and relay centres of the thalamus to the higher order levels of the brain.
- The **spinal cord** represents the central pathway for neurons as they conduct signals to and from the brain.

(b) PNS

The PNS consists of all neurons in the body excluding those in the CNS. Anatomically, the PNS may be thought of as an extension of the CNS in that the functional control centres for the PNS lie in the CNS. The PNS is divided into two systems: the somatic and the autonomic.

- The **somatic system** carries sensory and motor signals to and from the CNS. Thus it innervates sensory mechanisms and striate muscles.
- The **autonomic system** carries impulses that are concerned with the regulation of the body's internal environment and the maintenance of homeostasis. The autonomic system, therefore, innervates the heart, the smooth muscles and the glands. The autonomic nervous system can be further subdivided into two branches or subsystems (Everly and Rosenfeld, 1981):
    - The sympathetic nervous system is concerned with preparing the body for action. Its effect on the organs it activates is that of generalised arousal.
    - The parasympathetic nervous system is concerned with restorative functions and the relaxation of the body. Its general effects are those of slowing and maintaining basic bodily functions.



CNS – Central nervous system
SNS – Sympathetic nervous system
ANS – Autonomic nervous system
PaNS – Parasympathetic nervous system

Figure 2:9: Structure of the human nervous system

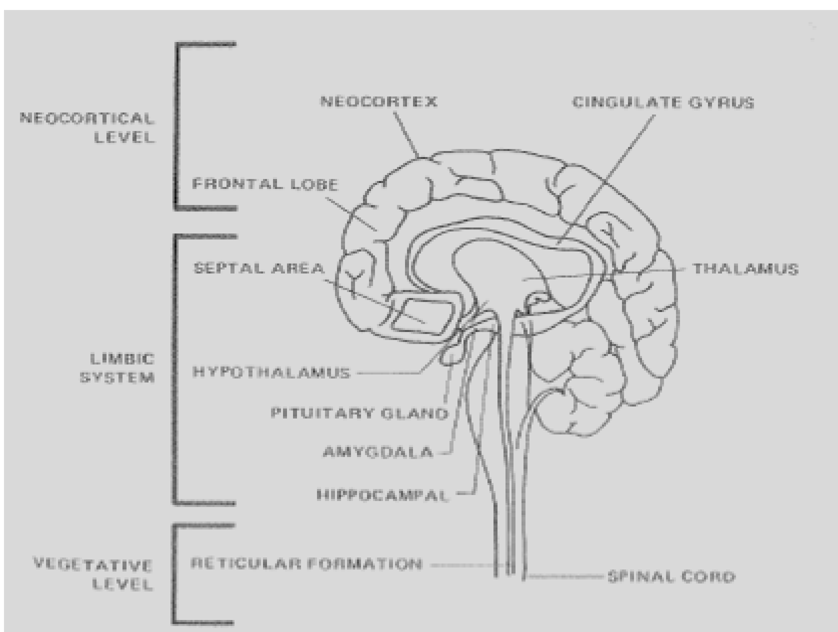


Figure 2.10: The human brain

In this study three functions that are primarily associated with the sympathetic nervous system, i.e. heart rate, heart-rate variability and blood pressure are monitored.

2.8.4.3 The stress response

Everly and Rosenfeld (1981) provide the following description to illustrate the stress response as a reaction to a psychosocial stimulus:

(a) The Initiation of the Stress Response – CNS Mechanisms

If any given, otherwise neutral, stimulus is to evoke a stress response, it must first be received by the sensory receptors of the PNS. Once stimulated, the sensory receptors send impulses along the sensory pathways of the PNS towards the brain. Once in the CNS, collateral nerves from the sensory pathways diverge from the main ascending pathways to the neocortex and innervate the reticular formation. The emotionally-coloured neocortical interpretation is fed back to the limbic system. If the neocortical-limbic system perceives a psychosocial stimulus to be a threat or a challenge, or to be in any way aversive, then emotional arousal will likely result. In most individuals the activation of emotional mechanisms stimulates one or more of the three major psychosomatic stress axes (see Figure 3.6 below). Therefore, we see that stress reactions to psychosocial stimuli result from the cognitive interpretation of the stimuli and the resultant emotional arousal, rather than from the stimuli themselves.

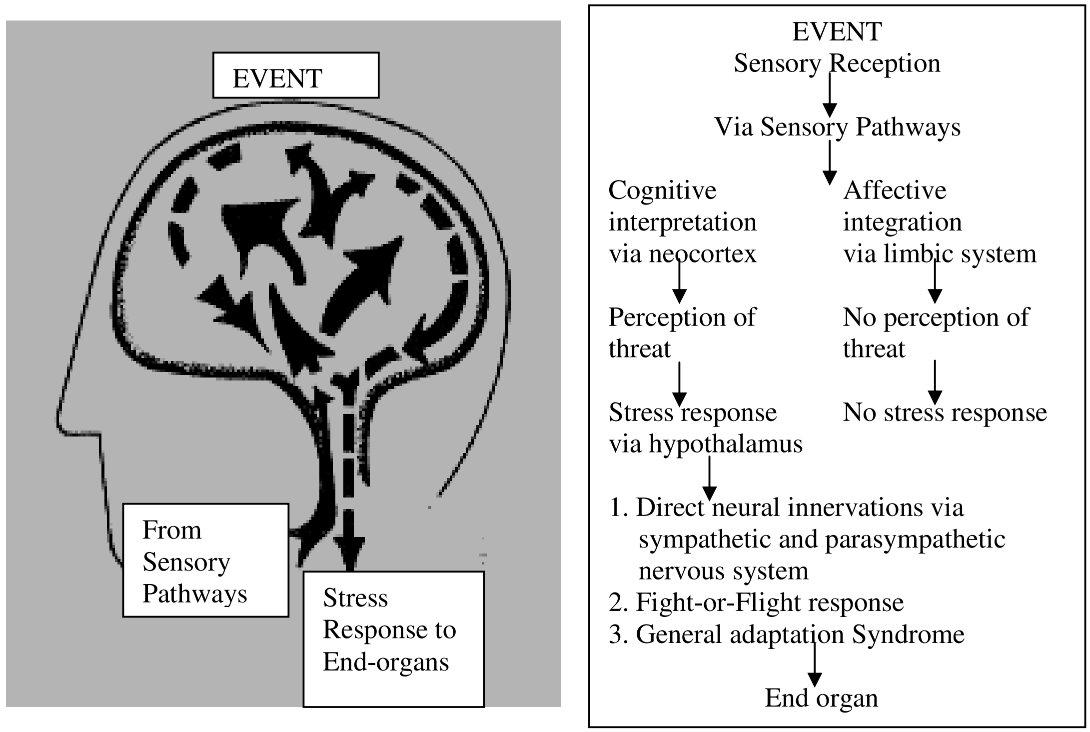


Figure 2.11: The stress response

All stressors will lead to a degree of psychobiological activation. The outcome, however, is strongly influenced by the individual's perception of the stressor. This would consequently influence work performance as well as the degree to which the physiological stress response is activated.

(b) The Psychosomatic Stress Axes

The stress response is of limited duration and is directed towards helping individuals cope with demands. It is therefore adaptational in nature. The non-specific adaptive reactions are directed towards mobilisation of the individual's reserves for energy, modulation of the homeostatic responses, and maintaining a high level of functional activity. However, when demands on individuals become excessive, the neuroendocrine activation associated with the stress response can become chronic and result in adverse effects on performance and health. Performance is, however, often maintained at the cost of health.

The difference between the specific stress response and the non-specific stress response lies in the degree of specificity of the stressor. Stressors that are very specific, such as cold and hunger, have their own specific stress responses that bring the internal homeostasis of the body back to normal through the normal negative feedback mechanisms of the body. However, any stressor that leads to psychobiological activation can give rise to the so-called non-specific stress response that involves, with minor variations, a relative non-specific neuroendocrine activation. This so-called non-specific stress condition or response can be seen as a new homeostasis meant to enable the individual to cope, mentally and physically, with the demands that initiated the response. Although the non-specific stress response involves virtually the whole body, as previously mentioned, two major neuroendocrine systems are involved in the peripheral expression of the response, i.e., the sympathoadrenomedullary (SAM) axis and the hypothalamo-pituitary-adrenocortical (HPA) axis (Viljoen, Claassen and Hazelhurst, 2004. See Appendix D).

The SAM and HPA axes are the two main stress axes that are activated after prolonged exposure to a stressor(s) and have therefore formed the basis of the physiological measurements of the verification study. The non-specific stress response varies widely from individual to individual (interpersonal differences) and even within one individual at different times (intrapersonal). This makes it very difficult to be definitive about the stress levels and stress responsivity of any individual. The discussion that follows will be limited to those physiological parameters that were possible to measure within the limitations set by the brief of this study, i.e., a non-intervention study within the daytime working hours of the TCOs.

(c) The Neuroendocrine or Sympatho-adrenomedullary (SAM) Axis

The SAM axis consists of the sympathetic nervous system and the adrenal medulla.

When mass activation of the SAM axis occurs it is referred to as the fight-or-flight response. The effect is to aid the individual in coping with environmental and other demands and entails a mobilisation of the body to prepare for muscular activity in response to a perceived threat. This mechanism allows the organism to either fight or flee from the perceived threat.

The pivotal system in this response is the adrenal medulla, which can be seen as an extension of the sympathetic nervous system. Stimulation of the adrenal medulla releases adrenalin and noradrenalin (catecholamines) into the systemic circulation. The effect of this release helps to increase energy availability and, in addition, leads to an increase in cardiac output, an increase in blood pressure, and a differential increase in blood flow (Viljoen et al., 2004. See Appendix D).

It is important to note that 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 could be interpreted 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 another, novel, stressor, the adrenal medullary secretory response may be significantly larger. This response is known as sensitisation (Viljoen et al., 2004. See Appendix D).

In terms of TCOs and their stress levels or, more specifically, their adrenal medullary stress responses, it would imply that they may adjust to the same workload when exposed to it over a period of time (habituation), but that the introduction of significant changes could lead to higher stress levels than before (sensitisation).

According to Everly and Rosenfeld (1981), specific somatic effects that have been observed in human beings as result of activation of the SAM axis in response to psychosocial stimuli are:

- increased arterial blood pressure
- increased cardiac output
- vasoconstriction
- increased muscle tension
- increased cholesterol levels.

(d) The Endocrine or Hypothalamo-pituitary-adrenocortical (HPA) Axis

The most chronic and prolonged somatic responses to stress result from the stimulation of the endocrine axes. Activation of the HPA axis during stress leads to an increase in, along with other neurotransmitters, cortisol, which in turn leads to increased access to energy stores as a result of conversion of other substances like lipids to glucose (Viljoen et al., 2004. See Appendix D).

The circadian rhythmic secretion of cortisol is characterised by peak secretion in the early morning just before getting up and during early morning activity. It levels off later in the day but may increase as a result of meals or any of a variety of stressors. One of the major effects of stress, including heavy workloads, unpleasant working environments or even continuous shift changes between day and night shifts, is, in fact, an alteration of the circadian cortisol secretory pattern.

Stress has several effects on the circadian rhythmicity of cortisol. Acute stress can cause acute increases in cortisol levels, which usually return to normal again within two hours. This is especially marked in the case of aversive stressors. Similarly to the SAM axis, incidents of acute heterotypic stressor application during periods of chronic stress alter the response. It would seem that acute stress superimposed on chronic stress is dependent on the familiarity with the type of stressor. It appears that the response to an acute stressor in the chronically stressed individual is 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 may be more intense. It is however possible that cross-tolerance to stressors may develop, especially with stressors that involve the same neurological pathways.

In theory this can be extrapolated to the working situation where high cortisol levels would not be significantly increased by increased levels of homotypic stressors such as increases in the amount of work to which the individual is accustomed. In contrast, the response can be exacerbated when atypical stressors are encountered. The implication is that individuals with chronically high cortisol levels will not show the expected increase from basal trough values when stress levels increase. This could be ascribed to the fact that the trough values are chronically higher than normal (Viljoen et al., 2004. See Appendix D).

The effects of chronically high cortisol levels will again be referred to in the discussion of the verification study results.

A discussion of the human stress response and the accompanying measures of stress associated with high workload would be incomplete without reference to allostatic load. In Addendum D the potential effects of allostatic load in the context of TCOs' job demands, and the resulting high mental workload, are discussed.

(e) Allostatic Load

Various types of allostatic responses to stress can occur. With the normal acute allostatic response, which is aimed at being of benefit to the individual, the response that usually occurs includes activation of the SAM axis and the HPA axis with subsequent increases in heart rate, blood pressure, blood sugar and other physiological changes. The response is terminated after the stressful event, and the physiology of the body returns to normal. As long as the allostatic response is limited to the period of work, or the demands on the individual are not excessive, protection via adaptation predominates over any adverse consequences. However, exposure to the allostatic response over extended periods of time can have psychopathological consequences. Chronically stressful situations and other factors may lead to abnormal allostatic responses which can, in turn, give rise to a high allostatic load.

The allostatic load can be seen as similar to the effect of the long-term or chronic so-called non-specific stress response. It is the price paid for the repeated or chronic activation of the allostatic response – a response which in the short term is intended to be advantageous. Another way of putting it would be to define it as the wear and tear that results from chronic over activity or under activity of the allostatic systems. The allostatic load is the product of accumulated stressors over an extended period. This includes the effects of minor day-to-day stressors and more dramatic or traumatic events. There can be no doubt that type of work, the working environment and working hours can be seen as major determinants of stress. In this the perceptions of individuals play a significant role. It is one of the major factors that determine whether they will become sensitised or habituated to a certain stressor, i.e., how they will cope with the different day-to-day stressors and for how long their neuroendocrine systems will remain activated. Other factors that can be seen as instrumental in modulating the characteristics of the stress or allostatic response are the general physical health and lifestyle of the individuals, i.e. exercise, diet, alcohol intake and smoking (Viljoen et al., 2004. See Appendix D).

From a medical point of view, as well as in terms of common human interest, an important aspect of mental load would be to establish whether it is possible to identify loads that may lead to pathological consequences.

The relationship between health and stress is both complex and confusing. When coping is possible, or when individuals evaluate the situation as one they can cope with, there is no evidence of pathology or of bodily changes that may be assumed to lead to pathology. When coping or control appears impossible, pathology develops, at least in acute animal experiments (Ursin and Ursin, 1979).

Therefore, if it is impossible to measure mental load per se, psychosomatic complications could indicate prolonged exposure to overload, even if it is indicated simply by the failure of coping processes. These psychosomatic complications depend not only on the mental load but on competence and subjective evaluations of load and capacity.

From a physiological viewpoint, a mental load must be assumed to be a load on processes within the central nervous system (CNS). Physiological indicators pick up activation rather than information processing. Activation is the end result of a wide range of psychological processes (Ursin and Ursin, 1979).

The use of physiological indices therefore rests on the assumption that it is possible to assess the amount of effort expended by operators in meeting the demands of the task. Changes in activation levels of subjects will ensue and these can be measured. The literature uses concepts such as arousal, activation, and effort interchangeably in this regard.

Activation theory is still a powerful model for the explanation of physiological and psychological mechanisms in wakefulness, sleep, and emotional states. This theory states that when there is information transmitted through the usual sensory pathways, there are also impulses sent directly to the reticular formation of the brainstem. The activation is not restricted to cortical activation, but also involves activity in the autonomic nervous system, the somatomotor system, and endocrine system. With improved methods for determining the plasma levels of hormones, it has become increasingly clear that the whole or at least a very large part of the endocrine system is subject to the influence of psychological factors. These phenomena are easily treated within activation theory (Mulder, 1979).

It is suggested that whenever physiological indicators are used to measure mental load, psychological and social factors must be assumed to be of decisive importance for the values obtained. Physiological indicators, therefore, pick up what may be called 'emotional' factors, or simply 'activation'. Activation depends on whether individuals expect to cope with their situation or not. The methods do not measure the information load. They give information on how individuals estimate the load, in particular whether they think they are able to cope with it, and to what extent failure or success is important to them. Therefore it is not the load that is measured but their evaluation of the load and their evaluation of their capacity to cope with that load. This depends on experience as well as personality characteristics.

The assumption that underlies the use of psycho-physiological measures of stress and mental workload is simple: as workload increases, so a corresponding increase in operators' levels of arousal is reflected in the activity of the autonomic nervous system. This level of arousal can be recorded by a number of psycho-physiological techniques. While no general pattern of changes in the various psycho-physiological variables in response to known changes in workload has emerged, it has been shown that certain measures demonstrate relative specificity to different workload components (Megaw, 2005).

Physiological measures such as heart rate, blood pressure, respiratory rate, eye blink, and pupil diameter are related to autonomic nervous system responses and hence are not under voluntary control or subject to deliberate bias or subjectivity. These measures can be taken without intruding on the primary task and can be indicators of mental workload and the stress emanating from the workload. They could also be especially helpful in the process of validating other techniques (Kramer, 1991).

Some of the physiological measures of mental workload that have yielded positive research results in mental workload assessment are described briefly below:

2.8.4.4 Measures of Brain Function

(a) Event-related Potentials

Event-related potentials (ERPs) are fluctuations in the endogenous activity of the nervous system that is recorded in response to environmental stimulation and that are associated with psychological processes or in preparation for motor activity. Various ERP components have been taken as indicative of information processing activity and changes in mental workload (Hancock, Meshkati, Robertson and Robertson, 1985). The ERP is time-locked to the stimulus event. Workload inferences are based upon the amplitude and latency elements of the elicited ERP wave (Meshkati et al., 1995).

ERPs represent the most valid physiological measure of workload, but limitations such as the considerable supporting instrumentation, sensitivity to environmental electrical noise and to the movement of the individual, the requirement of trained personnel for operation and interpretation, and sensitivity to individual differences, which require calibrations for each individual operator, make it a difficult measurement in practical terms and intrusive on the work environment.

(b) Electro-encephalograph Changes

Experiments have shown that mental activity affects the frequencies of the electro-encephalograph (EEG) but the relevance of this is confused by large inter- and intraindividual differences (Meister, 1985). This approach is extremely intrusive and creates an artificial work environment.

2.8.4.5 Measures of Heart Function

(a) Blood Pressure

Blood pressure is the outward force exerted against the walls of blood vessels. Four pressures were considered for the purposes of this study, i.e., systolic pressure, diastolic pressure, pulse pressure and mean arterial pressure. Increased blood pressure is a major complication of a stressful lifestyle. It is therefore a good reflection of the accumulated wear and tear of stress (allostatic load) on individuals (Viljoen et al., 2004. See Appendix D). However, the measurement of blood pressure as an indicator of acute stress often does not yield the information required to assess the magnitude of the stress response (Kaplan, 2000).

(b) Heart Rate

Heart rate is largely the product of direct or indirect sympathetic nervous system responses and the involvement of neurohormones. In contrast to blood pressure, the sensitivity of heart rate to stress-induced central nervous system activation is generally considered to render a justified index of acute psychological stress (Baker, Suchday, and Kranz, 2000).

Many examinations of heart rate fail to consider heart rate within the context of cardiovascular control mechanisms. When the environment (temperature, acceleration related to flight operations) is not benign, it may be impossible to discriminate cognitive effects on heart rate from environmental effects associated with physical effort and thermoregulation (Gawron et al., 1989).

Generally, increased heart rates are associated with increased levels of workload. Such increases have been found to vary with workload in several studies with pilots. However, not all investigators report consistent findings of heart-rate changes that occur with differences in workload (Wierwille et al., 1985). This inconsistency of results has caused a number of investigators to abandon simple heart-rate measures and to look instead at the variability of the heart rate as a possible measure of cognitive workload (Wilson and O'Donnell, 1988).

(c) Heart-rate Variability

Measures pertaining to heart rate and their derivatives are currently the most practical physiological method of assessing mental workload. Among such measures perhaps none is more thoroughly investigated than that of heart-rate variability (HRV). Several empirical investigations attest to the strength of this assertion (Steptoe, 1981 and Meshkati, 1988).

Heart-rate variability (HRV), which is based on small changes in the beat-to-beat or interbeat intervals (also referred to as R-R intervals), provides a non-invasive indication of autonomic function. HRV is normally measured by variations in the interbeat interval variability in heart rate. It is generally thought that increased sympathetic stimulation decreases heart-rate variability and that increased parasympathetic stimulation increases heart-rate variability (Woo, Stevenson, Moser, Trelease and Harper, 1992).

In a study on air traffic controllers, whose tasks show a number of similarities with those of TCOs, Kalsbeek (1971), using interbeat intervals, noted a gradual suppression of the heart-rate irregularity resulting from increases in the difficulty of the task.

Wierwille (1979), whose research also involved aircrew mental workload, is supportive of the above argument. He concluded that heart-rate variability decreased with increased mental load.

Several studies have since been conducted on the use of HRV as a quantitative index for mental workload (Kumashiro, 2005). The variance of the inter-beat interval times with increasing taskload has been confirmed in a number of experiments (Mulder and Mulder, 1987). Veltman and Gaillard (1998) investigated the sensitivity of physiological measures to mental workload on pilots in a flight simulator. They found that HRV was sensitive to large changes in task difficulty. Consequently, it was posited that such a measure could be used to reflect mental workload.

The following conclusions can be drawn from a review of the literature:

- Heart rate generally does not change with mental load. In some instances stress might cause a change in heart rate.
- Heart-rate variability may decrease with an increase in mental load.
- Blood pressure may increase with stress but is not a reliable measure of workload.
- The use of spectrum analysis of heartbeat intervals shows promise as a means of estimating workload.

(d) Respiratory Rate

This is one of the easiest measures to record and has been used extensively in studies involving emotion, stress, arousal and mental load (Meister, 1985).

2.8.4.6 Eye Function Parameters

(a) Eye Blink Rate

Eye blink duration and eye blink frequency are good indications of the onset of fatigue and/or when a person is under stress. It has been found that eye blink rate decreases under conditions of high workload both in driving tasks and in flight tasks (Hankins and Wilson, 1998). Laboratory studies have demonstrated that tasks requiring attention, especially visual attention, are associated with fewer blinks and blinks of a shorter duration (Wilson and O'Donnell, 1988). Infrared eye blink detection technology makes this a practical measure to use because it is not intrusive on the performance of workers. Developments include systems that detect and measure both eye blinks and slow closures since both types of eye closure reveal information about operators' alertness levels (Knipling and Wierwille, 1994).

(b) Pupil Diameter

Several investigators have observed that the diameter of the pupil correlates quite closely with the resource demands of a large number of diverse cognitive activities (e.g., mental arithmetic, short-term memory load, air traffic control monitoring load, and logical problem solving). This diversity of responsiveness suggests that pupilometric measures may be highly sensitive, although as a result, it is undiagnostic. Pupil diameter reflects demands imposed anywhere within the system. Another disadvantage is that relevant pupil changes are in the order of tenths of a millimetre, which means that accurate measurement requires considerable constraint of the head and precise measuring equipment. In addition to this, changes in ambient illumination must be monitored since these also affect pupil diameter. Because of its association with the autonomic nervous system, the measure will also be susceptible to variations in emotional arousal (Wickens and Hollands, 2000).

(c) Visual Scanning

Visual scanning, the direction of pupil gaze, although used as a measure of attention allocation, can contribute extensively to workload measurement in two different ways. Firstly, dwell time can serve as an index of resources required to extract information from a single source. Secondly, scanning can be a diagnostic index of the source of workload within a multi-element display environment (Wickens and Hollands, 2000).

2.8.4.7 Body Fluid Analysis

(a) Cortisol

A novel physiological approach to assessing mental workload was the investigation into psychoneuroendocrine responses to mental workload by Hyyppa, Aunola, Lahtela, Lahti and Marniemi (1983). They were able to find a significant decline in the cortisol and prolactin levels of subjects undergoing psychologically demanding achievement-orientated tasks.

Stress hormones measured in the urine, blood or saliva can be a good control measure to determine if operators have experienced stress over a particular period of time. This technique is reliable but costly, and often employees object to blood and urine samples being taken for analysis for fear that other substances may also be detected. Recently however saliva test kits have been made available to measure stress hormones in saliva samples.

Benefits and Shortcomings of Physiological Measurements:

Physiological indexes have two distinct advantages -

(1) They provide a relatively continuous record of data over time.

(2) They are not intrusive on primary-task performance.

In terms of their disadvantages, physiological measurements often require the attachment of electrodes (for ECG measurement) or some degree of physical constraint (pupilometric measurement), and therefore they are not unobtrusive in the physical sense. These constraints may influence user acceptance. Many physiological measures have a further potential disadvantage in that they are, generally, one conceptual step removed from the inference that the system designers would like to make, that is, workload differences measured by physiological means must be used to infer that performance breakdowns will result or to infer how operators will feel about a task. Secondary- and primary-task measures assess the former directly, whereas subjective measures assess the latter (Wickens and Hollands, 2000).

When using physiological measures, care must be exercised in their selection and application. In general, the sensitivity of these measures is task dependent. Physiological measures might be more useful in assessing the effects of time on task and task strain. Also, in terms of intrusiveness, physiological measures have been fairly successful indicators without influencing the primary-task performance of operators (Meshkati et al., 1995).

The use of physiological measures as indicators of mental workload is influenced by a combination of several factors, such as the cost of both the hardware and software required to operate the equipment, the training level of the personnel who administer the physiological tests, environmental conditions at the workplace and the willingness of employees to be connected to a physiological recording mechanism. Due to problems caused by various combinations of these factors at the present time, most physiological measures are impractical for use in the assessment of mental workload in complex machine-person systems. However, technological innovations, such as miniaturisation, telemetry and data-logging have reduced some of the problems, and as a result, the use of some of the techniques has become more feasible.

In this section the different groups of mental workload measurement techniques have been discussed. The purpose of this review was to obtain an understanding of the existing techniques and to consider which of these techniques would be appropriate and practicable for use in the Spoornet train control context.

From the available research it is clear that the various mental workload measurement tools have been applied in a variety of research environments. From the literature reviewed it is evident that the period between the late 1970s and late 1980s was the most productive in terms of mental workload research. During this period most of the papers and research results published were based on commissioned and sponsored research in military, aviation and aerospace contexts. Many reports, especially on the implementation of specific models or tools, are classified and not available in the public domain. A further significant observation is that since 1990 very little research has been conducted. The only papers available are on improved physiological measurements, and no new research has been conducted or reported on that focuses on new ways of assessing mental workload.

The only previous South African study on mental workload, undertaken by Cilliers (1992), was a doctoral study performed in the South African Air Force. The purpose of the study was to select a number of tools to assess the mental workload of fixed-wing fighter pilots. No new tools were developed and the study essentially entailed the correlation of various known workload assessment tools (all of which have been discussed in this chapter) with one another. No evidence of the implementation of the selected tools could be found, unless this is contained in a classified report that is not currently available in the public domain.

Another shortcoming is that the existing literature does not provide sufficient guidelines for the application of the tools in an environment other than for that which the specific tool was developed. Furthermore, there is limited information on the nature of tasks and simulation environments in which the research was conducted. These factors limit the degree to which the techniques can be applied in an environment unrelated to the research environment.

The groups of mental workload measurement tools and techniques reviewed here each have reported advantages and disadvantages. In conclusion, each of these groups will be critically evaluated against the set criteria and in terms of their appropriateness and applicability in the current context.

The following criteria for a mental workload assessment tool were identified as important in the Spoornet context:
Sensitivity
Selectivity/Validity
Intrusiveness
Operator acceptance
Affordability
Implementation requirements

The reasoning behind not utilising a simulated environment for the development and testing of a new mental workload methodology needs to be explained. The literature has many reports on the use of simulated environments for testing new methodologies, and it was the obvious choice for this study. Upon inspecting the training facility for the radio train order environment, it became evident that real-life scenarios could not be simulated and that fidelity would be a major concern. RTO training is the only training in the train control environment that cannot be simulated, unlike the situation where TCOs work with colour light signals that are monitored on a panel. To overcome this problem from a training perspective, newly qualified TCOs are never sent to an RTO section until they have gained sufficient experience in other environments. They then receive supervised 'on the job' training in an RTO control centre and are certified *in vivo* while under supervision before they can work in the RTO environment. The use of a simulated environment has therefore never been an option for this study. This has resulted in a number of limitations on methods that could have been used but due to the safety criticality of the tasks of TCOs had to be abandoned.

The rationale for not using existing mental workload measurement methods is provided below.

2.8.4.8 Measures of Primary-task Performance

Considering the context, this group of measurements would violate the criterion of intrusiveness. Given the safety responsibilities of TCOs, this is a critical requirement. The criterion of sensitivity would also not be satisfied as the task demands of TCOs are not consistent but rather occur in surges. For this reason degradation of performance would be difficult to detect.

2.8.4.9 Secondary Measures of Spare Capacity

This group of measurements, due to its nature, poses a safety risk and, in the context of the TCOs' tasks, would be intrusive. This group was the first to be rejected by Spoornet management and trade unions when the option of simulation was ruled out.

2.8.4.10 Subjective Rating Techniques

The very nature of subjective ratings, i.e., assessing one's own feelings of fatigue, mental effort, anxiety, stress load, or frustration, disqualified the use of this group in the Spoornet context. The reason for this decision was twofold:

- Spoornet is a highly unionised work environment. Union members will only participate in a study of this nature upon receiving a mandate from their union leaders. All trade unions were therefore involved as stakeholders from the inception of the project. Spoornet has been subject to retrenchments since the early 1990s, and during the period when this project was executed (2001-2004) the process had not been finalised and retrenchments were ongoing. Union leaders were adamant that their members should not be compromised in any way. Completing a checklist or rating scale on how employees experienced their shift, constituted such a compromise to union leaders. They were concerned that employees might be judged as incompetent and could therefore be retrenched on the basis of their experienced workload as captured on a self report form. Despite guarantees from the researchers that employees would not be identified, this group of measurements was not acceptable and therefore did not satisfy the user acceptance criterion.
- Secondly, subjective rating techniques are essentially the rater's perception of the difficulty of a task. Spoornet management was not prepared to accept the perceptions of individuals as a basis on which to make operational safety decisions, especially if such perceptions were tainted by concerns regarding job security. The requirement at the outset was that an objective assessment of mental workload was required. Furthermore, it should be quantified in order to prevent individual interpretation. Simply put, a number – and preferably a cut-off number within the limits of which mental workload is considered safe – should be attached to the mental workload of a specific TCO within a specific train control centre.

As indicated earlier in this chapter, the NASA-TLX is considered an industry standard as far as subjective ratings of mental workload is concerned. There is however no such standard for objective ratings. In the planning phase of the project it was decided to validate the newly developed MWLI against physiological measurements, as well as against the NASA-TLX, because of its status as an industry standard. In the verification phase of the project, which posed a tremendous number of challenges in terms of logistical arrangements (which is referred to in a later chapter) it became clear that the NASA-TLX could not be used. The reason for this was the fact that various train control centres had different shift patterns, depending on the size and activities at the centre. To simplify the measurement process, it was decided that all measurements would be taken between 06:00 and 14:00. For all centres this meant the start of a new shift (06:00) but not necessarily the end of a shift as some centres had 8-hour shifts and others 12-hour shifts. In order for the NASA-TLX to not be intrusive and pose a safety risk in this context, it could only be administered at the end of a shift. For many of the centres, the end of a shift meant 18:00 and the measurements were only taken over an 8-hour period starting at 06:00, irrespective of whether the shift was 12 hours in total. The reason for this decision was to ensure that unrelated factors, such as circadian rhythm, which would affect cortisol measurements, would be eliminated. Physiological measurements, as proven objective indicators of whether a workload problem exists, provide a more valuable validation methodology in this context and the NASA-TLX was therefore eliminated from the verification phase.

From this overview of the existing body of research in the field of mental workload it is evident that there is clarity as far as the factors that contribute to mental workload are concerned. A practicable index for the measurement and prediction of mental workload is, however, not available.

With this challenge at hand, and taking into account the criticisms lodged against some of the existing methods, an objective, user-friendly, non-invasive method of measuring the mental workload of train control officers at Spoornet had to be developed that could be used on operators while they were performing their tasks. The method had to be reliable and valid as well as verifiable and had to comply with the other criteria referred to previously for the selection of mental workload techniques.

The existing mental workload measurement techniques have to be considered in the context of the study in order to understand their applicability. This context is consequently provided.

2.9 CONTEXTUALISATION OF THE STUDY

2.9.1 Existing research

The problems that were faced in this study related to the existing body of knowledge as well as to the context within which the research was undertaken.

Since the inception of mental workload research in the 1970s, researchers have used a variety of definitions of mental workload. There is however not one single definition of mental workload that could be used to guide ongoing research. The definitions available have also evolved over the years, and engineering, psychological and physiological experts have all provided paradigm-specific perspectives.

Meshkati (1988) (Figure 2.5) has developed a single model of mental workload that incorporates all the factors relating to workload. It incorporates most of the definitions used by researchers and highlights the development processes involved. Although this model dates back to 1988, there is no more recent integrated model and it was therefore considered a useful basis for this study. As mentioned earlier, the literature indicates that no single definition of workload will suffice due to the complex nature of the concept. This model captures this multi-definition approach and also emphasises the multidisciplinary nature of mental workload. This study does not accept any single definition of workload as the paradigm from which a mental workload assessment model should be developed. The model considers the task variables related to workload, individual moderating factors, and various measurement techniques, although no specific recommendations are made for the objective assessment of mental workload.

The limitations in the existing research, which have also necessitated this study, relate to the fact that there is no clarity on measurement techniques that fulfil the set criteria. There is no readily available technique that could be used across a variety of contexts. The research also indicates that techniques applied and reported on in various contexts are situation and context specific.

In order to find a solution for mental overload in the train control environment, the air traffic control (ATC) domain seemed a logical benchmark.

Nunes (2003), in a study on assessing the impact of predictive aids on controller performance, came to the conclusion that the traditional need to design interfaces in ATC based on direct visualisation have been grounded in the notion of reducing controllers' cognitive demand. His study confirmed previously expressed concerns that such aids can have detrimental effects on the underlying mental model, which in turn can compromise operators' problem-solving and learning abilities. In this particular study the NASA-TLX was used as a means to measure mental workload experienced.

From this study it seems that the reduction of mental workload through the provision of predictive aids might compromise the performance of controllers. Personal communications with the South African Air Traffic and Navigation Services (AT&NS) (2004) indicated that no specific measurement or tool is used to detect and manage mental overload. Rather, mental overload is managed through limited exposure to a high load by keeping shift lengths to four hours over peak periods.

A study by Averty, Collet, Dittmar, Athenes and Vernet-Maury (2004) makes an interesting contribution to ATC mental workload measurement. This study reports on the development of the traffic load index (TLI). Mental workload assessment studies of ATC have used either objective measures, i.e., numbers and distribution of aircraft, or subjective factors, such as self-imposed performance and stress levels, with mixed results. The number of aircraft (N) was the most common index used. Due to safety constraints, subjective aspects of workload (strain) have never been assessed in real time during actual work sessions. In the study quoted here, it is hypothesised that N is not a perfect index and that a better index of mental load measurement could be obtained if the interventions carried out by controllers were integrated in the index. In the study data were collected on communications with pilots, radar material and the number of aircraft (a distinction was made between aircraft monitored and aircraft with a control problem). Time pressure and uncertainty was also factored into the TLI. This data was then quantified. Controllers were also requested to complete the NASA-TLX during the experiment. The experiment took place in the field but arrangements were made to complete the NASA-TLX between sessions. The results of the TLI were compared with N and the NASA-TLX. TLI is shown to be more highly correlated to TLX than N to TLX. Conversely, the correlation between TLX and N was lower, meaning that N does not exactly reflect the perceived workload. Future work with the TLI entails demonstrating its validity. Correlation with psycho-physiological variables to validate TLI is envisaged.

The similarities between the study by Averty et al, (2004) and the study reported on in this thesis are noticeable. These similarities are summarised as follows:

Table 2.5: Similarities between TLI and MWLI development

Factor	**TLI development**	**MWLI development**
Existing measurement(s) of mental workload (crude)	Number of aircraft (N)	Spoornet formula
Safety risks relating to field work – limits study	Air traffic control	Train control
Methodology: 1. Recorded communications quantified 2. Operational records 3. Objective data: Actions quantified 4. Subjective/additional data 5. Added measurement of workload	1. Communications with pilots 2. Radar information 3. Number of aircraft 4. Emotional processes linked to conflicts and time pressure 5. NASA-TLX	1. Communications with train drivers and others 2. Train schedules and plans 3. Number of authorisations and data transactions 4. Moderating factors 5. TLA
Validation	Psychophysiological measurements envisaged	Psychophysiological measurements used on limited population

Cilliers' (1992) work refers to two other South African studies, both in the military aviation context. The reports are referenced as confidential reports and were never published in the public domain. Cilliers' study was also conducted in the field of military aviation and she therefore had access to these confidential military reports. Besides these two confidential reports, Cilliers' (1992) study is the only study on mental workload that was ever undertaken in South Africa. Cilliers quotes and comments extensively on the existing body of knowledge and come to the same conclusion, namely, that no specific techniques are recommended, no guidelines exist on the tasks that were included in research projects, and few guidelines exist for simulated environments.

2.9.2 Research in the Rail Environment

Mental workload research in the rail environment has never been undertaken in South Africa. Internationally, a few published studies are available. The work undertaken by Devoe (1974) is one of the most comprehensive and informative studies undertaken for the purpose of identifying and understanding the tasks of railroad train dispatchers (the equivalent term for train controllers) and the environment and context in which they occur.

The past few years have seen a resurgence of interest in rail human factors and specifically mental workload. In the USA the Transportation Research Board (TRB) is the main contributor to research in this field. The author of this paper is a member of the TRB Taskforce on Railroad Operational Safety and therefore has access to contract research reports.

As mentioned before, between the late 1980s and 2003 very little new work on human mental workload was published. Previously developed subjective measurements, especially the NASA-TLX, have become industry standards, specifically in the aviation industry.

During 2001 two studies relating to dispatcher workload were sponsored by the Federal Railroad Administration (FRA) and comprehensive reports were produced (Popkin, Gertler, and Reinach, 2001 and Reinach, 2001). Elements of both these reports were incorporated in developing the MWLI.

The purpose of the Popkin, et al. (2001) study was, in the interest of railroad operational safety, to create a better understanding of the dispatching environment and its associated levels of workload, occupational stress and fatigue. The project goals were:

- Identify the sources and magnitude of workload, stress and fatigue associated with the railroad dispatcher’s job and working life
- Determine any related health or performance effects

- Refine procedures for measuring workload, stress and fatigue in the dispatcher's workplace.

There were three sources of workload data, including an observational technique, based on the Task Analysis Workload method (TAWL), subjective ratings and activity-count data. Salivary cortisol was used as a physiological measure of stress while actigraphy was used to record sleep patterns. The modified Task Analysis Workload (mTAWL) measurements provide a means to gauge variation in workload over the course of the shift.

The key findings of the Popkin et al. (2001) study were:

- Subjective workload was moderately associated with the reported number of trains dispatched, regardless of shift or location. Perceived stress also related to the number of trains dispatched.
- The mTAWL is labour intensive and not suitable for a research study.
- There was little evidence of high levels of stress from either subjective stress ratings or from salivary cortisol levels. 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 likely that data was collected too infrequently in this rapidly changing environment and may not have captured the changing workload and related stress. (The authors of the report noted that given that the data was collected from a small, non-randomly selected sample of dispatchers (N=20) from two dispatching centres, the results should be carefully interpreted.)

A technical memorandum (Reinach, 2001) was produced in response to concerns raised by a number of FRA audits over the safety of the U.S. rail network. Specifically, concerns were raised about the absence of a method that could be used to reliably collect dispatcher-related data (e.g., workload) as part of FRA safety and compliance inspections and audits of dispatching centres. A major consideration in collecting taskload data, which is the basis of an assessment tool, is that the tasks must be observable, quantifiable, quick, easy and unobtrusive to collect. The ultimate goal of the research was to produce a tool that FRA field personnel and others could use to reliably and quickly obtain taskload data. The technical memorandum was the first step in the development of a dispatcher taskload assessment tool. The rationale and ultimate goal of this particular study shows significant similitude with the Spoornet study.

Reinach (2001) identifies 67 unique dispatcher tasks, which were grouped into six categories. A number of factors that were expected to affect dispatchers' taskload were also identified. A number of recommendations for further activities were made before a taskload assessment tool could be developed.

Reinach (2006) elaborated on the above study and developed a performance model of railroad dispatching. It is evident that subjective measurements developed and refined for the aviation industry could not be superimposed on the rail industry. Taskload in this context was defined as the average time demanded of a dispatcher in carrying out all job-related tasks at a dispatching desk over one shift. The methodology of the study involved the collecting of taskload data through observations and questionnaires. The same 67 tasks from the 2001 study, identified and organised into six top-level task categories, were used. These were:

- actuation of signals, switches, blocking devices and bridge controls via centralised traffic control (CTC) or computer-aided dispatching systems;
- issuance and cancellation of dispatcher-authorised mandatory directives;
- granting of other track-related permissions, protections and clearances (non-mandatory directives);
- performing non-movement authority or non-permission/protection/clearance communications (This generally involves advisories, coordinating activities and the exchange of work-related information.);
- Performance of general record-keeping tasks;
- Review of reference materials.

These tasks represent the gamut of possible railroad dispatching activities across the United States, regardless of the dispatching technologies used or the nature or size of the operation. A number of other factors that either affect dispatchers' taskloads, or can be used to describe the circumstances in which taskload is measured, were also identified. Some factors are internal to the dispatchers, while other factors are external. In the

Spoornet study, similar factors were identified and referred to as *moderating factors* (see Chapter 3).

The following factors were identified in the TRB study:

- track-related factors
- railroad operation-related factors
- dispatcher-related factors
- other factors

The goal of this TRB study was to identify dispatcher tasks and data collection methods that would support the development of a dispatcher *taskload* assessment tool that could reliably measure observable dispatcher activity at different dispatching desks. This goal is very similar to the goal of the Spoornet study. Reinach's study concludes that:

- Taskload assessment has its limitations, particularly due to the highly cognitive nature of the job. The physical activities of dispatchers are observable but not cognitive activities.
- It is possible that dispatchers' performance may not be easily measured using only the physical work (taskload).
- The data gathered as part of this research provided a better understanding of the job of railroad dispatchers and contributed to the development of a preliminary model of dispatcher performance and safety that incorporates both physical and cognitive aspects of the job (see Figure 2.12).
- The proposed model, when fully developed and validated could provide a broader understanding of dispatchers' jobs. The model could be used to support a number of activities, including technological developments that might facilitate safety, more efficient dispatching, identification and development of criteria for more effective training programs, and monitoring the effects of changes in technology on the task of dispatching.

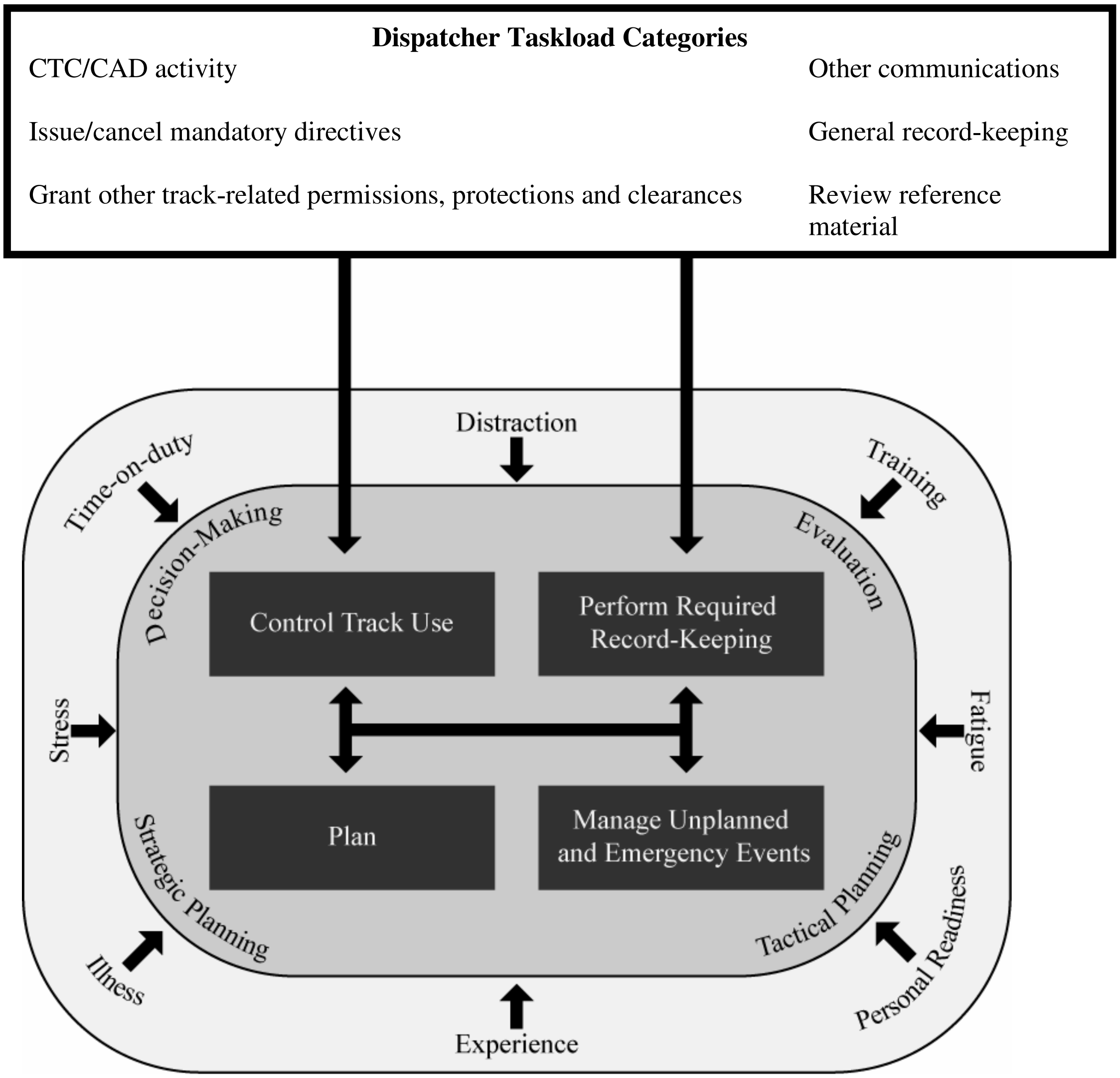


Figure 2.12: Preliminary Model of Railroad dispatching (Reinach, 2006)

In Europe the main contributor to human factors research in the rail environment is the Centre for Rail Human Factors at the University of Nottingham, (UK), which supports the Rail Safety and Standards Board in Britain with projects and research in the field of human factors. Two studies on mental workload from this centre were published in late 2005.

Pickup, Wilson, Sharples, Norris, Clarke and Young (2005) reiterate the same concern as the FRA, namely that mental workload in the signalling (or train control) environment is an area that needs to be better understood and practically assessed. They propose a conceptual framework and a suite of workload tools.

Pickup et al. (2005) acknowledge that mental workload is a complex field of study that starts with various definitions and a plethora of associated measurement methods. A further difficulty is that researchers have often employed one of a number of well known (self-report or subjective) MWL measurement methods, in the hope of setting their results against some norms, but as a result may be trying to use a benchmark with a measure that has been created for an entirely different purpose in entirely different circumstances.

The context of the work of Pickup et al. (2005) is very similar to the Spoornet context. A brief account of their research findings is given here in order to contextualise the Spoornet study.

The motivation for the research was the need of Network Rail (the owners of the rail infrastructure in the UK) to understand and measure the workload of signallers (elsewhere called dispatchers or controllers) in the context of technical and organisational system change. Network Rail wanted to derive a 'function complexity index' based on the principle that there should be a (ideally linear) relationship between the number of inputs received and the activities of signallers and their subjective perceptions of mental workload. This early approach was, thus, based on examination of the relationship between the demand for resources imposed by a task and the ability to supply these resources by the individual. The approach of this study was therefore based on the capacity concept, referred to in the operational definition. At some point (or *red zone*) that perceived level of workload would be considered unacceptable and, therefore, a maximum number of inputs/activities could be identified. This would allow a specification as to the number of trains or amount of track or other parameters individual signallers could cope with. In doing this, there is a move towards a set of quantitative criteria and an analytical model, so that system designs that provide appropriate workloads could be defined.

This need to quantify workload and to be able to predict the expected workload at a specific train control centre was the desired outcome of the Spoornet study.

A number of problems were identified with the Pickup et al. (2005) approach, of which the most important were:

- Difficulties in distinguishing between inputs and outputs (e.g., is setting a route an influence on workload or an indicator of workload for a signaller?).
- Different qualitative impacts of different inputs on workload, e.g., setting routes or handling emergency phone calls.
- The need to account for individual differences in work strategy and workload perceptions.
- A large number of discrete and combined tasks that are completed as part of the signalling job.
- Change in performance shown, either through performance decrement or strategy change, before maximum workload capacity is reached.
- As systems move from being mechanical to computerised, there is less active, observable intervention; workload moves from the physical to the mental and is internalised to maintain situation awareness; and it becomes increasingly difficult to use observation measures.

This work (Pickup et al., 2005) provides a detailed account of MWL in signalling and follows with a presentation of a conceptual model. Primary elements in the model – such as loading factors, effort, demand and effects – are examined in detail. Finally, existing and proposed new workload measurement tools are introduced and positioned within the framework.

Mental workload in rail signalling was found to relate to the following elements:

- the number, complexity and interaction of tasks that are performed – over a period of time or at one point in time,
- the load as subjectively experienced – over a period time or at one point in time,
- the number of functions that have to be performed in different situations and scenarios, and
- the compatibility of working arrangements with the functions that need to be completed.

The main distinction is between workload as imposed by the system, somehow measured independently of individuals and their ratings of the 'work' that is 'loaded' on them (Hart and Staveland, 1988) and whether this reflects perceived task difficulty or level of effort exerted.

These multiple meanings, particularly as MWL is seen as an index that measures what people have to do and how they feel about it, have been the cause of great confusion in the rail industry. This is compounded by the fact that the concept of workload has a highly intuitive meaning for most people. Thus, the context in which mental workload has to be understood usually directs the interpretation of its meaning and the choice of assessment methods.

The multiple views of the term mental workload in the rail industry have directed the focus of this study to consider workload in a functional context. An understanding of signaller function was gained by reviewing goal and task analyses carried out as part of previous projects and by carrying out new interviews and observations as described earlier. From the task analyses, interviews and observations, it was suggested that signallers at the highest level have three main functions (See Figure 2.13).

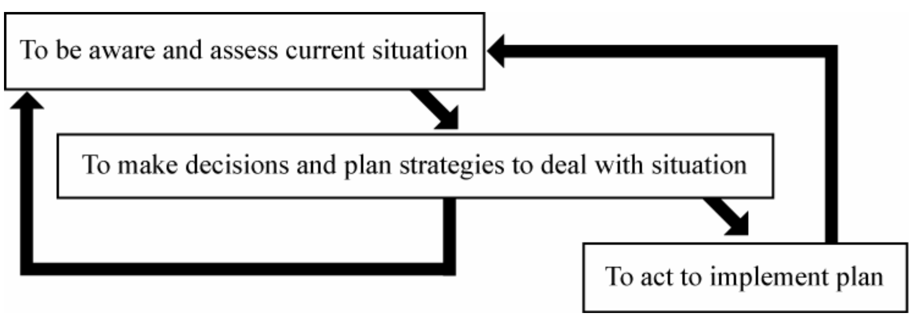


Figure 2.13: High level signaller functions (Pickup et al., 2005)

Workload tools previously used with rail signallers have included NASA TLX, timeline analysis and the Air Force Flight Test Centre (AFFTC) workload assessment scale. The interviews with rail human factors specialists and the personal experiences of those involved in the UK study indicate that these tools are not able to capture all aspects that can be considered relevant to workload of signallers. Often they are not appropriate for use in the field with civilian populations. Many tools only assess the workload of a single task (e.g., the NASA TLX, the Bedford scale and SWAT) whereas the studies of signallers and their environment suggest that multiple tasks are regularly completed concurrently. Many traditional workload assessment tools like SWAT only aim to assess dimensions that are related to high workloads (Reid and Nygren, 1988). The range of tools to be adopted by the rail industry should be capable of identifying situations of both underload (or at least low load) and overload, as both can impact upon health and safety and productivity. According to Pickup et al. (2005) the NASA-TLX is probably the most widely used workload measurement scale and extremely valuable in many situations. However, it was found to be impossible to use in real time in the field. Added to this, the wording appears to be a little obtuse for civilian and European audiences. More critically, it is not fully representative of the influencing variables most relevant to railway signallers.

These findings confirm the conclusion that was reached regarding existing tools and their appropriateness for use in the Spoornet context. This gave rise to the need to develop a customised tool for Spoornet.

Pickup et al. (2005) considered the development of a conceptual framework comprising the relevant dimensions of workload for the rail industry (and especially signalling) as essential support for a practical approach to workload measurement that may be usable in real situations. This conceptual framework has been developed gradually, expanded and refined in light of the relevant literature and also – critically – the authors' own studies of and with signallers. As a consequence, the evolving framework has been determined through a synthesis of empirical findings and theoretical interpretations and this synthesis is reflected in how the framework is explained.

The first simple conceptual framework of mental workload was based on the model proposed by Jahns (1973), which included input load, effort and performance. The starting framework reflected the effects of mental workload on both system performance and the wellbeing of signallers (Figure 2.14).

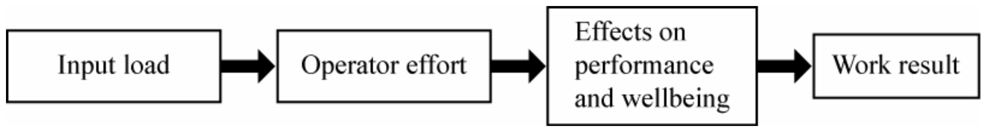


Figure 2.14: Adapted from Jahns' (1973) model (In Pickup et al., 2005)

The revised conceptual framework is shown in Figure 2.15 (below). Pickup et al. (2005) emphasise that this framework is not necessarily an operational model. It is proposed in order to develop and position a toolkit of methods with which to understand and assess mental workload. Thus, it explains the routes to measurement rather than the mechanisms that contribute to MWL. The framework also reflects the very different ways in which workload is conceived of and, therefore, measured by different investigators.

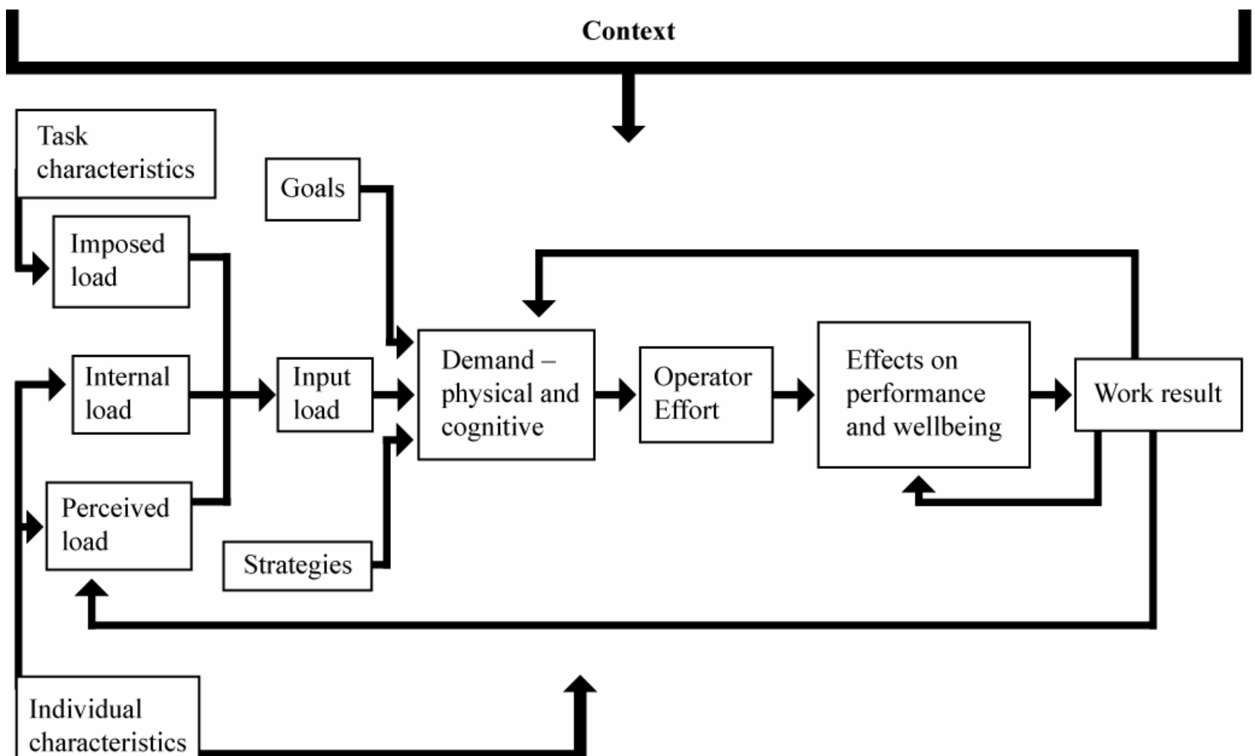


Figure 2.15: Developed conceptual framework of mental workload (Pickup et al., 2005)

The central elements of the framework are load, demand, effort and effects. Load is defined in terms of it being imposed and perceived. The interview analyses suggest different types of loading factors, that are categorised under several headings, for instance, organisation, environment (internal and external), situation (normal, abnormal or emergency working), team and job design, equipment, social situation, management and feedback on performance. There was considerable discussion during this study as to whether and how to represent context and individual differences within the framework. There is agreement that both context and individual differences may moderate the operation of the framework. However, questions such as whether imposed load and goals should be considered to be a part of the context, and whether experience, age, physical and mental wellbeing are part of individual differences or of context are still being debated. Goals and strategies were added as important intervening factors.

The analyses indicated that signallers viewed effort as a consequence of the demand created by loading factors. The perceived level of effort necessary to accommodate demands was related to:

- complexity of decision making required and options available,
- time sharing of activities,
- predictability of the situation,
- perception of the level of control within a situation,
- interruptions and distractions from the main activity,
- perception of likelihood of achieving intended goals (in relation to time, experience, organisational or environmental constraints),
- consequences of actions on the safety or performance of the system, and
- likelihood of being blamed for actions and existence of a blame culture.

The broader needs of workload measurement in the UK rail industry, especially those of Network Rail that were required for signalling, were referred to earlier. Obviously, no one tool would meet all those needs. For instance, assessing the number of functions that should be performed in a particular context requires a different form of measurement to that which would be required for assessing the load experienced at points in time by the person performing those functions. These different needs and the different understandings of the notion of workload have led to the conceptual framework embracing facets of load, goals and strategies, demands, effort and effects. The care taken to distinguish these notions means that care should also be taken to define measurement approaches and methods for each. In addition, various criteria against which to judge the value of workload tools have been defined by rail industry clients to include: allow assessment of acute load and also load over time; be diagnostic; be predictive; allow tracking of peaks and troughs of load; support direct assessment and be analytical; and reveal qualities of the performance. These are all in addition to the normal requirements for validity, reliability, acceptability, sensitivity, etc. No one measure will meet more than a few of these criteria. Understandably, there is no need to invent new methods and tools where existing ones will do. One use of the conceptual framework has been to clearly distinguish the purpose and focus of existing workload tools, to clarify what is available, even if it has been used and sometimes validated in other contexts. The conceptual framework also gave structure to the assessment of the potential value of the tools - for rail and for field, as well as for simulator use - undertaken with a sample by rail human factors specialists. The outcome of this exercise suggested that many tools currently available did not meet the needs of rail signalling, some had been used but with alterations in wording or method of administration, and one or two had potential but would need to go through trials and possible adaptation.

In terms of contextualising the Spoornet operational safety challenge, and in terms of appreciating the problem this research attempts to confront and resolve, the recent work undertaken in the railway environment and discussed here certainly provides a better understanding of the specific requirements of this industry. It also provides insight into the reasons why other well-known ‘industry-standard’ MWL tools are not useful in the railway context.