5

Stellingen

behorende bij het proefschrift van

F.R.H. Zijlstra

Efficiency in Work Behaviour
a design approach for modern tools

1. Arbeidspsychologisch onderzoek in het laboratorium is goed mogelijk en tevens in toenemende mate noodzakelijk.
(Dit proefschrift)

2. Indien we over menselijk gedrag spreken in termen van 'gedragseconomie' en daarbij vasthouden aan begrippen als 'psychologische kosten', moeten we ons realiseren dat men in psychologisch opzicht ook failliet kan gaan.
(W. Schönpflug, (1983). Coping efficiency and situational demands. In: G.R.J. Hockey (ed.), *Stress and Fatigue in Human Performance*. Chichester: Wiley)

3. De noodzaak tot introductie van het begrip 'psychologische efficientie' wordt ondersteund door het feit dat slechts in de helft van de gevallen de door de persoon als meest efficient gepercipieerde route ook daadwerkelijk de meest efficiente route is.
(P. Bovy, (1986). Routekeuzegedrag. *De Psycholoog*, Vol. 21(12), 616-620)

4. Effort blijkt in belangrijke mate samen te hangen met motivatie. In dat verband is het te verwachten dat als vervelend gepercipieerde taken ook meer effort (inspanning) kosten.

5. Het gebruik van Likert-achtige schalen met verbale labels leidt doorgaans tot een onderschatting van het meetresultaat.
(French-Lazovik, G., and Gibson, G.L. (1984). Effects of verbally labelled anchor points on the distributional parameters of rating measures. In: *Applied Psychological Measurement*, vol.8 (1), 49-57.)

6. Het feitelijk gebruik dat van PC's wordt gemaakt rechtvaardigt de roep om nog krachtigere processoren niet.
(R.A. Roe, et al., (1993). *Mentaal Informatie Werk*. Work and Organizational Research Centre, Tilburg. rapport.)

7. De 'Kuhniaanse' paradigma wenteling is niet van toepassing op maatschappij wetenschappen.

8. Ook een vijfjarige ingenieursopleiding zal onvoldoende blijken te zijn indien de gedragswetenschappelijke aspecten in de opleiding onderbelicht blijven.

9. Het probabilistische karakter van sociale wetenschappen is veel ingenieurs een doorn in het oog.
(Delta, 24(39). 17-12-92, p.8)

10. Indien een ernstige kritiek op het werk van Taylor is dat zijn mensbeeld te mechanistisch is, moeten we ons ernstig zorgen maken over het mensbeeld van ontwerpers die hun systeem graag 'monkey proof' willen maken.

11. "It is mei sizzen net te dwaan".

Efficiency in Work Behaviour

A Design Approach for Modern Tools

Efficiency in work behaviour

A design approach for modern tools

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus,
prof.ir. K.F. Wakker,
in het openbaar te verdedigen
ten overstaan van een commissie
aangewezen door het College van Dekanen
op dinsdag 23 november te 14.00 uur

door

Ferdinand Rudolf Hendrikus Zijlstra,

geboren te Bolsward,

doctorandus in de psychologie.

Dit proefschrift is goedgekeurd door de promotoren:
prof.dr R.A. Roe
prof.dr G. Mulder

CIP-GEGEVENS KONINKLIJKE BILBIOTHEEK, DEN HAAG

Zijlstra, Ferdinand Rudolf Hendrikus

Efficiency in work behaviour; a design approach for modern tools /
Ferdinand Rudolf Hendrikus Zijlstra. - Delft: Delft University Press. - Ill.
Thesis Technische Universiteit Delft. - With ref.
ISBN 90-6275-918-1
NUGI 659
Subject headings: efficiency / work behaviour / mental effort.

Grafisch ontwerp en lay-out:
Elle van As

Voorwoord

Een proefschrift is een proeve van bekwaamheid met betrekking tot het kunnen uitvoeren van wetenschappelijk onderzoek. Het zal reeds door menigeen duidelijk gemaakt zijn dat dit naast wetenschappelijke kwaliteiten, ook organisatorische kwaliteiten en uithoudingsvermogen vereist. Met name die laatste twee eigenschappen zijn niet alleen in de professionele sfeer van belang, maar ook daarbuiten. Dit impliceert tevens dat niet alleen de auteur van het proefschrift er de handen vol aan heeft (gehad), maar dat er ook vele anderen hun steen(tje) hebben bijgedragen. Dit is ook hier het geval.
Een woord van dank aan een ieder die op eigen wijze heeft bijgedragen aan het tot stand komen van dit werk is dan ook op z'n plaats, hetgeen bij deze dan ook gebeurt: dank u!

Fred Zijlstra — Delft, september 1993

Ter nagedachtenis aan mijn ouders

Contents

Chapter

1 Towards the design of modern work tools

1.1 Introduction

As long as people have used tools and equipment for their work they have endeavoured to improve them and make them more suited to the task in hand. Since the days of Frederic Winslow Taylor improving work tools and working-methods has been regarded as a science. In January 1912 Taylor presented the results of his research on 'the science of shovelling' to the Social Committee of the House of Representatives in the United States of America (Taylor, 1972). He described how people had always been shovelling all kinds of materials, without anyone ever having told them how to do this. In his research at the Bethlehem Steel Works, carried out in 1885, Taylor noticed that each worker had a shovel of his own and that each person used his shovel for heavy ore as well as for lighter materials. When they shovelled heavy ore the workers had about 17,5 kilograms on their shovels but when they shovelled lighter materials they only lifted about 1,5 kilograms. From extensive research with different sizes of shovels Taylor learned that a worker's productivity level was highest when he had about 10 kilograms on his shovel. If the weight was increased the worker had to rest more often because of the overload and if the weight was lighter the worker had to rest quite often because of muscular fatigue caused by having to move faster. Moreover it appeared that the best way of shovelling depended on the kind of materials being shovelled. Nevertheless all the workers shovelled all the materials in the same way.
Taylor also reported on the work of Frank Gilbreth. Gilbreth had analyzed the work of bricklayers. He noticed that bricklayers required 18 operations for each brick. Some of these operations, like bending to pick up the bricks were very tiresome. After some time Gilbreth succeeded in finding another, more efficient, way to do the job. Only five operations were now needed for each brick and the most tiring operations could be left out.

Although Taylor is best known for his 'Scientific Management' study (Taylor, 1911), which was chiefly seen as a plea for a differentiation between 'head labour' and 'hand labour': so-called 'Taylorism', he actually tried to improve workers' equipment and working methods. Taylor tried to help the labourers to find optimal ways to execute their tasks by abolishing traditional, non scientific ways of working, and replacing them with more scientifically based, efficient ways of working.

Taylor was the first to study actual human work behaviour. His results could only be obtained by extensive research into people's work behavioural patterns: their working-methods, their skills and how they actually used their equipment. In short, he employed a work psychological approach. He became famous for his 'time and motion' studies (Taylor, 1907, 1972). These methods have been extensively used by engineers. By observing people at work and registering how much time the various operations took, Taylor came up with the 'one best way' to carry out the task in each case. Going on his ideas and the results of his studies Taylor formulated directives that could be used by engineers to (re)design the tools and equipment. Efficiency in work was the key concept in this approach and high productivity was the ultimate criterium.
However, one of the drawbacks of Taylor's ideas was that the workers no longer had any say in their own working-methods, strategies, working speed, and so on. History has shown us that this is, in another respect, rather inefficient because it creates poor and meaningless jobs and demotivated workers. Moreover, workers usually have valuable knowledge and insight into the work process which may be helpful when the general aim is to optimize productivity and improve the equipment. In addition it prevents workers from regulating effort expenditure which can cause 'workload problems' (cf. Chapter 5). Nevertheless Taylor's work may be regarded as a valuable contribution to work psychological research as far as the improvement of work tools and equipment goes.

The study described in this book also focuses on the improvement (redesign) of tools. However, there are some differences between this approach and Taylor's approach. First there is a difference in the objectives and points of view. In this study I will be concentrating on 'modern' cognitive work tools, like computers. These tools differ to some extent from 'traditional' tools, as will be illustrated in the next paragraph.
Secondly, Taylor's main objective was to optimize output. To this end he selected employees who were strong and fit enough to do the work and he prescribed the tools and working-methods that should be used. Taylor's assumption was that there is always 'one best way' to carry out a task. This idea remained popular for quite a long time. The assembly-line for mass-production, which still exists, is a living proof of this. Also in the way that 'modern' tools are designed we can still trace elements of Taylor's ideas. Designers of automated systems and application software for computers usually design the software and interface according to their own ideas about how a task is, and should be executed, which implicitly may lead to working-methods being prescribed. However, according to modern psychological insights into working there is no 'one best way' to carry out a task. Workers should be able to choose their own working-methods and strategies (this point will be elaborated in Chapter 2). Consequently it can be argued that the user of the tools should remain in control of the working-method.

The point of view that will be dominant in this study is the notion that designers of tools - especially of cognitive tools - should bear in mind that their products should support the user in carrying out his task and should not prescribe how he should do it. Knowledge of human work behaviour may help designers of tools to improve their products.

With the emergence of modern information technology tools and equipment are becoming more and more powerful. Nowadays computer equipment is used in most office situations and industrial processes. The introduction of computers has led to a change in the demands being placed upon the worker (see also Zijlstra & Roe, 1988; Aronson, 1989). Various technologically highly developed systems are being produced for use in offices, to control (industrial) processes or else to perform highly complex tasks like flying aeroplanes. But these systems always have to be operated, controlled and supervised by human operators.
In order to ensure that these modern tools and equipment are used effectively and efficiently one has to be sure that they are well enough adapted to the tasks for which they are intended and correspond with the worker's working-methods. Furthermore designers should be aware of the physical and cognitive capabilities and limitations of the people who have to operate, control, and supervise these systems. Training and instructing the operators may not be of much use when the way in which these systems have to be operated is not sufficiently adapted to their ways of working. Examples of such mistakes can be found in errors made in the past and in the present and especially emerge when analyses of accidents, like plane crashes, are made or when malfunctions at nuclear power plants are brought to light. Faults in 'dialogues' between humans and (personal) computers can have serious consequences as well, e.g. documents and files can be lost. This means that in order to improve modern tools we must, like Taylor, study the way in which these tools are used. Like Taylor I believe that efficiency is a crucial concept in this context. However, while Taylor focused on the efficiency of the overall production process, I will be concentrating on the 'costs' created for an individual in realizing output, i.e. completing the task. Taylor's criterium which is more closely related to 'productivity' will therefore be superseded by a *psychological* efficiency criterium.

1.2 Modern Work Tools: A New Challenge

There have been considerable changes in the work domain since the days of Taylor. In a socio-economic respect Dutch society has changed from being an agricultural society to being an industrial society and has more recently turned into a service society. Recent employment figures (CBS, 1991; Roe, et al. 1993) show that almost 70 % of the Dutch working population is employed in the 'service sector'. A great deal of the work in this

sector can be described as 'office-work' where the core activities concern handling information so it can therefore be called: 'Information Work' (Meijer, 1989, Roe et al. 1992) or 'Knowledge Work' (Porat, 1977; Schäfer, 1988).
Furthermore there has been considerable technological development. Information and communication technology is something that has been developing with amazing rapidity (see also Nickerson, 1986; Ten Horn & Zijlstra, 1990; Roe, 1990). Technological development has had its impact on the tools that are used in office work. One of the most important tools nowadays is the 'personal computer'. Today more than 50 % of the working population appears to use a computer regularly during the working day. Within the group referred to as 'information workers' this percentage is even much higher: about 88 % (Roe, et al., 1993). Consequently when computer systems do not function optimally this may lead to great time loss and to considerable financial losses not to mention the personal distress that might be caused to those who have to work with systems that do not function optimally. Recent figures show that a considerable number of people still have many complaints about various software applications. These complaints vary from: feeling not in control of the system (11 %), to not understanding how the system is actually processing the files and documents that have been worked on (26 %), to establishing that the system does not always operate according expectations (33 %) (Zijlstra, et. al., 1989, de Ronde, 1992). This illustrates the relevance of a study that aims at improving modern tools.

Unlike the tools of Taylor's time the tools of today require predominantly cognitive operations. Working with a computer means dealing with 'virtual' objects that is to say, for example, with texts that only 'exist' on a screen or in a digitalized form on a disc and are not tangible unless they are printed out. The text can only be manipulated by pushing 'buttons', which can be perceived as 'second-order' control (Wickens, 1984).
Computer systems are operated by means of application software. Both the software and the hardware make the system into a tool. However, the functionality of the tool may vary according to the software that is implemented, while the hardware and software put together will determine the usability of the user-interface. So we are in fact dealing with 'virtual' tools and 'virtual' work objects. Hence the reason that work has become more abstract and therefore more complex and now makes considerable mental demands upon the operators.

In this study I will restrict myself to these 'virtual' tools, and in particular to the usability aspects of the user-interface.
However, this does not mean that the approach and instruments described in this book could not be implemented to (re)design other equipment, e.g. large scale computer systems. To my mind there are striking parallels with

man-machine systems of the type that are used in the process-industry, in aeroplanes etc.

1.3 Human costs as a design criterion

The design process can be regarded as an iterative process that constantly forces the designer to make choices (Eekels, 1982). This automatically applies also to designing computer systems and their interfaces. It is practically impossible to foresee all the detailed consequences of every choice with respect to the interface of the interaction between the worker and the system. Sooner or later the designer therefore has to evaluate his design solutions. Usually design products are extensively appraised in the final stages of the design process when the designer wants to test his product under conditions that resemble those that will be encountered in real situations.

Most designers of large scale complex computer systems are well aware of the fact that the systems they design have to be operated and supervised by human beings. Hence the reason that they are interested in questions like: 'what are the limitations of the mental capacities of operators?' 'to what extent can operators be charged with such a task, and for how long?' The mental demands that are imposed upon the operator are considered to be a relevant criterion for the evaluation of Human-Machine Interfaces. As far as large-scale Man-Machine Systems are concerned this has been recognized for several years now, especially in the aviation-industry (cf. Kantowitz et al, 1983, 1984). For instance at NASA research institutions various research programs have been set up to measure the workload that is imposed on pilots when they fly technologically advanced aeroplanes. In this respect it is interesting to note that in the past Cooper and Harper developed a rating scale (the Cooper-Harper scale - Cooper & Harper, 1969) that was intended to evaluate the 'flyability' of planes. They had no explicit definition of 'flyability' but they asked experienced pilots to evaluate aircraft against this scale (see also Moray, 1982) because they realized that if they were going to evaluate these planes and possibly improve upon them they would have to establish how pilots managed to handle them.
As far as small-scale office systems are concerned the idea of taking people's mental demands into account is rather new. Literature on the evaluation of Human-Computer Interfaces (HCI) centres around the rather hybrid concept of 'user-friendliness'. From the very start this concept of 'user-friendliness' has lead to contradictory views on Human-Interface design. The following statements concerning the timing of error messages on the screen are illustrative in this respect: Martin (1973) recommended that error messages should not be given too abruptly, because "..... a split-second error response in midthought is jarring and rude". On the other hand

Shneiderman (1979) "......suggests that human performance improves if errors are indicated immediately and the disruption of user thought processes by immediate interruption is not a serious impediment". These quotations illustrate that where computers are concerned 'user-friendliness' has sometimes been taken too literally. Even nowadays lots of systems prompt the message when started up: "Good Morning, mister Jones". Later on Stevens (1983) warned that the 'dialogue' between humans and computers should not be made too friendly. It might mislead users and give rise to expectations on the user's side about functions that we can never expect computers to meet.

A concept like 'user-friendliness' may derive from the fact that computers have a diversity of potential uses. In contrast to the more 'classical' appliances, there is no clear relationship between the shape and function of the tools any more. Traditional tools and machines had just one function: a hammer is made to hit a nail on the head, a typewriter can only be used for typing, a coffee-machine is for making coffee and so on. An object's functionality was expressed in its design. In other words, there was a clear relationship between form and functionality. However, with the introduction of computers all this changed. Computers are multi-functional, they can be used as word processors, calculators, archives, instruments in laboratory experiments and for many other purposes. The fact that computers can perform so many functions - depending on the software that is used - makes them rather vague, indistinct, and amorphous for a number of people. The fact that people are fascinated and discouraged at the same time by the great potential of computers may be what has led to the emergence of a (meta-)metaphor in which a computer is personified as a colleague or companion (usually a male one because of its technical origins). It is the dominant perspective in common parlance, also in commercials, where computers are usually addressed with the pronoun 'he'. One can 'interact' with a computer, have a 'dialogue' with it and the computer can be said to 'understand' or not 'understand' what one means. Computers are called 'intelligent' systems even though they merely perform predefined arithmetic operations. The term 'intelligence' really refers to a set of cognitive resources including creativity, problem-solving, conceptual thinking, flexibility and so on that are reserved for human beings, just as the notions of 'interacting' and 'dialogue' refer to inter-individual activities that are unique to human beings.
Designers of computer systems have (un)intentionally adopted this perspective of the computer as a person[1] (Kerr and Hiltz, 1982). This meta-metaphor may have given rise to the emergence of the concept 'user-friendliness' which is nowadays very often used as the ultimate criterion for

1 Ironically cognitive scientists view a human as a processor of information and quite often represent the human information processing system as a digital computer.

evaluating systems. The diversity of the idiosyncratic interpretations for this concept has amongst other things led to the sort of situations referred to earlier in this section. It must be clear that it is very difficult to attain generally accepted quality standards for computer systems on this basis.

The problem of establishing quality standards for computer systems may be approached by regarding working with such systems as a double-task. The mere operation of the system (determined by the user-interface) may be regarded as a secondary task while the task itself (e.g. compiling and editing a text) which relates to the functionality of the tool constitutes the primary task. One could say that the more attention one devotes to the primary task the better it will be, but with all the attention that is needed to operate, monitor, or supervise the system a person cannot devote all his attention to the primary task. It would be more efficient if the secondary task did not require so much attention as that might improve primary task performance, or enable operators to work at lower 'human cost'. Since human attention span is limited, it becomes clear that systems should be designed in such a way that the demands on the mental reserves of the operator are as low as possible. One could say that the optimal interface is the interface that yields the best performance and imposes the minimum of load on the operator. In this respect designing an interface can be termed as an optimalization problem. An optimal interface is an interface that can be operated efficiently by an operator, i.e. provide the best performance at minimum 'human cost'.

1.4 Approaches to the design of tools

Technology is one of the most influential factors upon human work behaviour and consequently upon the results of such work. This means that from a work psychological perspective several requirements should be specified which the designer of man-machine systems, or work tools, has to take into account. These requirements relate to conditions for optimal human work behaviour, such as criteria for the quality of the operator's task.
With respect to the design of technical systems several approaches can be distinguished:

1. *Priority for technology.* This means that trying to automate all that it is technically possible to automate and leaving the remaining tasks to human beings. This can be regarded as the first stage in the history of system design based on Taylor's ideas about division of tasks which in turn were based on a 'mechanical' model of man (cf. Fitt's list, e.g. Mackay & Whittington, 1983) where it is assumed the operator will react when prompted by a signal or sign. It appeared that the resulting tasks were monotonous and had very adverse effects on the workers (high absenteeism and sickness, low productivity, etc.).

2. *Optimal division of tasks between Man and Machine.* This approach is directed towards the division of tasks within the Man-Machine System. For each system that has to be developed one should decide what tasks can best be performed by the machine and what tasks should be assigned to the operator in order to optimize total system performance. From a work psychological point of view the criterion should be that 'meaningful' and 'complete' tasks are created (cf. Hackman & Oldham, 1975; Hacker, 1986).
3. *Facilitation of task execution.* More recently a third approach to system design has been formulated (Roe, 1984, 1988): The Action Facilitation Approach. The Action Facilitation Approach (AFA) consists of a set of guidelines for Man-Computer Interface design stemming from work psychological principles, as formulated within Hacker's Action Theory (1978, 1986). In this theory the process of 'action control' is of fundamental importance (see also Chapter 2). It is assumed that workers actively plan their actions which implies that they should decide themselves how and when to act. The AFA guidelines aim at pointing the design process in such a direction that systems will facilitate the planning and supervisory activities of workers.

For a growing number of people the computer is becoming a multi-purpose tool. Users may engage in several different activities like text editing, file searching and mail processing in a single work session with the computer. From a work psychological point of view there is no 'one best way' to carry out these tasks. Each individual actively arranges his own work activities and tries to find the optimal way to carry out the task in question. This means that in contrast to Taylor's approach the worker should be able to choose his own working-methods and strategies. From this point of view workers should have autonomy or 'decision latitude'. Formulated in terms of design: workers should be able to 'design' their own interaction with the computer.
The aim of the Action Facilitation Approach is to facilitate the work process of the individual worker. The AFA guidelines should help designers in this respect.[2] It has been hypothesized that implementation of the AFA design guidelines will lead to 'action facilitation', which can be operationalized as 'maintenance, or even improvement, of performance at lower individual costs'. Or, to put it another way: action facilitation has to do with improving the efficiency of individual human actions.

This opens the way to a work psychological approach to design and, in particular, towards the evaluation of computer systems. From the work psychological point of view the user's personal efficiency is the main criterion alongside of which computer systems, or perhaps it would be better

2 It should be noted that autonomy for the worker should not be interpreted as 'non-design' for the professional designers.

to say interfaces, should be evaluated. This gives rise to a conceptualization of efficiency at the level of the individual known as: *psychological efficiency*. This concept refers to the 'costs' imposed on an individual in relation to the performance-level ('benefits' of work behaviour) required (by the organization). This concept of efficiency will be elaborated upon in the second chapter.

1.5 Aim and plan of this study

The central theme of this study is the improvement of modern work tools. Taylor's approach, which was largely based on 'time and motion' studies is, in my opinion, no longer valid. To start off with the psychological processes that regulate human work behaviour are not observable and were not taken into account in Taylor's approach. This point is extra relevant to 'cognitive' work tools which require predominantly cognitive operations. At best one can only observe the outcomes of cognitive operations and then only to the extent that they require some motor response. Consequently Taylor's method of observing and timing the worker's operations is not relevant when it comes to studying cognitive work. Since cognitive operations are not observable we need some theoretical notions about the cognitive processes that are involved in regulating human work behaviour. These theoretical notions will be discussed in Chapter 2. However, Taylor's ideas still seem to have some adherents amongst designers of systems, albeit in a modern style (cf. Card, Moran, and Newell, 1983).

This study aims at presenting a scientific contribution to the (re)design of tools. This contribution should consist of an examination of the value and applicability of a psychological efficiency concept as an evaluation criterion for modern tools.
This general goal incorporates several sub-goals:

a. Development of a methodology for the evaluation of interfaces in which psychological efficiency is the main criterion. The concept of efficiency will be elaborated upon and the applicability of this concept will be examined.
b. The development of a procedure or instrument to measure the 'human costs' involved in task execution. Since computer work is to a large extent mental work, this means measuring the 'mental costs'.
c. An empirical validation of the Action Facilitation Approach. The Action Facilitation Approach has been shown to be applicable to interface evaluation and leads to useful suggestions for systems redesign (Van der Velden et al., 1989). But until now the Action Facilitation Approach has not been empirically evaluated.

The plan of this study relates to the above mentioned sub-goals.

In Chapter 2 some general principles of behaviour economics will be discussed and I will focus on the concept of efficiency that will be used in this study. The next step will be to examine whether these efficiency concepts can be incorporated in existing engineering theories.

As has been said earlier in this Chapter I will focus on the aspect of 'human costs'. Chapter 3 will therefore be dedicated to elaborating on the psychological costs. Efficiency relates to the trade-off between obtaining a goal (completing a task) and the costs associated with attaining such a goal. Since computer work is to a large extent associated with cognitive functioning and mental processes the relevant costs in this respect have to do with mental demands, mental workload, and mental effort.

In Chapter 4 the focus will be on subjective ways of measuring mental effort. Most subjective methods are relatively easy to apply and are fairly cheap. The development of a rating scale to measure mental effort will be discussed together with some results of the initial validation of this scale.

Chapter 5 will be dedicated to the experimental validation of the rating scale for measuring mental effort. In this chapter an experimental study will be described in which the validity of the rating scale has been examined according to the theoretical model of effort discussed in Chapter 3.

This instrument for measuring mental effort has been applied in a technical context. Chapter 6 will describe a study which aims at evaluating the interfaces of two word processors with respect to the degree to which they facilitate human work behaviour or allow people to work efficiently. The word processors in question were especially selected before the study began and it was assumed that they differed in few of the critical aspects formulated within the Action Facilitation Approach. The study consists of two parts. In the first part the word processors are evaluated by experts by means of check-lists; the second study of the study constitutes an experimental evaluation. From this experiment we hope to get information about the applicability of the efficiency concepts and the methodology that is employed in this experiment. The study reported in Chapter 6 was designed in such a way that it can be regarded as a study on the empirical validation of the Action Facilitation Approach.

Finally Chapter 7 will contain concluding remarks based on the preceding chapters. The work psychological and cognitive psychological aspects which will be set out in this book will be brought up again in Chapter 6. Problems that have arisen will be discussed and put in perspective in the last chapter.

Chapter

2 Efficiency in work behaviour

2.1 Introduction

One of the aims of this study is to develop an interface evaluation methodology. With evaluating the traditional approach has been to use check-lists. Amongst the complaints generally levelled against check-lists is the claim that they always contain insufficient items, that the most crucial items are often missing and that it is unclear whether all the items in the list are of equal importance.

More recently it has been acknowledged that it is better to use a 'criterion oriented' approach to evaluation. This means specifying a criterion for the evaluation of the product in advance. Such an approach is more flexible than the check-list method because it allows different criteria to be specified for different systems and situations and takes into account the different needs of various groups of users. This approach is also more complicated because it means that before a system can be evaluated one has to establish what the users actually want or expect from the product. These expectations may be context-dependent and may also depend upon individual needs and preferences.

In order to avoid allowing such an approach to lead to sheer individualism and idiosyncrasy one also has to find some general characteristics which may serve as criteria. I think I have found such a criterion in the concept of efficiency. In my view people will continue to use a system (or tool) when they perceive that system as a real aid to the carrying out of their task, in other words: when the system enables them to carry out their task more efficiently. I assume that in their work behaviour people generally strive for efficiency.

In contrast to the situation in the days of Taylor we now have at our disposal of a whole body of knowledge about human work behaviour. Some of the principles of human behaviour are widely acknowledged. Examples of such principles are that a part of human behaviour, notably work behaviour, is goal-directed (Watson, 1919 - although he called it 'purpose' directed; Russell, 1921) and that this goal-directed behaviour is hierarchically organized (Miller et al., 1960). These notions can be found in many of the psychological theories presented nowadays (cf. Sabini et al., 1985). One of the theories which extensively elaborates these notions is Hacker's 'Action

Theory' (German: 'Psychologie der Handlungsregulation' - 1973, 1978, 1986). This theory falls into the work psychological domain but it is rapidly becoming popular in the area of system design. At the moment Hacker's Action Theory is one of the most complete theories on work behaviour. It is therefore interesting to see whether this theory can offer a basis upon which we can elaborate notions of efficiency. Another reason for looking into Action Theory is the fact that the Action Facilitation Approach, already mentioned in the previous chapter, has its roots in this theory.
In this chapter we shall be examining Action Theory to establish its usefulness for our purposes but first we shall focus on the concept of 'psychological efficiency'.

2.2 Efficiency in behaviour

An interesting observation from daily life is that people prefer to take the shortest or most time-saving route in traffic or to pick the shortest queue at the post-office, etc. The most obvious explanation for this phenomenon is that people are aware of their 'limited resources', whether they be financial or less tangible - like time or the amount of energy available - and so endeavour to be economic with regard to their resources. There have been studies (Sperandio, 1972; Teiger, 1978; van Aalst, et al., 1986) which have demonstrated that people actively regulate their effort-expenditure by choosing different working strategies in order to be able to complete the day's work. This can also be regarded as striving to behave economically.

The general utility of 'behaviour economic' concepts is supported by many examples taken from daily life, from the work domain (making a telephone call while a document is being printed out) and from observing how tools are used. A very illustrative and relevant example of trying to be as efficient as possible can be found in a study by Winsemius (1969) on the occurrence of accidents. He describes a situation in which a welder had to use various tools in order to construct a particular object. At a certain moment while using a welding-iron he briefly needed a screwdriver. Instead of putting down the welding-iron to pick up the screwdriver he kept it in his hand and put the screwdriver in the same hand. When a sudden movement was made the welding-iron rotated and the welder was burnt. The welder had thought that it would be more efficient if he could shortcut the process of switching off the welding-iron, putting it down, picking it up again later, switching it on again but this time it did not work out as planned.

This notion of 'behaviour economy' was introduced by Schönpflug (1985) in connection with theories on coping with stress. He applied the logic of economics to behaviour regulation. Coping with stress is conceived as problem-solving behaviour. Coping adequately behaviour may result in reducing the level of stress while coping inadequately may lead to sustaining or even increasing the level of stress. Conceived thus stress is interpreted as losing resources or is seen in terms of the psychic costs that are associated with having control over one's situation. As Schönpflug formulates it: 'States of stress are states of limited and overcharged resources and individuals involved in stressful encounters operate to save resources' (Schönpflug, 1985: p. 85). This means that behaviour economics has to do with reducing the 'costs' that are connected with a particular pattern of behaviour. This concept of 'behaviour economics' will be a basic element in the approach to efficiency in the evaluation methodology that is to be described in this study.

2.2.1 Efficiency and 'behaviour economics'

Work behaviour may be conceived of as series of specific actions and operations that are undertaken in order to reach certain (predetermined) goals. It is assumed that a goal may generally be reached in one of different ways: in this respect one might speak of a set of action alternatives which together make up an action space. Apart from having a conceptual action space known as 'A', which comprises all the action alternatives that are conceivable in connection with a given goal one may define an individual action space, let us say 'A_i' which embraces the specific subset of action alternatives which are covered by the cognitive map of the individual 'i'.
Action alternatives may be seen as cognitive representations of possible actions. They may have various shapes, e.g. they may be linear or branched.

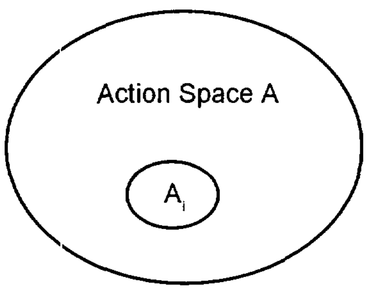

Figure 2.1: *Graphical representation of the Action Space (A).*

While at the level of actual activity just *one* course of action may occur, the concept of an individual action space seems useful for understanding the choice of an action alternative during action preparation (i.e. planning), or the shift of alternatives during action execution. This can be illustrated with the example of a cyclist taking a certain route from his house to the office. The will cyclist usually be aware of the fact that there are several possible routes but he will only actually choose one particular route. Only when his preferred route is blocked will he start to look for another way to reach his goal.

Being engaged in performing work tasks the individual expects to encounter certain outcomes. First there is the 'result' of the work process: the work product, or output. This may be conceived as the realization of the 'goal'. Secondly, there are the rewards (financial, social, etc.) to be obtained upon attainment of one's goal. Thirdly, there is the opportunity to learn, i.e. to consolidate or extend one's repertoire of skills. Furthermore, there will usually be some negative, but unavoidable, 'outcomes' as well: working will cause fatigue and sometimes stress. The extent of these negative consequences will depend on the effort one has to make which in turn will depend amongst other things on the difficulty of the task and the person's level of aptitude for the task (see also Chapter 5). These latter effects, which tend to be perceived as unpleasant are usually of a temporary nature. However, certain specific conditions, like having insufficient rest, may cause irreversible damage to a person's health (Meijman, 1989).
The worker's abilities in terms of his knowledge of the relevant procedures, skills, etc. and also his psycho-physiological condition (or rather his momentary state) will determine to what extent he will get tired at work.
On the other hand these factors also determine which strategy the worker will follow during task execution, provided the working methods have not been prescribed in detail and the worker does in this respect have decision latitude. This means that the worker selects the action alternative that seems most appropriate to him at that particular time bearing in mind his own capacity at that moment.
Since people differ in their various individual characteristics like in their knowledge and skills it may be obvious that various workers will choose different strategies, i.e. action alternatives, to reach similar goals (equifinality).

From a behaviour economics point of view it is assumed that people will tend to avoid or will diminish the negative consequences and optimize the positive result(s) of their actions. In the previous section some examples from daily life were mentioned. These notions imply that people have some awareness of the 'costs' of their goals and that they strive towards reducing these 'costs'.
This model implicitly assumes that people make rational choices with respect to the various behaviour alternatives that are available to them, i.e. represented in their cognitive map. It is assumed that during task execution workers will choose those action alternatives, or perhaps one should say strategies that in their view involve the least cost.

These conceptualizations show some resemblance with Vroom's 'Expectancy-theory' (1964), and Edwards' 'Subjective Utility Theory' (1954). These theories were used to explain an individual's motivational processes. However these models concentrate on the expected 'value' or 'utility' of a certain benefit (i.e. the goal that is to be obtained). The costs that are

associated with obtaining this goal are not explicitly mentioned in the mathematical formulas developed to calculate the strength of the motivational processes.
The value of these types of models is that they explain how processes might work, but they are difficult to apply to predicting individual behaviour. Individual behaviour depends partly on the demands of the situation (task), which can vary. Furthermore individual behaviour is determined by the personal preferences and abilities which can change during the course of time.

In conclusion one may state that an individual's choice behaviour is dynamic and therefore difficult to predict. Moreover there is clearly an individual component in choice behaviour which means that people will differ in the alternatives they choose. These differences are created by differences in the task situations and varying preferences and abilities and therefore in the varying perceptions of what might be the most appropriate or efficient strategy. This implies that efficiency in work behaviour has to be conceived as 'psychological efficiency'. It means that efficiency should be defined on on an individual level. Psychological efficiency relates to the individual's perception and/or experience of what is the most efficient way to carry out the task or, to be more specific, which strategy (or action alternative) is required to obtain the goal (i.e. 'benefits' of work behaviour) that involves the lowest 'costs'.

To illustrate the different perceptions of efficiency we could examine various groups of users of information systems. Users are expected to differ with respect to the amount of effort they are willing to invest in learning how to use a system.
For instance, several writers have suggested that compared to technical specialists managers are less likely to be willing to invest much time in learning how to use computer-based tools (Damodaran, 1981; Eason, 1981; Stewart, 1981). Managers often have someone else to operate the computer on their behalf. Specialists may be willing to invest considerable time and effort in learning how to use computer-based tools provided that these tools are really going to help them to perform their tasks.
Clerical workers are less likely than either managers or specialists to have the freedom to be able to choose not to learn how to use computer systems. If their job requires that they use a computer they will have to learn how to use it. How long it takes them to learn and how effectively they use the system will depend very much on the extent to which they perceive these tools as being really helpful (Nickerson, 1986).

2.2.2 Defining psychological efficiency

Efficiency refers to a ratio between costs and benefits. With respect to efficiency in work behaviour it must be clear what are the benefits and what

are the costs of work behaviour that should be taken into account. Since we have established that work behaviour is primarily 'goal-directed' the 'benefits' of work behaviour should in general be defined as being connected with goal attainment i.e. with completing the task. Task completion is usually accompanied by realizing a certain prescribed output. This output may be defined in terms of quantity and/or quality, though quantifying output may present many difficulties in certain professions. This implicitly means that when a goal is not achieved, i.e. when a task is not completed, one cannot speak of there being efficiency. Therefore effectiveness (attainment of the goal) can be regarded as a necessary, though not the complete prerequisite for efficiency. Being efficient also implies limiting the 'costs' associated with goal accomplishment.
Indicating the costs of work behaviour may be more problematic. Although some authors refer to the concept of 'psychological costs' (cf. Schönpflug, 1986a), by which they usually mean: involving 'resources', this point needs some clarification. The basic question in this respect is, which 'resources' do people account for when they estimate their 'costs'.

From the previous section we can understand that groups of users differ in the extent to which they are willing to invest time and effort to learn to work with a system. This may be a good indication as to the 'resources' we have to consider in this respect. Effort can be seen as a 'resource' in an economical sense, because one's 'energetic resources' (physical or mental) are scarce and one must choose how to invest one's effort. So, the effort that is spent can be termed the 'psychological cost'. The amount of effort that has to be invested can be conceived as being related to the complexity, or rather the usability of the system. It may express the complexity of the actions that have to be carried out, i.e. the number of operations that have to be carried out, the knowledge that is required to operate the system, the amount of information that has to be processed, etc. in short, all factors that may be a burden to the user and make the secondary task (i.e. operating the system) more demanding. Effort investment relates to the extent to which people get tired from their work. It appears that fatigue reduces work behaviour efficiency (Meijman, 1991). This means that the psycho-physiological state of the individual should be taken into account as well. This point will be elaborated in the next chapter.

As has been mentioned in the previous paragraph, effort investment may vary from individual to individual, even within the same task domain. So, the psychological costs can be defined as the amount of effort an individual has to invest in order to operate a particular system. What logically follows from this psychological efficiency is the cost/benefit ratio for each individual, i.e. the perception of the amount of effort the user has to invest in relation to the goal he intends to obtain. In work situations the goal is usually laid down by the organization as far as output quantity, quality, or both matters go. This

also implies that individuals do not usually decide about the level of performance (i.e. benefits of work behaviour) and therefore do not have much say in this respect. The choices they usually have are related to working methods and strategies, i.e. behaviour alternatives. And these are associated with the 'costs' of work behaviour. For this reason I will primarily focus on the costs of work behaviour and will not take the 'benefits' into consideration.
It has been hypothesized that from the options available to him an individual will go for the behaviour alternatives that allow him to complete the task in question which demand the least amount of psychological cost. This will be the criterion for evaluating computer system interfaces in the interface evaluation study described in Chapter 6.

2.3 Approaches to interface evaluation

During the last ten years evaluating interfaces has become an important field of research. The currently predominant view of human-computer interaction is, that it is essentially an information processing activity. The worker and the computer are thus said to engage in an active dialogue in which the two 'parties' exchange and interpret information. According to this view the ease with which users can achieve their intended goals will depend primarily on the degree of 'cognitive compatibility' between the human and the computer in the way it is presented to him. Most of the researchers in this field have adopted a cognitive psychological point of view (in so far as any theoretical models are used - cf. Howard & Murray, 1987), or certain pragmatic rules of thumb derived from these models (cf. Gardiner and Christie, 1987).

Cognitive psychological models have a serious weakness with respect to interface evaluation as they are built upon knowledge obtained from experimental research in the laboratory. The tasks used in most of this kind of research are 'context-free', that is, they are abstractions of real-life tasks and are therefore - though they may vary in complexity - simplifications of real tasks. This reflects the difference in emphasis between the cognitive psychological and work psychological disciplines. Cognitive psychology focuses on elementary aspects of human information processing while the work psychology approach focuses on human work behaviour which has to do with more comprehensive situations of the sort we may be confronted with in daily working life. The tasks that are described in Chapter 5 may be regarded as examples of tasks that are used in the cognitive psychology domain.
The tasks of daily life working situations are generally more complex in that they demand a range of cognitive operations in various combinations. This makes it almost impossible to give an exact description of the tasks in terms

of all the required cognitive operations and use them for experimental purposes. For research purposes the best way is to make some kind of taxonomy of behaviour requirements in order to be able to then classify human task performance (cf. Miller, 1962; Posner, 1964; Fleishman & Quaintance, 1984). The work of Newell and Simon (1972) on human problem solving behaviour has been especially important in this respect. They distinguished between two kinds of tasks: a) tasks that require problem solving behaviour and b) tasks that have to be carried out according to predefined strategies and/or algorithms. Newell and Simon regarded problem solving as goal-directed behaviour but the sequence of activities which should eventually result in obtaining the goal that has been set is not yet clear. Unlike 'problem solving tasks' the latter sort of tasks do not require problem solving behaviour because the sequence of operations needed to carry out the task is familiar.

If a user of a system is not familiar with the strategy that is needed to operate that system he may be said to be in a problem solving situation. Whether a person is in a problem solving situation or not is determined by the way the person (with his knowledge) interacts upon the task.

In my opinion we cannot simply string together models of the several cognitive processes involved in achieving any given task and call this a theory of human work behaviour. In particular the motivational processes of attachment - or devotion - to the task, which might severely affect cognitive functioning are not dealt with in many models of information processing. Some writers do however acknowledge that motivational aspects affect the effort 'allocation policy' (cf. Kahneman, 1973). Another point is that physical activity and fatigue (i.e. the energetic aspects of work behaviour) may also influence the processing of information. This point is not taken into account either in most cognitive psychological models.

Although the earlier 'data-limited' models of human cognitive functioning regard the human as a passive 'consumer' of the input information Treisman (1969) and Rabbitt (1979) have shown that with extensive practice subjects can discover and use ever more efficient ways of processing information. (Note: 'data-limited' means to say that information processing is limited by the quality of the data, not by the resources invested). This aspect is relevant to the acquisition of skills (Anderson, 1981, 1982; Leplat, 1989).

Therefore the later 'resource-limited' models do account for changes of strategy: processing the information is limited by the amount of resources that are available for processing. These models postulate having a central decision mechanism which is in control of system input characteristics and which can reset these characteristics from time to time as well as change the allocation of resources. Studies of speed-accuracy trade-offs suggest that there is considerable flexibility in the allocation of resources and they illustrate how different strategies may be operated at different levels of information processing, all depending upon the demands of different tasks.

Any comprehensive model of information processing that will eventually be developed is likely to be highly complex. It should involve interdependence between the cognitive processes which will vary according to the requirements of the task and its context and will be affected by individual characteristics such as: available knowledge and style of information processing. Such a model would be difficult to use for evaluation purposes. In real life the task in hand requires various combinations of problem solving behaviour and routinized procedures.
Nowadays there is a great variety of ways of approaching system design. As explained earlier I will be adopting a work psychological stance since the above mentioned aspects (motivation, energetics, strategies) are covered by the work psychological approach. I will therefore be discussing an engineering approach from the work psychological angle: Hacker's Action Theory.

2.4 An engineering approach: Action Theory[1]

Action Theory (Handlungstheorie, later also 'Tätigkeitstheorie'), as described by the (formerly East German) work psychologist Hacker in his 'Allgemeine Arbeits- und Ingenieurspsychologie (1973, 1978), may be thought of as a heuristic framework in which various models and hypotheses from cognitive psychology, learning psychology, and ergonomics are integrated. The foundations of this theory stem from the cybernetic TOTE-concept (Test-Operate-Test-Exit) of Miller, Gallanter and Pribam (1960) and the ideas of Soviet psychologists like Rubinstein (1962) and Leontjev (1964, 1979) with regard to activity and motivation.
Action Theory tries to integrate knowledge from various (psychological) disciplines about the determinants, processes, and consequences of work behaviour. Action Theory has clear connections with Newell and Simon's (1972) work on problem solving, with performance theory and ergonomics (Sperandio, 1980; Bainbridge, 1979) and with socio-technical job enrichment theories (Trist & Bamforth, 1951; Trist, 1981).
On the other hand when it comes to talking about the relationship between the individual and the material world (objects) around him a lot of elements derive from Soviet Russian psychology. Central to this concept is the contention that the individual consciously creates and changes his own environment by acting (German: Tätigkeit). This process of making and changing one's environment is embedded in a social context: it involves the participation of other people and it makes people feel that they belong to this society, which provides sense to their lives.

1 The description of Hacker's Action Theory has also been drawn from Neuberger (1985), Roe (1988) and Roe and Zijlstra (1991).

From this broad concept of 'activity' (G: Tätigkeit) Hacker narrows down to the more specific concept of 'action' (G: Handlung - introduced by Lewin, 1926):

"Unter 'Handlung' verstehen wir die kleinste psychologische Einheit der willensmässig gesteuerten Tätigkeit. Die Abgrenzung dieser Handlung erfolgt durch das bewusste Ziel, das die mit einem Motiv verbundene Vorwegnahme des Ergebnisses darstellt. Nur kraft ihres Ziels sind Handlungen selbständige, abgrenzbare Grundbestandteile (Einheiten) der Tätigkeiten" (Hacker, 1978, p. 62).

An 'action' has a hierarchical-sequential structure, an 'action' itself can be broken down into: partial actions (G: Teilhandlungen), operations, and movements. What is characteristic of these lower levels is that their goals are not consciously aimed at and they are not 'psychical' acts in the sense that they have their own motives and goals. The function and meaning of these lower-level operations and movements is supposed to be derived from the action of which they are a part.
In Hacker's Action Theory the process of action control (Hacker uses the term 'regulation' of action) is of fundamental importance. In Hacker's view the work process is only partly determined by its organizational and technical context which leaves several possibilities for the worker to intervene. In other words: there is not 'one best way' to carry out a task, but the worker has some degree of autonomy in arranging his activities. This degree of freedom within a task may be said to form an operator's 'decision latitude' (G: Handlungsspielraum). This means that the operator must, to some extent, decide himself how and when to act (or react). He will therefore need some directives upon which to base his decisions. The cognitive representations (internal model) of the variety of behaviour alternatives (G: Operative Abbild Systeme - OAS) that each individual has at his disposal and the associated outcomes will provide a model. These internal models that consist of already existing programs about how to go about reaching certain goals are crucial to the regulation of actions. These programs are stored in the memory system and can be recalled when required.
These internal models should not be regarded as exact (photographic) representations of reality but as individual (and thus selective) simplified, higher-level, schematical plans of the processes that are to be controlled. They work something like this: "if you want to buy a book, you should go to town and find a bookshop; you can get there by foot, by bicycle, by bus, by car, etcetera". Situation specific adjustments and differentiations are always possible.
The internal models are supposed to control the execution of actions and through the experiences one gets (feedback) these models will be completed, modified, corrected or differentiated.

It is not assumed that every single action is meticulously represented in detail, like in a computer program. Action regulation can take place at various levels depending on the worker's experience with regard to the task, regulation of actions can take place at various levels. Hacker defines three regulation-levels, which correspond to the findings of Rasmussen (1981) in his research carried out with process-operators:

(1) Intellectual level of regulation (knowledge based). This is the highest level of (conscious) regulation of actions. Action programs are being developed on the basis of intellectual analyses of the situation and syntheses of possible strategies. Rather complex and abstract action programs are produced. With this approach new and unexpected situations are encountered. The resulting action programs are executed consciously and predominantly fall into a serial mode, interpreting feedback step by step. This intellectual level of regulation bears some resemblance to Newell & Simon's (1972) concept of 'problem solving behaviour'.
(2) The perceptual-conceptual level of regulation (rule based). At this level preparing actions does not require extensive and detailed analyses of the situation but the execution of previously established action programs is triggered by the perception of cues and signals. For instance, on the basis of certain symptoms a doctor will diagnose influenza without having to think very deeply and will prescribe some kind of appropriate medicine. In Newell & Simon's terms (1972) there is no problem solving behaviour required in these situations.
(3) Senso-motoric level of regulation (skill based). At this level the actions are almost completely automatized and any movements made are hardly consciously regulated any more. This type of regulation requires very little attention and is characteristic of automated or routine operations or movements (known as psychomotor skills), for example: changing gear when driving a car, operating a keyboard.

According to Action Theory the actions performed constitute the final product of regulatory processes. From a functional point of view these processes require several steps:

(a) Redefining the work instruction (task) to make it an accepted task with a personal goal status.
(b) Orienting to conditions in the environment such as: the availability of information, work aids, or persons and to personal conditions such as: knowledge and skills.
(c) Design of sub-goal sequences and variants of action programs, based on an analysis of the goal, interpretations of directives, etc., and on the available previously designed programs.
(d) Choosing, within the framework, an action program variant that will serve as an action program or having an actualization of an available action program.
(e) Step by step implementation of the action program under continuous monitoring and supervision. Monitoring implies watching the sequence of operations and checking the correspondence of intended and obtained

outcomes (feedback). Supervision implies anticipating future operations, planning new courses of action and making decisions and choices which have a more strategic impact. It ensures that the action plan is adapted, interrupted or terminated whenever necessary while being implemented.

Steps (a) to (d) constitute the preparatory or orientation phase of an action. Step (e) embodies the phase of action-execution. In a serial mode of regulation these phases take place alternatively, while in a parallel mode they occur simultaneously, with a certain degree of overlap. This means that during the execution of certain operations (or actions) other actions are mentally prepared. This far-reaching anticipation occurs especially in activities that demand little attention and allow parallel cognitive processing. This can be regarded as 'expert behaviour'.

Although Hacker does not refer to any kind of monitoring process, it may be useful to assume that such a process is active. The monitoring of the actions may be conceived as cutting across all three levels of regulation. It is a continuous process of gathering feedback on the consequences and progress of the action. When necessary, for instance when an error is detected, the information is directed to a higher regulatory level in order to develop a strategy for error-handling. It might be assumed that monitoring is regulated at the level next highest to the one where the actual operations are executed. Operations on the lowest level, i.e. the 'skill based' level, are executed as Test-Operate-Test-Exit (TOTE) units and proceed almost unconsciously (cf. Miller et al., 1960; Hacker, 1986 p. 140), which means that only the results of the various operations can be monitored. This takes place at the 'rule based' level. Monitoring at the highest level, i.e. the 'knowledge based' level relates closely to the supervision process (Roe, 1988) of all activities. This supervisory process is regulated at the highest level which implies that it is under cognitive control and includes the process of anticipating future actions (starting action preparation), carrying out status assessment and executing progress control (i.e. strategic planning). It is likely that this process also plays a part with respect to the energetical aspects of work behaviour (cf. Chapter 3) i.e.: establishing whether additional effort should be invested or not, etc.

An important axiom in Hacker's Action Theory which is especially relevant to the present study, is that people tend to maximize their action efficiency. This is manifested in conscious efforts to choose the best possible behaviour alternatives, to rationalize the actual action structure during action preparation (by leaving out redundant operations, combining or synchronizing operations, choosing optimal sequences) on the one hand and by learning processes when executing actions on the other hand. Such learning includes developing smooth patterns of movement (routinization), associating certain patterns of stimuli with specified meanings (signal formation, compare

Shiffrin and Schneiders' (1977) 'Consistent Mapping') and forming increasingly compact representations, especially with the help of language (chunking), etc. Learning to recognize errors, anticipate them and correct them while acting, is another important aspect of the process of acquiring skills (Anderson, 1981, 1982; Leplat, 1989).
The three above-mentioned levels of regulation may alternate during one single work process situation, depending on the nature of the operations to be performed and on the action programs that are available. In principle the aim is to let regulation take place at the lowest possible level which is the least effort-demanding and therefore the most efficient option. Thus trying to lower the action regulation level is another way (apart from choosing the optimal behaviour alternatives) of minimizing the psychological costs. This may result from the above-mentioned learning process. Consequently lack of skills or knowledge and unexpected conditions or outcomes will require a higher level of regulation. For example, errors made during senso-motoric regulation will be detected and corrected at the perceptual-conceptual or intellectual level. When unexpected situations occur analyses will be made and solutions will be sought at the intellectual level. If correction efforts fail, or difficulties appear insoluble, help may be sought either from a colleague or by consulting a manual or an expert. Such a decision may influence the cost level, for instance: in terms of the psychological costs and the time and money involved and it may therefore be regarded as a strategic decision. This process is regulated at the intellectual level.
The more experience is gained in executing a certain task and the further the learning processes progress the lower the level of regulation will be. Acquiring skill in a complex task derives from the interaction of various mechanisms. One could say that automation allows the field of control to expand and that inversely this expansion facilitates automation. So, in acquiring (cognitive) skills the worker is able to attain his goals at lower cost. While the inexperienced worker executes his actions predominantly at an intellectual level, the more experienced worker will apply perceptual-conceptual and senso-motoric regulation. The skilled worker may act more efficiently because the lower types of regulation will demand of him less effort which will enable him to anticipate future operations and actions.

Action regulation also depends on a person's actual state. What is essential to Action Theory is the pro-active concept of acting which implies that human beings are constantly looking ahead and actively planning their actions; anticipating things to come. In overload or stress situations when the subject is tired the mental capacity available may no longer be sufficient. Consequently the subject may not have the capacity (see also Chapter 3) to actively anticipate the actions that are to come; the subject's acting in the present thus becomes re-active. Essential parts of the action program may be left-out ('short-circuited'), signals may not be noticed or may not be properly interpreted, checks on completed operations may be skipped, problems may be solved by partial analyses and decision procedures and so on (see also Meijman, 1991). In cases of under-stimulation (cf. Sperandio, 1978) compensatory activity will be sought, such as 'playing' (i.e. consciously seeking intellectual activity) when control tasks are monotonous. This may draw attention away from the actual task. In both cases more errors are likely to be made and consequently action effectiveness will decline.

Action Theory acknowledges that the individual has several psycho-physiological mechanisms (resources) at his disposal for the preparation and execution of actions, each of which have specific capacities and limitations. Apart from sensory and motor mechanisms, attention, thinking, and memory functions are particularly relevant. Mechanisms regulating emotions and motivation should also be mentioned.
Above we have referred implicitly to the attention mechanism that regulates the magnitude of mental capacity and to the assignment of capacity to perceptual, cognitive and motor processes. These aspects will be further dealt with in Chapter 3.

2.4.1 Critical evaluation of Action Theory

What is characteristic of Action Theory is the postulate that human work activity can be regarded as a process by which an individual transforms his environment purposefully while using psychological regulation mechanisms. Acting is seen as a cognitive regulated, goal-directed process based on programs and flexible strategies. The progress of a course of action is continuously monitored by feedback loops (included in TOTE units) which enable an operator to modify his plan of action, if necessary. Specific action programs that are often used, can be routinized and regulated at a lower level of consciousness (thus requiring less attention). There may be several ways to one goal which is why an operator always has to choose an action program from several alternatives.
Another important aspect is the assumption that subjects have to interpret and internalize the instructions given in the task. For this purpose a distinction is made between the 'objective task' (in fact the set of instructions) and the 'subjective task' (the individual's interpretation of the instructions. In German terminology this is known as 'Auftrag' or 'Aufgabe'). The 'subjective task' determines which strategy and action plans the worker thinks will apply to the situation and which will therefore influence his perception and expectations about the level of efficiency he will be able to reach. The strategy and action plans will also be selected with the idea of being as efficient as possible. To some extent this helps to highlight inter-individual differences in work behaviour, when individuals work on identical tasks.

Action Theory aims to offer theoretical notions about human work behaviour which can be applied by engineers who design machines and work places. For that reason Hacker has derived criteria, which can be used for analyzing and designing work situations, tools and the systems (to be used at work). Some examples of these criteria, formulated as questions (and operationalized in the 'Tätigkeitsbewertungssystem' - TBS, Iwanowa and Hacker, 1984), are:

- Does the worker have decision latitude (degrees of freedom)?
- Does the worker have a clear idea about the goal of the task?
- Are the coherence and the inner structure of the work transparent enough for the worker to reconstruct the task sequences himself?
- Is there any chance for the worker to further develop his qualifications (by learning something)?

These criteria reflect the approach of the worker involved in active planning, in goal-directed behaviour and in controlled acting. The idea of various 'goal-action plans' being linked up makes acting flexibly conceivable. This concept of a kind of higher order organization of work behaviour is crucial to Action Theory and is what makes it so valuable: this is what is lacking in most other theories. For this purpose Hacker combined theoretical concepts from Soviet theories about consciousness and activity with Western cognitive psychology to form one heuristic framework. From here he formulated theoretical notions about human work behaviour. What makes Hacker's theory so useful to practical engineers is: the guidelines and criteria he formulated on task design. In fact when Hacker developed his theory he had an open eye to the problems that engineers encounter when developing and building new machines.
An important point of criticism is, that Hacker did not operationalize all his concepts as clearly as certain people, especially psychologists working in the field of cognitive and experimental psychology, would have liked him to or indeed are used to. Nevertheless various authors (cf. Hacker, 1986; Hacker, Volpert and von Cranach, 1982; Ulich, 1984; Volpert, 1975, 1990) have published series of laboratory and applied research projects to demonstrate the usefulness of this theory.
Some writers (cf. Neuberger, 1985) have criticized Hacker's individualistic approach claiming that he concentrated too much on the plans and strategies of individuals thereby neglecting the group dynamic effects on these aspects.
Although Hacker refers to the tendency to optimize the efficiency of actions (Handlungsefficienz) by means of rationalizing the existing action structure he does not explicitly bear in mind that people's strategies and plans may also be based on 'economical' principles. Yet the theory offers some possibilities for elaborating this point.

2.5 The Action Facilitation Approach

Action Theory is becoming more and more popular in West European countries. This can perhaps be attributed to the high applicability of parts of the theory in the technical domain. Action Theory was, in fact, designed for engineers and designers. The theory describes how human actions are psychologically regulated. The theory seems to be useful for engineers who have to design new machines and places of work.
The Action Facilitation Approach is an elaboration of this aspect of Action Theory. This approach, as formulated by Roe (1984, 1988), is intended to be used for (re)designing computer interfaces and was later generalized to 'within job design' (Roe & Meijer, 1990). The assumption frequently made is that the computer is a *tool* employed by a user to carry out certain work activities. With this assumption in mind AFA draws upon the principles of human work activity, as specified above, to find ways to support the user. The general thinking behind the Action Facilitation Approach is, that by supporting the user in planning and carrying out his work activity there will be an improvement in the overall performance of the man-machine system, both in effectiveness and efficiency.
The Action Facilitation Approach consists of 1) a set of principles of human action, derived from Action Theory, that serve as a psychological model of the worker 2) a set of guidelines, based on these principles of human action, that indicate dimensions of support and 3) a set of interface design recommendations that serve to operationalize each of the support dimensions. It is hypothesized that implementing these design guidelines will lead to 'action facilitation'. Action facilitation has to do with supporting the worker in the carrying out of his task and may be operationalized as improvement or maintain of performance at lower individual cost.

The design principles of the Action Facilitation Approach, as described by Roe (1988), should be considered as supplementary to classical ergonomic principles.
From the Action Facilitation perspective what is of utmost importance is that the division of tasks between man and computer is such that the individual user can develop and implement his own action programs. Computer software, work procedures, work aids, etc. should not prescribe tasks and constraints that interfere with action programs that are optimal from the user's point of view. Instead they should be adjusted to the user's working strategies and support his regulatory activity as much as possible.
The main tenets of this approach are summarized below:

- The orientation to the task, task execution, and the development of action programs should be facilitated. This should be done by presenting relevant information about the status of system modes and the progress of the action execution and by offering help in analysis, problem solving and decision making. In general the development of personal action plans should be stimulated by offering relevant information on how one can proceed with the system.

- Feedback should be presented in a way that corresponds with the structure of the action program and the level of regulation, which to a large extent depends on the user's degree of experience.
- The system should be flexible enough to allow for changes in the action program over the course of time as well as in the level of regulation which usually occurs when the user is confronted with unexpected conditions and/or errors. Modifications in the action program also might be called for when the user wishes to reach a higher efficiency level. This might require more condensed formats for representing commands, system messages, etc.
- Apart from being merely flexible aids should be there to help increase the efficiency of performance by supporting the parallel execution of operations (like printing one text while modifying another) or to abolish (when possible) repetitive operations, etc.
- To facilitate supervisory activity, especially anticipating, information should be presented on the parts of the action program that have already been executed, or are waiting to be executed. This information should not interfere with the actual task performance but should rather be imparted simultaneously.
- Endeavours should be made to keep the workload at an optimal level. Through proper layout and carefully choosing what information is displayed it may be possible to avoid overloading. Furthermore the system should allow users to regulate their own workload by giving them the opportunity to choose their own work strategy, e.g. by postponing or cancelling certain tasks or by offering possibilities for the simultaneous execution of more than one simple task in order to prevent under-stimulation arising.
- With respect to the above mentioned principles an effort should be made to accommodate (individual or group) differences between system users in terms of qualifications (knowledge, skills, abilities, and working-styles).

It is believed that applying these design principles will enhance efficiency in work behaviour. Therefore efficiency can be regarded as a core concept within the Action Facilitation Approach.

2.6 Towards a theoretical framework for efficiency in work behaviour

Action Theory offers some opportunities for incorporating our efficiency concepts from section 2.2. Hacker states explicitly that human action is characterized by a tendency to optimize action efficiency. This is manifested in the way the action structure is rationalized during action preparation on the one hand and in the way learning processes take place during action execution on the other hand. In practice this means that the worker is always looking for better or more efficient ways to perform his task. This process is stimulated by feedback gained during action execution. Monitoring and supervising have an important role in this respect. These processes provide relevant information during action execution: feedback on the progress of the actions and the errors that are made. Trying to recover the errors stimulates a learning process. This learning process enables one to find the most efficient way to accomplish one's task.

The process of acquiring new skills is also accompanied by lowering the action regulation level. This means that the action program takes shape and can be executed with less attention than was previously required. Regulating skilled actions requires less attention (and therefore less mental effort, cf. Chapter 3). The amount of mental effort that is needed to execute a task may therefore be indicative of the worker's level of efficiency as long as the person's psycho-physiological state remains unchanged (cf. Hockey, 1986b; Heemstra, 1988; Zijlstra et al, 1989). This relates also to the efficiency of tool handling or operating a system: the less attention is needed for operating a system, the more attention can be devoted to the primary (or actual) task. The complexity of an interface, e.g. the number of function keys, the structure of the menu-dialogue, the transparency of the system, user guidance, etc., will have their influence on efficiency too.

In discussing the earlier mentioned efficiency concepts I continually referred to situations where the goals had already been attained, in other words where the actions had been effective. This clarifies the relation between effectiveness and efficiency, i.e. effectiveness is a necessary but not the only prerequisite for efficiency. When actions are not effective (i.e. goals are not reached), they cannot by definition be efficient.
However, some nuances have to be seen in this statement. In real work situations things might be a little more complex, for instance tasks may vary in length. Sometimes, like for process operators, it is even impossible to determine when the goal has been attained and the task is finished. Furthermore there are situations where a person thinks that he has completed the task while the employer may have a different opinion. Action Theory does take into account the fact that there is a difference between the prescribed task (i.e. output level, working-hours and other constraints) and the subjective interpretation (redefinition) of the various aspects of the task. This implies that there may also be a difference between the subject's and the organization's interpretation of whether a task has been completed or not. This constitutes an additional difficulty in determining effectiveness and efficiency, because it is not always evident whether or not a task is

completed. This also depends on a person's own interpretation of the task and its constraints, and his personal norms and values. As we are interested in psychological efficiency what is decisive with respect to completion of the task is when the individual thinks that he has attained his goal.

Since it is impossible to know each individual's preferences with respect to certain situations in advance we have to satisfy ourselves with more abstract notions of behaviour-economics (cf. Battmann, 1988). Therefore we assume that people strive to spend as little time and expend as little energy as possible on their task and thus work as efficiently as possible. In this respect the amount of effort that is needed for task execution (i.e. to accomplish one's goal) might be a valid criterium.

2.7 Conclusion

In this chapter the concept of efficiency in work behaviour has been elaborated upon.
Furthermore it has been argued that Hacker's Action Theoretical approach (1978, 1986) is suitable for evaluation studies. Hacker's Action Theory offers a holistic framework that explicitly takes efficiency in work behaviour into account. Action Theory also covers the 'non-expert', error-prone behaviour that is to be encountered in new, unfamiliar tasks.
The design guidelines that have been formulated within the Action Facilitation Approach are an elaboration of the model of work behaviour created by Hacker. The central point in this approach is that the worker decides upon his own working method and strategy. It is argued that this is better than prescribing exactly how a task should be executed. Such an approach does not offer enough opportunity for the various working methods and strategies that are needed to improve efficiency. The design guidelines of the Action Facilitation Approach support the idea of the operator controlling the work process. This provides opportunities for changing working methods and strategies and thus for improving efficiency.

With respect to the operationalization of efficiency the individual perception of task accomplishment 'costs' has been stressed. It has lead to the acknowledgement that efficiency is related to the amount of effort that, according to his perception, an individual has to put into a task in order to accomplish the prescribed output.

Chapter

3 Mental effort: theoretical viewpoints

3.1 Introduction

The way in which complex Man-Machine Systems are designed has a great impact on the (mental) load that is imposed on an operator. Working with such systems taxes a worker's information processing capacities.
Hence the reason that system designers need to be able to discriminate between design options and select those that will ease the operator's task. Being able to estimate the mental load that is imposed upon the operator requires having a knowledge of human work behaviour (see previous chapter) and an understanding of the cognitive processes and mental operations that are needed to decide and act. This is the aspect that will be dealt with in this chapter. To have insight into mental workload means first of all having knowledge of the way in which human information processing works.

In this chapter some theoretical viewpoints concerning mental workload and effort will be described. The concept of 'mental effort' will be central in this chapter.
In general we can distinguish between two different perceptions of this concept. In the first approach mental effort is strongly related to *attention* (cf. the 'search-light' metaphor, Cowan, 1988). The more attention demanding a task is, the more effort will be required. This approach stems from research on mental workload and focuses primarily on human limitations to process information. These limitations are determined by the structure of the information processing capacity.
In the second approach effort is conceived as a *process* that regulates the psycho-physiological *state* of the person, and keeps this state in a condition that is optimal for carrying out a task. This approach stems from research into the effects of various stressors on mental functions. Here effort is seen as a compensating process that prevents performance dropping below a certain critical level.
This chapter will present a brief overview of several theories and models related to both these conceptualizations of effort.

3.2 A general model of human information processing

Figure 3.1 presents a schematic representation of the information processing system (based on that of Wickens, 1984). Stimuli entering the 'Short-term sensory store' are passed on to the next stage; the perception stage. At this stage information is encoded, i.e. critical features are extracted, and after the information has been categorized it is passed on to the next stage: the decision and response selection stage. Here the information is interpreted and related to the knowledge available in the memory system and on the basis of the outcome of this process a decision is made on whether and if so, how, to react. Response patterns from previous situations may be called upon and updated when necessary. These responses (or consequences) may constitute new input for the information processing system: symbolized by the feedback loop.

The box in the upper part of Figure 3.1 symbolizes a limited amount of attention. This can be visualized as a 'reservoir' for retaining attention for the several stages of the process. This reservoir reflects the limitations of the human information processing capacity. Old and/or familiar information requires no attention and can therefore be processed very quickly because the relation between stimulus and response has already been established. This means that there is no need for extensive decision and response selection. For example red traffic lights immediately call for the response 'stop'.

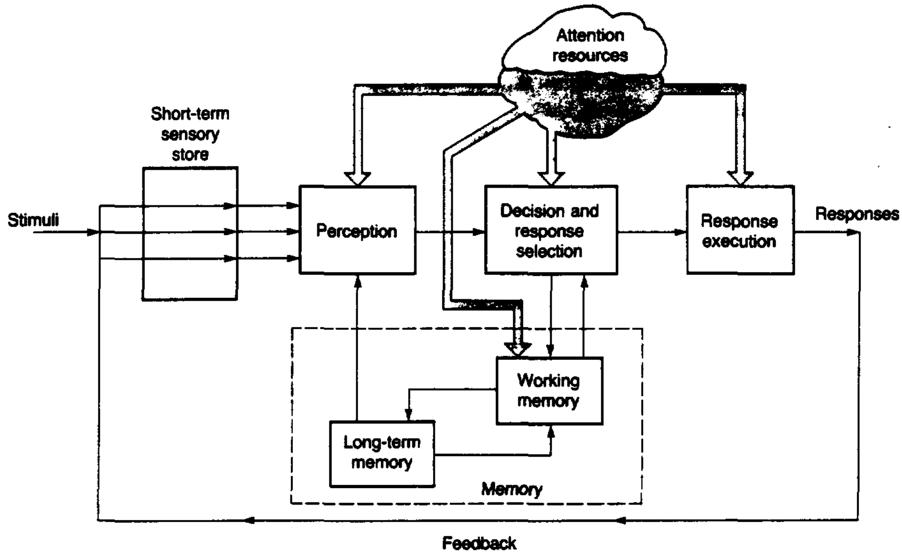

Figure 3.1: *the information processing system*

New situations, or new information have to be dealt with in a different way. Such situations call upon the so-called 'problem solving behaviour' capability (Newell & Simon, 1972) which is very attention demanding.
These conceptualizations bear some resemblance to Hacker's notions of levels of regulating human action (cf. Chapter 2): the perceptual-conceptual (or rule based) level and the intellectual (or knowledge based) level.
The attention available is supposed to be allocated to the various stages of the process according to certain allocation policy (Kahneman, 1973).

There are other models, (see Cowan, 1988; Hockey, 1986a), but all converge on the point where several stages of information processing are assumed. The allocation of attention to these stages is controlled by a so-called '*central processing unit*'. This unit is intended to supervise information processing. Another aspect is that most models assume a limited attention resource, or, a limited 'central executive processor' capacity. This central processor is supposed to focus attention (in some models the metaphor of a search-light is used - Cowan, 1988) and thereby control the processing of information: so-called 'controlled processing'. This system is regarded as being responsible for attention-demanding operations, like selecting information from the memory system, activating certain selected representations in 'working memory', searching for information in 'long-term memory', recombining information and so on.
Another important aspect concerns the memory system. The memory system (long-term memory and working memory) is supposed to contain a lot of information: knowledge, action programs, etc. Modern theory about the storage of information in the memory holds that knowledge is stored in neural networks in the 'long-term memory'. When knots in this network cease to be activated (by means of recall) for some time, they will loose their activation capacity which means that information will be forgotten.
Furthermore it is assumed that some processes are not controlled by the central processing unit; these processes seem to proceed automatically: 'automatic processing'. Stimuli that are continuously available (such as background noise), or enter at predictable moments, lead to habituation of the system. These stimuli do not usually become the centre of attention any more unless they are deliberately brought to attention again by means of conscious processes. New stimuli attract attention involuntarily.

According to these models of information processing human behaviour is assumed to be the result of automatic processes and processes which are controlled by the central executive system (the central processing unit). Or, as has been formulated in Chapter 2: human behaviour involves several regulation levels.
In Hacker's Action Theory (1978) three regulation levels are distinguished. Rasmussen (1983) presented a comparable model, also with three different levels, that was based upon research done amongst operators in the process

industry. It is not quite clear how these models relate to the above mentioned models of information processing. The cognitive models focus primarily on describing and explaining the specific processes that are involved in the processing of information at micro level while Hacker's theory focuses on describing the organization of work behaviour, which may in turn be conceived as operating on a macro level when compared to these cognitive psychological models. So, there is a difference in scope. Although it is rather speculative, it may be assumed that activities that are regulated on the intellectual level (the knowledge level) or the perceptual-conceptual (the rule based level) may include controlled and automatic processing of information as well. Since Hacker's and Rasmussen's theories were formulated other authors have elaborated their models and have suggested variously differentiating into four (Frese, et. al., 1989), five (Volpert, 1982) or even more levels. Frese, for instance, suggested that there might be an additional (higher) level which is concerned with abstract thinking.

3.3 From Mental Workload to Mental Effort

The concept of mental workload refers to the degree to which the information processing system is loaded, or occupied. This concept became popular because in certain situations it seemed that operators were 'overloaded' or 'overcharged' and so started to make errors. These are situations in which the demands of the task exceed the amount of human capacity available (Gopher & Donchin, 1986). What is inherent in this notion is the assumption that the capacity of the information processing system is limited. The nature of these limitations is therefore something which is of interest to designers of man-machine systems.

Originally workload was conceived as being synonymous with the *demands* of the task. Later on a distinction was made between '*objective*' workload and '*functional*' workload. The latter was conceived as the effect of the task demands on the individual and is presumed to depend on the capacity that the operator has available (Meijman & O'Hanlon, 1984; Meijman & Mulder, 1992).
Human work capacity depends on the various physiological systems and mental processes which human beings have at their disposal, or which they can acquire (cf. abilities and skills, etc.). These systems and processes have to be mobilized and organized in order to be able to act. Meijman & O'Hanlon (1984) call this the individual's 'performance *potential*'. There are individual differences with respect to people's performance potential. For instance, while for some people lifting and carrying a 50 kilogram weight poses no problems others can only manage to do that if they put great effort into it. The (task) load may be the same for both but how much energy it will cost will depend on a subject's performance potential.

Moreover the same task can even constitute different workloads for one person at different moments because the actual load is subject to a person's psycho-physiological state (condition) which is in itself variable. Factors such as fatigue, prolonged workload, circadian effects, shiftwork, intake of alcohol or medicines can affect a subject's psycho-physiological state, i.e. one's work capacity. When the subject is not in an optimal condition he may be able to perform at an acceptable level, but the costs, in terms of invested effort, will be higher. This means that it will be harder for him to work when the actual psycho-physiological state is not optimal in relation to the energy that is required to execute his task. Consequently a worker's efficiency also depends on his psycho-physiological state.

Workload can easily be viewed a static concept. However, workload is determined by the interaction between the characteristics of a task and an individual worker's capacity. The individual capacity appears to be very dynamic, it may change from hour to hour or even from minute to minute when motivational factors are involved.
Hence the introduction of the concept '(mental) effort'. Effort refers to the degree to which an individual mobilizes his performance potential. A task imposes a certain workload on the individual, or one could say, charges the performance potential but it does not automatically 'trigger' the expenditure of effort. The expenditure of effort is preceded by a decision to invest effort in order to execute a certain task. This means that the motivational component also counts. A person could also decide to withdraw from the task (for certain reasons) and thus not invest the required effort. If a person accepts a task he has to exert effort in order to handle the accompanying task demands. The amount of effort that has to be invested depends on the complexity of these task demands and on that individual's actual psycho-physiological state.

3.4 Theories on mental effort

In the relevant literature two approaches to this concept can be distinguished (see also Mulder, 1986).

3.4.1 Mental effort as 'executive resource control'

One approach stems from a research tradition which was introduced by Donders (1868) and was later on revived in the work of Sternberg (1969). It focuses on the structure of the process of information processing. Donders started to build models of this process. He assumed that performing various tasks requires a different sequence of cognitive operations, of so-called 'modules' such as: 'encoding', 'searching in memory', 'decision making', and 'response preparation'. The general assumption is that after information has been processed within one of these modules it is presented to the next module as a discrete package of information. Thus, according to Donders information processing is a serial and discrete process. Donders developed an elegant method for measuring the processing time of each of the various modules. He created three small tasks. One was a very simple task in which a subject only had to press a button when a light came on. According to Donders a subject only had to observe and react; he called this the A-task. The second task was more or less the same except that two lights were used. The subject had to press the button on the right side when the right light burned and the button on the left side when the left light went on. This task was called the B-task and contained the modules: 'observe' and 'react' and also 'discriminate' and 'choose'. The third task, the C-task, was to press only the right button when the right light burned. This task contained the modules 'observe' and 'discriminate'. By measuring the reaction times of each task and subtracting these from each other Donders was able to calculate the duration of each module. Therefore Donders' method came to be known as the 'subtraction method'. A hundred years later Sternberg (1969) used a modification of Donders' method: the 'additive factor method', which was based on a similar kind of logic.

Since then other models have been presented. What is central to most of these models is that they all require several stages of information processing. Cognitive psychology research focuses on the 'architecture' of the information processing system and constitutes an endeavour to identify and explain the limits, shortcomings and bottleneck points of these models. One of the main assumptions with these models is that information is always processed in one and the same way. Incoming information always goes through the various stages in the same sequence and can never be passed on to a next stage unless it has completely passed through the relevant preceding phase. New information can only be dealt with after previous information has been processed. This makes these models rather inflexible, which is one of the major criticisms. It would be (theoretically) impossible to perform a task which addresses more than one stage at the same time, at least this would have been impossible with the earlier 'single-channel' models (cf. Broadbent, 1958). Later models, for instance the 'cascade-models' (cf. McClelland, 1979) do have a theoretical opening for this possibility.

Another problem in these early models is that they do not account for differences in strategies (for instance: speed/accuracy trade-off). This, together with the fact that these information processing models have a rather complex structure, has led to the search for alternatives.

Using the 'resource-' or 'energy-metaphor' (analogous to physical workload models) provided one solution. Stages of information processing are supplied with resources (cf. Kahneman, 1973). The concept of a resource

relates closely to that of 'difficulty'. Tasks that are more difficult and are carried out at the same performance level demand more resources. Resources can be seen as sources of energy which are at the disposal of the 'central processor' of the process. In cases of shortage of energy there will be no processing of information. According to Wickens (1984, 1986) the concept of processing resources is put forward as a hypothetical construct to account for variations in the performance of task; in situations of time-sharing.
Kahneman (1973) introduced a general undifferentiated source of energy and related this resource-concept to 'arousal'. Kahneman assumed that the total capacity depended directly on an individual's psycho-physiological state: their so-called 'arousal-level'. How capacity has been allocated becomes apparent from changes in physiological activity, such as pupil-dilatation.

Later on 'multiple-resource models' were proposed (Kantowitz and Knight, 1976; Gopher and Navon, 1980). In the multiple-resource approach it is argued that instead of having one central 'pool' of resources with satellite structures (eyes, ears, hands, voice), humans possess several different capacities with resource properties. This allows various processes to be at work simultaneously as long as these processes do not draw on the same resources. If these processes do draw on the same resources parallel tasks will interfere with each other.
Navon & Gopher (1979) came up with an economical model in which a person who has to perform one or more tasks has to try to optimize various resources. The great advantage of these 'resource-models' was that parallel processing, or time-sharing, could be explained. It is assumed that tasks demand resources when they are executed and the availability of resources is limited. Most tasks call upon various modules. Some tasks can be time-shared and other cannot (depending on their nature) because they draw upon the same modules and will therefore interfere with each other. For example, an experienced driver can talk with a passenger and give his attention to the task (of driving and attending to the traffic) at the same time as long as the traffic situation does not change drastically.
Another problem with the 'single resource theory' is that with changes in the structure of the task the interference between tasks can grow, lessen, or even disappear completely, depending on which modules are involved. According to 'single-resource models' the effectiveness of time-sharing decreases when the demand for resources - in both tasks - exceeds the resources available. The assumption is that the more complex or difficult a task is the more resources are needed. Therefore it is obvious that it is very difficult to perform two complex tasks in time-sharing. Time-sharing occurs when tasks are well learned and routinized as is the case with actions which are regulated at the lower levels (see Chapter 2).

At the moment the 'multiple resource' model of Wickens (1984) is considered to be the most adequate model. Wickens assumes that there are three simple dichotomous dimensions in a hierarchical structure of resources (see Figure 3.2): two stage-defined resources (early versus late processes), two modality-defined resources (auditory versus visual encoding) and two resources defined by processing codes (spatial versus verbal).

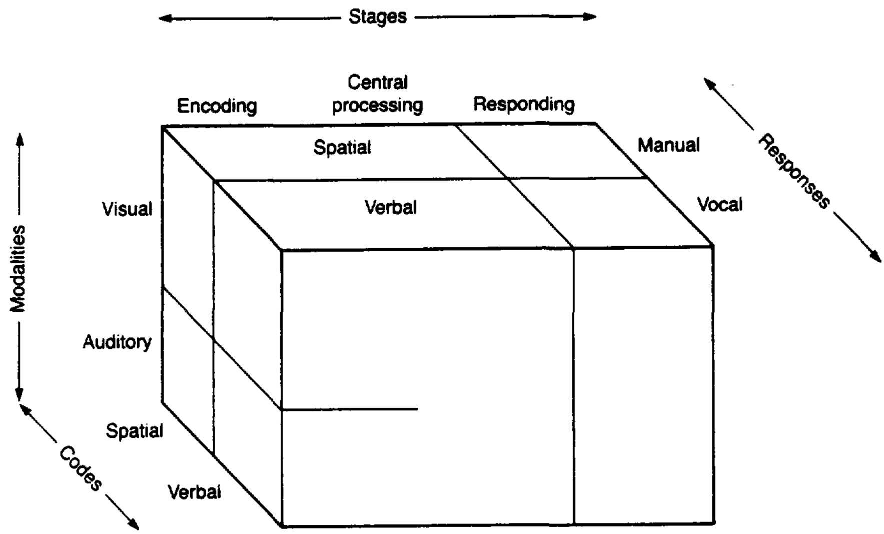

Figure 3.2: *Wickens' model of the structure of processing resources*

To the extent that any two tasks demand separate rather than common resources from any of the three dimensions three things will happen: 1) time-sharing will be more efficient; 2) changes in the complexity of one of the tasks will be less likely to influence performance in one of the other tasks; 3) it will not be possible to reallocate resources between the tasks, that is, to utilize the resources of one task for another (when different resources are depended upon).

The three dimensions of Wickens' model do not claim to account for all structural influences on dual-task performance and time-sharing efficiency. A literal interpretation of the model suggests that any two tasks demanding separate resources should yield perfect time-sharing. Since that does not quite appear to be the case, Wickens suggests that "there may well be a layer of 'general capacity' that is added, like frosting on a cake, to the top and front of the separate resources. These resources, competed for by all tasks, would then prevent perfect time-sharing of all but heavily data-limited tasks" (Wickens, 1984: p. 305).

The 'multiple resource theory' has been supported by neuro-physiological findings. In 1975 Pribram & McGuinness published their well-known paper in which they presented evidence for three neuro-physiological mechanisms. The first mechanism (arousal) determines activity in the early, input-related modules. The second mechanism (activation) determines activity in the later output-related modules and finally there seems to be a third mechanism (effort) operating on a higher level which coordinates and regulates both other mechanisms.

Another important development, apart from the innovation of the 'multiple resource', is to be found in the work of McClelland (1979), Norman & Bobrow (1975) and Posner & Snyder (1975). They performed experiments and concluded that a distinction can be made between automatic, unconscious and parallel processing of information on the one hand and controlled, conscious and serial processing of information on the other hand. This last way of processing is supposed to be effortful, or attention-demanding, while the first is not.

Shiffrin & Schneider (1977) and Hasher & Zacks (1979) have all investigated, in a laboratory situation, the conditions under which information is automatically processed. It appeared to be very important for a one-to-one relationship to be established between stimulus and response during the preceding period. As a consequence it becomes unnecessary to search exhaustively in the working memory, which saves time and effort. Schneider & Shiffrin (1977) call this 'consistent mapping' between stimulus and response.

When stimuli continually change thus requiring different responses it is not possible to establish a clear and obvious relation between stimulus and response. This situation is called a 'varied mapping' situation.

According to Logan (1988-a, 1988-b) automatic processing imposes a minimum of demand upon the central executive system because the information required can be directly obtained from memory, in contrast to non-automatic processes which sometimes have to use very complicated algorithms.

So, to summarize, it is clear that the primary focus has been on the structure and architecture of the information processing system. To explain parallel activity in the processing stages the 'energy metaphor', together with various resources, was introduced. According to the theories and models presented in this section, effort should be concieved as 'attention-demanding' processing. Tasks are attention-demanding (and thus require effort) when they call upon the various processing modules. This should be apparent from the increase in time that is required to process the information (reaction time).

The energy-metaphor is useful for describing fluctuations in attention (Kahneman, 1973). Attention is likely to flag when 'energy' in the various 'resource' areas has been consumed. Kahneman (1973) suggested that the amount of energy available relates to the individual's psycho-physiological state.

Other researchers stress the importance of the dynamic aspects of learning and establishing a relation between stimulus and required response. This means that tasks become less attention-demanding (i.e. require less effort) when subjects do not need to carry out extensive searching in the memory-system.

3.4.2 Mental effort as 'state-regulation'

The second approach to mental effort relates this concept to the individual's psycho-physiological state: a state which changes with people's circadian rhythm the most obvious example of this being the difference between sleeping and being awake. When under the influence of psycho-pharmaca, alcohol, or lack of sleep a subject's state will change as well. Each state can be characterized by a different pattern of (physical and cognitive) activity. A period of prolonged working may also influence the psycho-physiological state: the resultant state being called 'fatigue' (cf. Meijman, 1991). Changes in psycho-physiological states may also affect the functioning of information processing modules. A performance decrement in these modules can be avoided if more effort is exerted. Effort is seen as something which influences the mechanisms which determine the state of an organism. In this connection Broadbent (1971) distinguished two 'arousal' mechanisms: a 'lower mechanism' which is activated for instance by noise and which is de-activated by lack of sleep; and a 'higher arousal mechanism' which is active when tasks are of relatively short duration but which displays decreased activity when tasks are lengthy. This latter mechanism is supposed to be activated by knowledge of results, rewards, etc. According to Broadbent this mechanism compensates for sub- or supra-optimal activity of the lower mechanism. Consequently conditions like lack of sleep or noisy working conditions will not immediately affect people's performance levels. Especially where tasks are of a shorter duration the higher mechanism will be able to keep performance at a constant level by providing compensatory activity. With prolonged work and without knowledge of the results this would be almost impossible.

In the models of the previous paragraph effort could be interpreted as being synonymous with attention; the models in this paragraph offer a somewhat different perspective and introduce the notion that effort is required when a person is involved in actively controlling deficiencies in his actual state (see below).

Under certain conditions a subject is capable of compensating for a sub-optimal state but compensating has its costs: the human cost of adapting to task-demands. This idea has been elaborated by O'Hanlon (1981), Sanders, (1983), Hockey & Hamilton (1983), Hockey (1984).

According to Hockey and Hamilton (1983) several 'states' of the human information processing system can be defined. Such cognitive states are characterized by the availability of different modules in the information processing system and can be influenced by external stressors. For example, noise can cause increased selectivity in the area of attention but the availability of working memory may thereby deteriorate. Alcohol will for instance have quite different effects on the information processing system. Sometimes the actual state does not correspond with the state that is required to adequately perform a given task. It is assumed that a task is performed optimally when the required modules are all available. Hockey's model of 'state control' (1986) is presented in Figure 3.3.

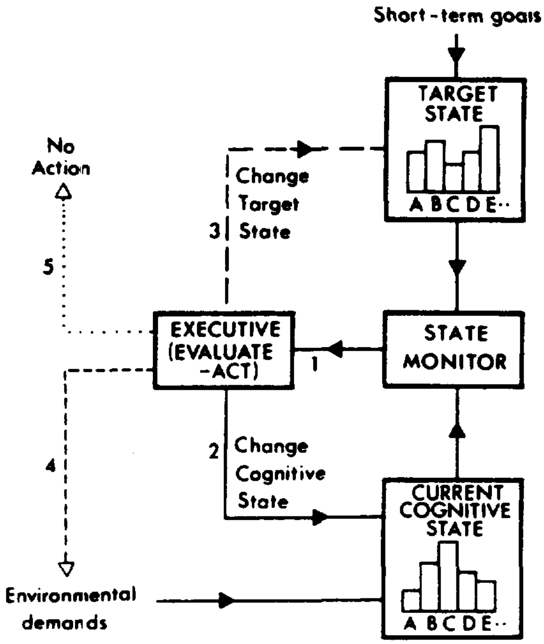

Figure 3.3: *Hockey's control model of state regulation*

What is central to this model is a mechanism which continuously monitors states and registers possible mismatches between the actual cognitive state at that moment and the state which is necessary for optimal task execution. It is assumed that each task requires a particular pattern of activity. This required state is referred to as the 'target' state. Changes in the cognitive state brought about by stressors or internal factors might produce a mismatch between the actual state and the target state. The monitor mechanism reports to the central executive system and so it is that the state is evaluated. In cases where the differences are great performance will deteriorate: mistakes will be made, signals will not be observed, etc. According to this model there are four different possibilities open to an individual in the event that the actual state and the target state are mismatched.

1. The executive system can change the parameters of the current state so that the current state is adjusted.
2. The individual may reduce the mismatch by changing the target state (for instance reducing the aspiration level)
3. The individual may try to control the environmental influences on the cognitive state (this is only an option with physical stressors, such as noise, heat, etc.).
4. Theoretically there is also a strategy of 'inaction' (doing nothing). This might be an option when the mismatch between actual state and required state is very small, or extremely large. The latter can be interpreted as resigning from the task.

Hockey and Hamilton (1983) assume that the central executive system directly manipulates the subject's actual state by intervening in the arousal level. It is assumed that the need to carry out goal-directed activity under unusual circumstances is more likely to involve the effortful, active mode of control since such activity is, by definition, unfamiliar and little practised. The central energetical construct in the use of active control of resources is that of effort.

Effort is involved whenever an attempt to resolve a mismatch of target and current state takes the form of active manipulation of cognitive resources. Effort, in this model, is the psychological process which corresponds to the use of active control of the cognitive processes involved in information processing (option 1 in the above-mentioned model).

To conclude, it can be stated that in this line of research effort investment is regarded as being synonymous with actively changing one's own (cognitive) state. Everybody can imagine how one feels after dinner: a little tired, a little lethargic. When one still has work to do one tries to overcome this sluggishness by actively changing one's state. It is as if extra energy is needed to get started.

It may be clear that in this process a motivational component is involved as well though this is not explicitly stated in the model. The individual has to make a conscious decision to continue working even though he does not feel fit. This includes deciding to exert some effort to overcome his lethargy.

3.5 Towards an integration of both approaches

Sanders (1983) stresses in his 'cognitive-energetical stage model' the relation between the 'energetical aspects' and the 'structural aspects' of the process. This model is based on Sternberg's 'additive factor model' (1969) which describes the current of information through four successive information processing stages: 'preprocessing' of the stimulus, 'feature extraction' for characterization and recognition of the stimulus, 'response choice' the stage when the individual decides whether to react and if so, how to react, 'motor adjustment' the stage where the motorial component of the response is organized. These stages, or modules, are supposed to be related to the three 'energetical mechanisms' (resources), already mentioned by Pribram & McGuinness (1975), 'arousal' which determines the availability of feature extraction while 'activation' controls the availability of motor adjustment and thirdly 'effort' which has two functions. The primary function of the 'effort mechanism' is to control the response choice and the secondary function is to coordinate and regulate both the arousal and the activation mechanisms. Finally Sanders (1983) introduces an evaluation-mechanism. This mechanism evaluates task performance and is influenced by motivational factors such as incentives and knowledge of results.

Sanders' model may be regarded as an attempt to combine the two above mentioned approaches of the effort-concept. In his model one can recognize the two major lines of development of the effort concept: the various stages of the central executive system and the energetical aspects of arousal and activation. A recent model, presented by Mulder (1986), integrates more explicitly both conceptualizations of effort (see Fig. 3.4). This model assumes that 'computational' modules (another term for the various stages of the information processing system) and energetical mechanisms exist. It has a distinct resemblance to physiological processes. The timing of the computational modules is to some extent visible in the latency of components of brain-potentials (Wijers, 1989). It is assumed that the executive system becomes active when there is a need for central processing of information. This executive system also regulates the system's state. It is believed that the frontal cortex is heavily involved in this regulatory activity.

The 'state' concept can be regarded as the binding and central element in both approaches to mental effort. In fact one may also regard the change from controlled (and therefore attention demanding) information processing to automatic processing as a change in the internal 'state' of the information processing system. In the case of automatic processing the information processing system may be regarded as being in a different 'state' from the controlled processing situation. With this viewpoint the 'state' concept is integrated in both conceptualizations of effort.

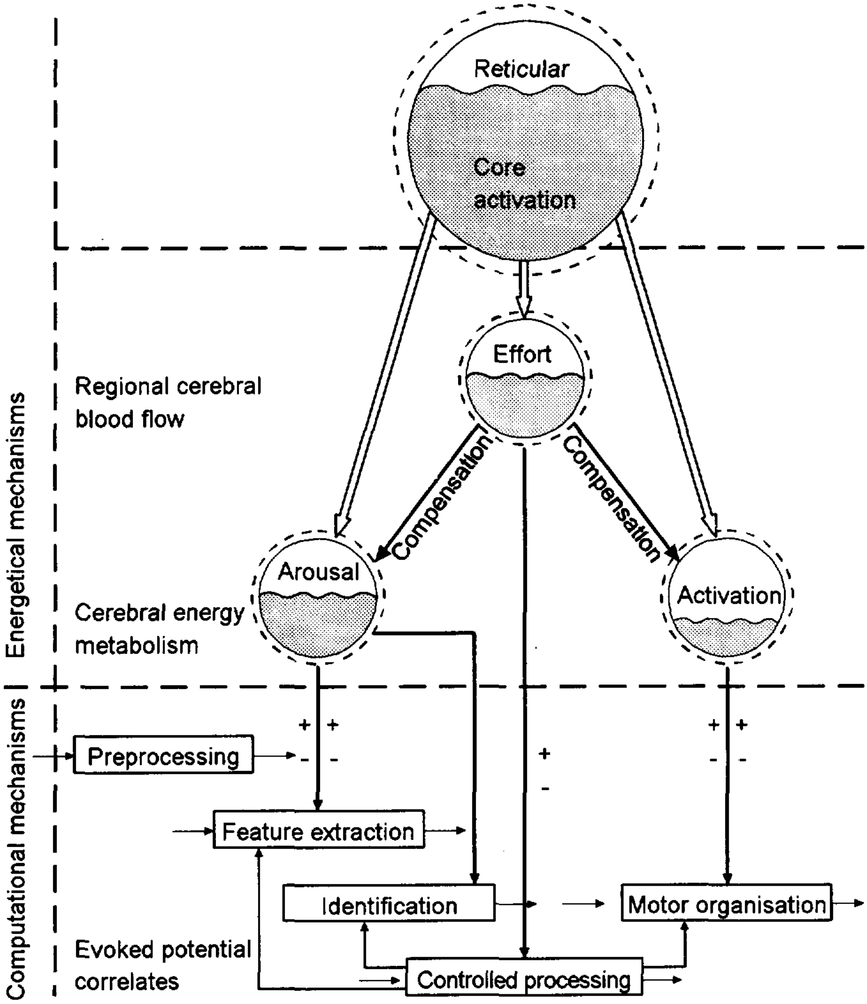

Figure 3.4: *The integrated model of mental effort (Mulder, 1986)*

Another binding element in both approaches is the energetic metaphor. This leads us to the conclusion that effort should preferably be thought of as 'energy', rather than as being a 'mechanism' (the Sanders' model - 1983). Effort can be conceived as the 'fuel' that makes the processes and mechanisms run. Where physical tasks are concerned we could compare effort with such things as oxygen and glucose intake. Though we know that the brain needs oxygen and glucose we still do not know exactly how this relates to the differential activation of the various cognitive systems in the brain.

3.6 Effort as a work psychological concept

The previous paragraphs lead to the question: how can the effort concept can be fitted into work psychological theories?
Whenever a person performs a task, even the most simple of tasks, he has to invest a degree of effort in order to successfully execute this task and the amount of effort required partly tells us what 'costs' are involved in executing the task. It is important to stress that there is a motivational aspect involved as well. The individual has to be *willing* to exert effort. This means that effort is under cognitive control and can in other words be seen as part of the *supervisor* mechanism already referred to in Chapter 2 (section 2.4). The supervisor mechanism is responsible for the ability to anticipate future actions and control the continuation and progress of the actions etc. In other words, this supervisor mechanism checks whether the set goal comes within reach or whether more activity is required. It is therefore plausible to believe that the supervisor mechanism also has control over the effort expenditure level. This means that people are able to select the work strategy that is most economical in terms of effort expenditure in order to equally distribute effort expenditure over the day. There is empirical evidence to prove that people do indeed use such strategies to ensure that their work capacity is evenly distributed throughout the day (Teiger, 1978; van Aalst, et al., 1986).

3.7 Conclusion

An important conclusion of this chapter is that there is a conceptual difference between 'workload' and 'effort'. This distinction is relevant whenever the aim is to estimate the 'costs' for a worker created by task demands. The amount of effort that has to be exerted depends on the complexity of the task and on the psycho-physiological state of the individual. If there is a difference between the individual's actual state and the state that is required to execute a certain task then extra effort is required on the part of the individual if the actual state is to be adjusted. Also in the case of automatic (non attention demanding) processing of information it appears that changes in the psycho-physiological state of the individual are very relevant to the level of exerted effort, as is the case with controlled (attention demanding) processing.

Designers should take into account the fact that effort is a dynamic concept. It is not only the complexity of the system that is relevant but also, in this respect, the worker's actual state. This state may vary during the course of the day.

Chapter

4 Measuring mental effort: the construction of a scale

4.1 Introduction

Several kinds of methods can be distinguished that are used in research on mental effort. In this chapter we will be focusing on the subjective methods because these appear to have some pragmatic advantages such as, ease of application, cheapness etc. Special attention is given to rating scales. First some general remarks will be made about subjective methods and then a few specific instruments will be discussed. Since these instruments appear to be focused on measuring workload instead of effort the third step will be to construct the Rating Scale Mental Effort (RSME) which will be described in section 4.4. The last part of this chapter is devoted to a study in which the RSME has been applied. This study presents some initial information about the validity and usability of the RSME.

4.2 Indicators of Mental Effort

In current research on mental effort several indicators are being used. These indicators can be roughly classified in three categories (for a more extensive review the reader is referred to O'Donnell & Eggemeier (1986)):

A. Physiological indicators
   - those that are associated with information processing: pupil dilatation (Kahneman, 1973), heart rate variability (Mulder, 1980), event related brain potentials (Donchin, 1981);
   - those that are associated with changes in physiological state: muscle tension (van Boxtel, 1993), adrenaline level (Frankenhauser, 1975).

B. Subjective indicators
   - rating scales that are used to rate the task load: Subjective Workload Assessment Technique - SWAT (Reid & Nygren, 1988), NASA-TLX (Task Load indeX - Hart & Staveland, 1988);
   - questionnaires that are used to estimate the operator's psycho-physiological state: mood questionnaires (Thayer, 1967), SEB (questionnaire on subjective fatigue - Meijman, 1991).

C. Performance measures on 'primary tasks' and 'secondary tasks'
   Performance measures on 'secondary tasks' are used to 'demonstrate' the availability of certain 'resources' or 'processing modules'. With respect to performance on primary tasks, or rather 'main' tasks, decline in performance is believed to be an indicator of workload.

It is clear that there is a large range of methods and instruments available for measuring workload. Each of these methods has its advantages and disadvantages in relation to the relevant psychometric criteria and pragmatic aspects. It is beyond the scope of this study to discuss these (dis)advantages in detail. For an extensive review the reader is referred to O'Donnell and Eggemeier (1986). However some general but relevant remarks can be made. Most physiological measurement procedures require specific knowledge and expensive equipment. This makes them less suitable for our purposes. We need an instrument that will measure the costs of work behaviour and can be applied by designers. Most designers are not experts on psychophysiological measurement procedures and do not have the necessary knowledge and equipment at their disposal. Subjective measurement procedures are usually easy to apply and are very cheap, this makes them from more suitable for our purposes from a pragmatic point of view.

4.3 Subjective methods in research on mental effort

Subjective methods have been rather controversial and suspect for some time. Especially in the sixties and seventies people disapproved of the more subjective methods. With the term 'subjective' we mean methods and instruments that ask for the subject's opinion, experience or evaluation. In fact one should make a distinction between the 'subjectivity' of the *method* and the 'subjectivity' of the *phenomenon*. For instance, a questionnaire is always a subjective instrument because it requires a subject's answers to certain questions but there is a difference between the type of questions that relate to 'subjective' topics like satisfaction, motivation, political beliefs, etc. and topics that are of a less phenomenological nature, like intelligence. In the context of this research it is important to notice that with regard to workload one should also distinguish between the 'objectively' assessable workload - in terms of amount of information that has to be processed - and the 'subjective' representation of the externally imposed load.
Many researchers questioned whether instruments that are based on subjects' expressions can provide the required information on mental workload and effort. They have argued that mental workload and mental effort refer to the appeal that is made to the information processing processes and associated resources. These processes are thought to proceed unconsciously which therefore makes them in principle unaccessible for introspection or for conscious evaluation.

During the last decade the attitude towards subjective methods has become more positive. Some researchers maintain that subjective methods cannot be omitted from the research into mental workload, as Johansson et al. stated (Johansson et al. 1979, page 105): "If the person *feels* loaded and effortful, he *is* loaded and effortful. Whatever the behavioural and performance measures may show". Although this puts the primacy of subjective methods very radically I think that in research practice the subject's ratings may indeed come closest to tapping the essence of mental workload. Such ratings constitute the only source of information about the subjective impact of a task.
The importance of subjective experiences goes beyond the issue of subjective ratings. The phenomenological experiences of human operators affect subsequent behaviour and thus also their performance and physiological responses in a given situation. If operators consider that the workload of a task is excessive, they may behave as though they are overloaded, even though in all objectivity the task demands are low. They may adopt strategies that would be appropriate in a high workload situation (e.g. shedding tasks, responding less accurately), experience psychological or physiological distress or adopt a lower criterion of acceptability for performance. It may also affect the willingness of the operator to spend time on the task and the confidence he has in the outcome of his decisions.

From a methodological point of view there is also some evidence in favour of subjective methods. Several comparative studies on different ways of measuring mental workload have been done (Williges & Wierwille, 1979; Casali & Wierwille, 1983; Gopher & Braune, 1984). The general conclusion of these studies is that different rating scales differentiated best between workload conditions compared to other methods like physiological methods, performance measures (such as secondary tasks, methods etc). Moreover, it appeared that rating scales were least obtrusive when taken during task execution.

Nevertheless what remains an important question is, whether methods which are grounded on human experiences of workload and effort can indeed provide the information they pretend to provide. In other words the construct validity of these instruments is of special importance which means that it is important to know whether the construct of mental effort is described in terms of computational stages and their associated resources or in terms of the 'state control theory'.
In this connection it is worth noting that Derrick (1981, 1988) showed, with the use of a one-dimensional scale, that subjective evaluations of the load of tasks are based on: 1) the competition between several resources; 2) the adequacy of the feedback and 3) the appeal made to the entire processing capacity. None of these three components refers to 'state'-like constructs.

4.4 Rating scales

Nowadays the rating scale is a very popular instrument in research on mental workload. Such instruments ask the operator to rate the complexity of the task or the amount of effort that needs to be exerted, while the task is being executed. A well-known instrument, already referred to in Chapter 1, is the 'Cooper-Harper Scale' used in the aviation industry. This scale was originally developed to evaluate the 'flyability' of aeroplanes. Experienced test pilots had to rate the flying characteristics of planes on a 10 point scale (Cooper & Harper, 1969). This scale assumes one underlying dimension which ranges from 'easy to fly' to ' very difficult to fly'. This scale is very specific to a pilot's task and therefore not very suitable for other purposes. Another well-known rating scale was the one constructed by Borg (1962, 1976). This scale contains 15 anchor points varying from 'very, very light' to 'very, very heavy' and was constructed according to the psychophysical 'ratio-estimation' method. This scale is used for rating the exertion of effort in physical tasks as well as in mental tasks (Borg et al., 1971).

An often heard objection to these rating scales is, that they assume one-dimensionality. On theoretical grounds it is hard to imagine that mental load could ever be represented one dimensionally. For this reason several multi-dimensional rating scales have been developed, one of which is 'Sheridan's dimensional scale' (Sheridan, 1980; Sheridan & Simpson, 1979). Sheridan assumes that the subjective experience of workload incorporates three *dimensions: 1) the information processing load, 2) the complexity of the task* and 3) the emotional stress of the task. People have demonstrated that they are very capable of evaluating a task according to these three dimensions and the evaluations of the different raters corresponded very well.
Another instrument that is comparable to Sheridan's analysis scales is, the 'Subjective Workload Assessment Technique' (SWAT), developed by Reid and Nygren (1988), Reid et al. that distinguished three dimensions: 1) Time pressure, 2) Mental effort, and 3) Psychological stress. They came up with these dimensions after extensively reviewing the relevant literature. Reid and Nygren noted what many scientist believed to be critical components of the perception of mental workload. The 'time pressure' factor has been operationally defined as 'time available' and 'task overlap'. This latter aspect refers to the possibility of time-sharing (or parallel processing). The mental effort dimension involves processes such as performing calculations, making decisions, attending to information sources, placing information in short-term memory, retrieving information from long-term memory, etc. In other words this dimension encompasses the concept of mental capacities and refers to the consumption of resources. The stress dimension includes a number of operator variables such as motivation, training, fatigue, health, and emotional state. This dimension is defined as "anything that contributes to an operator's confusion, frustration and anxiety".

Although it has not been explicitly mentioned there seems to be some relation to the two aspects of the effort construct differentiated in the previous chapter. The second dimension is clearly related to 'effort as the central executive processor', while the third dimension shows some relationship to effort in the 'state control' approach. The first dimension is of a different nature, though time pressure does seem to be more related to an individual's state than to the central executive processor.

Reid and Nygren (1988) reported on a few experimental studies that aimed at validating the SWAT. Most of these studies were experiments carried out in a military setting, e.g. during a fighter air defense mission flown in an F-16 simulator.
Low workload was defined as: 'an F-16 chases three enemy aircraft making an "S" weave escape. High workload was defined as: 'seven enemy aircraft approach the F-16; two of the aircraft split in opposite directions to catch the F-16 in a pincher maneuver'. It may be obvious that since descriptions of such situations have a strong emotional connotation (people are fighting for their lives) they are not easily transferable to everyday situations.

A more extensive instrument is the NASA-TLX (Task Load Index) described by Hart & Staveland (1988). With this instrument six dimensions are distinguished: 1) mental task demands, 2) physical task demands, 3) time pressure, 4) success of the performance, 5) effort and 6) frustration level. These dimensions are selected from an original set of nine rating scales (task difficulty, time pressure, performance, mental effort, physical effort, frustration level, stress level, fatigue, and activity type). By combining 'task-related' scales (task difficulty, time pressure, and activity type) and doing the same for the 'behaviour-related' scales (physical effort, mental effort, and performance) and 'subject-related' scales (frustration, stress, and fatigue) they end up with six dimensions. In order to select these six dimensions Hart et al. used criteria such as: sensitivity to differences between tasks, associations with subjective ratings of overall workload and subjective importance of raters.
Subjects have to rate each task according to these six dimensions and evaluate the contribution of each factor (its weight) to the workload of the task. These weight-ratings have to account for between-subject variation, caused by differences in the way in which the various raters define workload. In addition, the weights themselves should provide diagnostic information about the nature of the workload imposed by the task. When their experiments on scales were completed Hart et al. reported considerable disagreement between subjects about which factors best represented their concept of workload though some consistent trends could be observed. According to Hart et al. the weight-rating results suggest that there may be two patterns of workload definition: one based on task and performance related factors and another based on the impact of tasks on the performer.

In their validation experiments Hart & Staveland examined a substantial number of laboratory tasks, such as tasks with single cognitive activities, continuous single-axis manual control tasks, dual tasks with cognitive and manual control activities, simulations conducted in a motion-base, single-pilot, simulator. It appeared that the factors 'task difficulty', 'stress', and 'mental effort' were consistently related to subjective workload for all subjects and tasks. Other factors (time pressure, fatigue, physical effort) were closely related to workload under certain experimental conditions (or one could say: for certain tasks), but not in other conditions. This may lead to the conclusion that the factors 'task difficulty', 'stress' and 'mental effort' are the most important factors in the TLX.

Of the remaining factors 'time pressure' appeared to be the most important one in relation to 'workload'. During the validation studies it appeared that various sub-scales correlated highly with each other (see Hart & Staveland, 1988, pg. 156, table 4). According to this table the dimensions 'Task Demands' and 'Mental Effort' correlate .74. From stress research it has also become clear that the factor 'time pressure' is closely related to the factor 'stress' and might be regarded as one of the determinants for stress reactions.

As has been explained there is a conceptual difference between measuring 'workload' and 'effort' (see also Chapter 3). Workload is the (static) external load which is imposed upon the human operator and consequently relates to the 'objective' complexity of the task demands. Effort-investment however relates to the active involvement of an operator's performance potential. Effort-measurement should therefore account for changes in the state of the operator i.e. prolonged activity, loss of sleep or any other stressor that affects the human operator's psycho-physiological state and reduces his work capacity.
This aspect links up with the two different patterns of workload-definition that are described by Hart et al. The 'task and performance related factors' can be regarded as representing the 'workload' conceptualization, while the 'impact factor' comes close to the 'effort' conceptualization. This latter factor has to do with the costs that are associated with performing the task. Hart & Staveland called this the *experience* of workload.

Several points of criticism may be levelled against the TLX. First of all it may be noted that Hart et al. claim that the NASA-TLX is a multi-dimensional rating scale. However, the scale consists of six separate one-dimensional scales. Each of these scales aims at measuring a construct and some of these constructs may be regarded as 'multi-dimensional' as well. For example, the factor 'stress' has many dimensions and there are a lot of factors that contribute to inducing 'stress', one of which is the dimension 'time pressure'. Since several dimensions appear to be highly correlated

(Hart & Staveland, 1988: p. 156-162), a factor analysis might bring down the present number of dimensions from six to less.
On the theoretical level the question as to whether these six TLX dimensions are exhaustive is still open. Moreover, not all six dimensions seem to be valid for all tasks, as became apparent in Hart et al.'s validation studies. This means that there may be tasks, or situations, in which other factors may have a big influence on (the *experience* of) workload.
In addition Vidulich and Tsang (1986) claim that '..dimensionality is the key-issue in subjective workload assessment..', but in their opinion it is questionable whether subjects are able to distinguish and combine the relevant dimensions adequately. Therefore they suggest that a multi-dimensional technique is not required for a subjective measurement of workload in a homogeneous set of circumstances.

In my view it may be argued that (generally) in research into mental effort it is sufficient to use a well-constructed one-dimensional rating scale in combination with an extensive psychological analysis of the task that is to be evaluated (cf. Alluisi, 1967). Generally speaking it is very difficult to evaluate the workload of a task without considering the context of the task. A psychological analysis of the task should provide the researcher with a clear insight into the, for the operator, relevant and potentially demanding aspects of the task: especially when the task does not belong to a homogeneous task domain. In this respect it may be noted that Hart & Staveland (1988, p. 161) conclude "..... that a phenomenon exists that can be generally termed 'workload', *but its specific causes may differ from one task to the other"* (italics mine).
After the task has been analyzed the operator should give his ratings on the different aspects of the task according to a one-dimensional scale. Such a procedure offers a more differentiated representation of the load profile of the task (see also Meijman et al., 1986; Zijlstra & Meijman, 1988).
In fact the study of mental workload should always be preceded by a detailed analysis of the task which is to be evaluated. Too often, in workload-research, one aims at evaluating the job or task as an entity without being explicit about which task aspects contribute to the resulting workload. Even a multi-dimensional instrument is not a guarantee that all factors contributing to the workload are being taken into consideration.

4.5 Construction of the Rating Scale Mental Effort (RSME)

4.5.1 Choice of procedure

People often generate evaluations about the difficulty of ongoing experiences and the impact of those experiences on their physical and mental state. They rarely quantify or verbalize these fleeting impressions but when they

do, they naturally describe their experiences with verbal terms and modifiers (e.g. "high", "easy", or "moderate") rather than with numerical values. Furthermore in their evaluations subjects hardly differentiate between various aspects of the tasks like time pressure and certain cognitive operations. Only when subjects are explicitly asked to justify their evaluations will they refer to different task demands ("it was very easy, *because* I had lots of time to complete the task").
The above-mentioned 'Cooper-Harper scale' and 'Borg-scale' are very simple rating scales, but apparently they have been applied quite successfully (Borg, 1962, 1976; Moray, 1982; Casali & Wierwille, 1983; Skipper et al., 1986). These scales employ verbal statements as anchors. The success of these scales, and the fact that they are very easy to apply, and hardly interfere with the task has inspired the development of such a rating scale for measuring effort.
However, verbal labels cannot be located on a continuum (or scale) without considering the genuine meaning of such verbal descriptions. Very often rating scales in the behavioural sciences (cf. the Cooper-Harper scale and the Borg-scale) employ the following (or similar) set of verbal anchors: "very ....", "above average", "average", "below average", "very ...." whereby the label "average" is often chosen as midpoint of the scale. However, it appears that the term "average" may not always be a true midpoint (French-Lazovick & Gibson, 1984). For instance in characterizing human behaviour, the term "average" is felt to be rather denigrating (people do not like to be qualified as "average"). This connotation with the word "average" also goes for other objects and so it is that "average" has come to often be regarded as belonging below the midpoint of a scale.
Using the term "average" as a midpoint may result in shifting the distribution of the scores obtained towards the higher end of the scale. It is therefore important to pay careful attention to the verbal labels chosen and to their location on the scale.

There are several techniques for finding scale values, for instance Thurstone's pair-wise comparison technique (Swanborn, 1982). These techniques result in scales with interval characteristics. A problem with these techniques is that they require a large number of comparisons. One technique with properties that are in this respect attractive is the psychophysical scaling method. The measurement methodology of psychophysical scaling has been refined by the Harvard group led by Stevens (1966). Psychophysical functions describe the relationship between variations in the amplitude of a defined physical quantity and the psychological perception of these changes, examples are: brightness, loudness and so on. In a typical psychophysical scaling experiment subjects are asked to express their subjective perception of the physical stimuli in an intensity (or size) ratio, whereby the stimuli are compared to a reference stimulus. This procedure has also been used to scale non-physical stimuli (Hamblin, 1974; Saris et al., 1977; Lodge,

1981). An important aspect of this procedure is that according to the 'cross-modality paradigm' (Lodge, 1981) the same items can feature several times within different 'modalities', like 'line production' and 'numerical estimation'. This enables one to compute indices of equivalence which may be interpreted as indicators of test-retest reliability: with 'line production' for the first estimate and 'numerical estimation' for the second estimate (or visa versa). This method results in scales with (log)interval measurement characteristics (Saris et al., 1980; Lodge, 1981; Swanborn, 1982).
When exact numerical values that locate the positions of the verbal labels on a continuum have been determined answers can be given to questions like whether or not these anchors have more or less the same meaning for different persons. This is necessary if such a scale is to be used inter-individually.

4.5.2 Scaling verbal labels

Within a broader study into the workload of bus drivers employed by the municipal transport company in Groningen an attempt was made to develop a scale with verbally labelled anchor points that referred to the degree of effort expenditure and had realistic intervals between the anchor points. This scale was intended for use during task execution: subjects were expected to rate the amount of effort that they had to expend at that moment. Later on these ratings were translated into numeric scores. It would be helpful if the labels could be located on the scale in such a way that the distances between the labels reflect the relative magnitude of the intervals on an interval scale.

In the Dutch language there are two different words for the word 'effort': the words 'moeite' and 'inspanning' which come closest to the concept 'effort'. It was decided to use both in parallel versions of the scale and therefore the scaling procedure was followed for both versions: version I for 'inspanning', and version II for 'moeite'.

Subjects

In order to construct the scale two different groups were formed consisting of people with different educational and professional backgrounds as it was believed that a person's level of education might influence their interpretation and evaluation of the verbal labels used. The first group consisted of 39 psychology students from Groningen University and the second group consisted of 25 bus drivers of the municipal bus company in Groningen. Both groups were divided in two and each half of each group took part in the construction of scale version I while the other half of the group took part in the construction of version II. Altogether 31 subjects received items belonging to version I and 33 subjects received items belonging to version II.

Procedure

The subjects were asked to rate the items mentioned below using the response modalities 'line production' (LP) and 'numeric estimation' (NE). Line production is: drawing lines of a certain length to express the strength of one's impression; 'numeric estimation' is: assigning numbers to express the strength of one's impression. In order to get acquainted with this scaling procedure and with the two response modalities, the subjects were briefly trained.

The items were (17) verbal labels of which the scale values had to be determined. These verbal labels indicated different degrees of effort (" very effortful", "rather effortful", "moderately effortful", "not effortful", etc.). In other words, the items are supposed to represent a single underlying dimension or, amount of perceived effort (see Table 4.1).

Table 4.1: *Items of both versions*

		Version I	*Version II*
A	=	'Enigszins inspannend'	'Enige moeite'
B	=	'Behoorlijk inspannend'	'Behoorlijk wat moeite'
C	=	'Erg inspannend'	'Erg veel moeite'
D	=	'Niet zo inspannend'	'Niet zo veel moeite'
E	=	'Zeer inspannend'	'Zeer veel moeite'
F	=	'Nogal inspannend'	'Nogal wat moeite'
G	=	'Een beetje inspannend'	'Een beetje moeite'
H	=	'Totaal niet inspannend'	'Totaal geen moeite'
I	=	'Ontzettend inspannend'	'Ontzettend veel moeite'
J	=	'Nauwelijks inspannend'	'Nauwelijks moeite'
K	=	'Niet al te inspannend'	'Niet al te veel moeite'
L	=	'Vreselijk inpannend'	'Vreselijk veel moeite'
M	=	'Tamelijk inspannend'	'Tamelijk veel moeite'
N	=	'Helemaal niet inspannend'	'Helemaal geen moeite'
O	=	'Niet inspannend'	'Geen moeite'
P	=	'Heel erg inspannend'	'Heel erg veel moeite'
Q	=	'Matig inspannend'	'Wel wat moeite'

These items were presented in a randomized order in a booklet with one item to a page. The subjects had to express a ratio between the 'reference item' and the other items. The last item (item Q) was chosen as a reference item. An arbitrary value of 50 was assigned to this reference item.

Two response modalities were used (Line Production and Numeric Estimation) according to the cross-modality paradigm (Lodge 1981). (For a detailed description of the technique and procedure the reader is referred to Lodge (1981) and for a detailed analysis of the results to Zijlstra & van Doorn, (1985), and van Doorn & Zijlstra, (1988)).

4.5.3 Results of scaling procedure

The product-moment correlation between 'line' and 'number' scores for each individual is used as a 'parallel' reliability measure. In Table 4.2 these correlations are presented for both versions.

Table 4.2: *Product-moment correlation between Line Production and Numeric Estimation scores for each individual.*

	Version I	Version II
mean	.951	.951
lowest value	.791	.876
highest value	.989	.995
no. lower .85	1	0
no. of subjects	31	32

One must not forget that one subject out of the 64 respondents did not complete the test because he found the task too difficult. From the remaining group all the respondents except one had correlations above a lower boundary of .85. Therefore we concluded from these results that almost all respondents can give reliable and consistent ratings for these items.
The next step is to see whether the scales of all the respondents are identical. This is an important part of the study because even if individuals give reliable responses they will not necessarily order the items in the same way. This is why the individual scales were compared with an overall group scale that was constructed. The correlations are presented in Table 4.3.

Table 4.3: *Product-moment correlations of individual scales with the 'group scale'.*

	Line Production	Numeric Estimation
mean	.958	.955
lowest value	.883	.804
highest value	.997	.994
number of resp.	63	63

Apparently all the respondents had individual scales that did to a large extent correspond with the group scale because all the individual scales correlated very highly with the average overall group scale. This means that there is a consensus about the relative scale positions of the items. All respondents, irrespective of educational or professional background, placed the items in the same relative position on an imaginary continuum. In other words, differences in the respondents in educational or professional backgrounds of the respondents did not significantly affect the evaluation and interpretation of the labels presented.

To construct a scale based on the items the geometric mean of all the subjects' item scores was calculated. (According to Lodge (1981) the geometric mean is a standard measure of central tendency with magnitude estimation data). The scale values of the labels are presented in Table 4.4. The items that were used to construct the scale were selected so that there would 'equal appearing' intervals on the scale. Figure 4.1 presents the resulting scale (version I): the Rating Scale Mental Effort.

Table 4.4: *Scale values of the items. Items marked with * were used in the resulting scales (see Figure 4.1).*

	version I	version II
Enigszins inspannend (A)*	37	38
Behoorlijk inspannend (B)*	71	70
Erg inspannend (C)*	86	102
Niet zo inspannend (D)	23	28
Zeer inspannend (E)	94	119
Nogal inspannend (F)	61	67
Een beetje inspannend (G)*	26	15
Totaal niet inspannend (H)	2	2
Ontzettend inspannend (I)	114	138
Nauwelijks inspannend (J)*	13	10
Niet al te inspannend (K)	24	24
Vreselijk inspannend (L)*	112	137
Tamelijk inspannend (M)*	57	58
Helemaal niet inspannend (N)*	2	2
Niet inspannend (O)	3	2
Heel erg inspannend (P)*	102	113

Although it can be seen that respondents assigned roughly the same relative scale positions to the items involved this does not necessarily mean that the absolute scale values are identical. This scaling procedure usually results in a wide range of raw item scores. However, these differences constitute no violation of the assumptions for the (log)interval measurement level (Saris et al., 1984).

When both versions (I and II) are compared it is evident that, although each version was developed independently, the ratios between the items in both versions are the same, for example item B was almost twice as high as item A in both versions. Therefore one might say that this scale has ratio measurement characteristics but since no nil-point was defined it is better to call it a scale with interval characteristics. As long as the nil-point is not defined, the absolute values of these item scores are meaningless. An effort degree of 'zero' makes no sense, even the easiest of tasks requires some, though perhaps not much, effort. In fact this is also expressed with the items 'Totaal niet inspannend' and 'Helemaal niet inspannend' (completely undemanding) and these items did receive a higher than 'zero' rating.

From what has been stated above it will be clear that it is difficult to interpret the absolute scores of respondents. For the time being all analyses are accepted which compare relative size of scores or average scores. This means, for example, that changes in perceived effort can be compared between the groups of respondents and the mean scores of groups of respondents for different tasks will be robust.

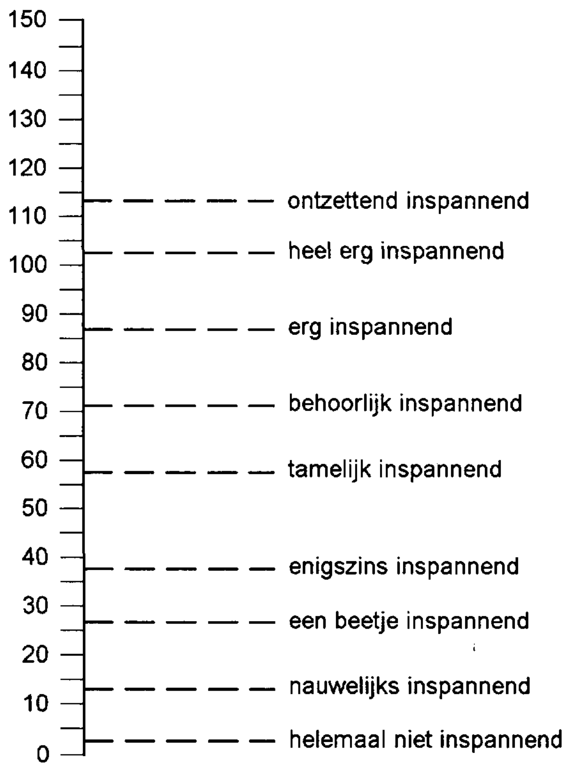

Figure 4.1: *Rating Scale Mental Effort*

These scale construction procedures were replicated with other groups and resulted in exactly the same scale values. The scale was also constructed in German according to the prescribed procedure and was done in cooperation with a German research group (Eilers, et. al., 1986). The scale values of the corresponding items did not deviate significantly from the Dutch items: the positioning of the items on the scale was similar.

4.6 Initial validation of the Rating Scale Mental Effort

Part of the procedure in developing an instrument is connected with assessing the utility of such an instrument. If an instrument is going to be used for research purposes it should meet certain requirements. Several of these requirements are listed here below:

1. Sensitivity. How sensitive is the instrument with respect to changes in task load, psycho-physiological state, etc.
2. Reliability. This has to do with the instrument's accuracy, that is to say, if the measurement procedure is repeated under different conditions will the results be the same? This is sometimes referred to as 'test-retest' reliability or 'reproducability'. It is important therefore to know something about the stability of the trait or construct that one is after.
3. Validity. Several kinds of validity can be distinguished (Drenth & Sijtsma, 1990): predictive validity, concurrent validity, content validity, construct validity, etc. What is important is, that one collects empirical evidence to be sure that the instrument really measures what one intends it to measure. This aspect may have implications for the diagnosticity and selectivity of an instrument.
4. Intrusiveness. The degree to which the measurement procedure interferes with the subject's behaviour or irritates him.
5. Costs. This is a pragmatic, but nonetheless very important aspect. The more costly a measurement procedure is, the more important the above mentioned aspects will be.

Regarding the dissimilarity of definitions pertaining to mental workload and mental effort, special attention should be given to the construct validity of the instruments that are used to measure these concepts. In the light of what has been described in the previous chapter, it is relevant to ask whether the current indicator for mental effort reflects mental effort as the activation of a 'central computational mechanism' or as an 'executive resource control'. Special attention will be given to this aspect in the following paragraph.

In this paragraph the first experiment in which the Rating Scale Mental Effort was employed will be discussed.
This experiment was part of a more extensive research project into the workload of city bus drivers at the municipal bus company in Groningen (Mulders et al., 1982, 1988; Kompier, 1988).

In this project 27 bus drivers agreed to participate in an extensive research program. The drivers were examined on several different days. At least two of these days were working days and one was a duty-free day (for a detailed description of the design of this study the reader is referred to Kompier, 1988). In this study the drivers were required to rate their effort expenditure at various moments during their working day.
Furthermore they had to perform specially developed laboratory tasks that were designed to measure workload effects on a routine working day. These laboratory tasks had to be performed three times: at the start of the working-day (at about 9.00 hours), during the lunch-break (13.00 hours) and at the end of the working day (17.00 hours). This procedure was repeated on a second working day. First the results of this laboratory part of the study will be described.

These laboratory tasks are so-called visual memory-search tasks (Sternberg, 1969; Massaro, 1975; Mulder, 1980). During such a task a set of stimuli is presented on a video screen for a short period. This set of stimuli consists of a variable number of letters (1 - 4), the so-called 'display set'. The subjects are asked to indicate whether the (or one of the) letter(s) presented belong(s) to a previously shown set of letters, which should be memorized (the so-called 'memory-set'). The subjects have to respond by pushing one button for a 'yes' response and another button for a 'no' response to show that one or more of the letters of the presented display-set does or does not belong to the memory-set. The memory-load of the task can be manipulated by varying the number of letters in the memory-set. This means that the difficulty of the task or, one should say, the task load, is varied. (For a detailed description of this task see: van Dellen et al., 1985; Aasman et al., 1988).

In the experiment to be described here memory-sets of two letters and of four letters were used. Furthermore an experimental condition was created in which a subject was requested to count how often certain letters from the memory-set were presented. This resulted in four different task load levels with increasing information loads:

1. 2 letters in memory and pushing a button in response;
2. 4 letters in memory and pushing a button in response;
3. 4 letters in memory and counting in response;
4. 4 letters in memory and counting and pushing a button in response; this task may be regarded as a dual-task.

The task load levels, each consisting of a high number of trials, were presented in a random order. After each task load level the subjects had to rate their effort expenditure by means of the Rating Scale Mental Effort (RSME). Other measurements were also made, like reaction and heart rate

variability. A task like the one described above takes about 20 minutes to complete.
The results of the ratings on the first working-day are presented in Figure 4.2^{a} and the results gathered on the second working-day are presented in Figure 4.2^{b}.

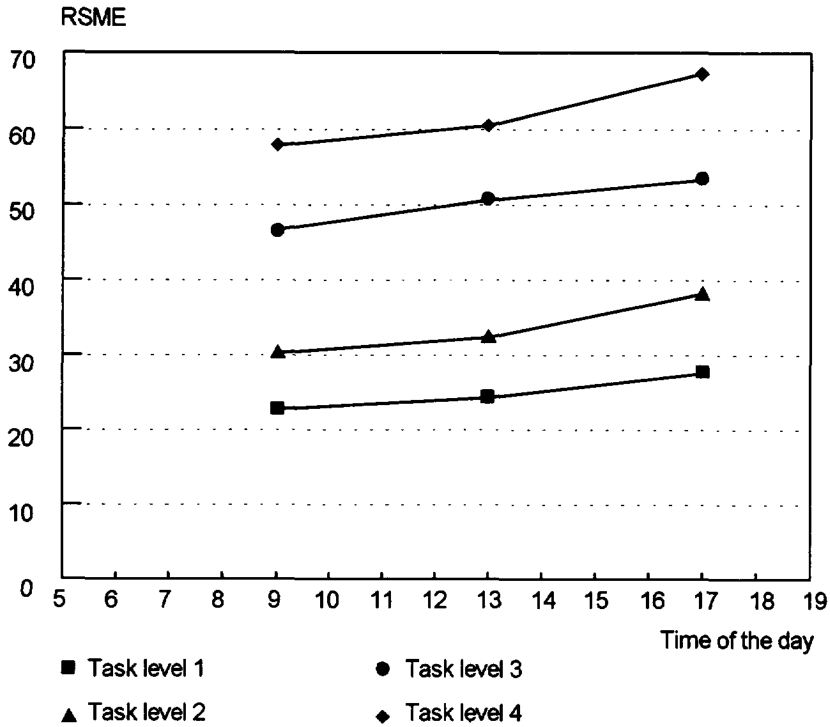

Figure 4.2a: *Laboratory task, working-day 1*

According to both figures the RSME is able to differentiate systematically between various information load levels. The mentally more demanding tasks are given higher scores than the mentally less demanding tasks. Analysis of Variance showed significant effects for the factors 'task level' ($F(3,24) = 31.2$; $p < .001$, and task level 1 < task level 2 < task level 3 < task level 4) and 'time of the day' ($F(2,25) = 7.4$; $p = .003$, and score on time 9.00 < score time 13.00 < score time 17.00). No significant effect for the factor 'working day' was found. This is important, because comparing the effort-scores on both working days (Fig. 4.2^{a} vs. 4.2^{b}) gives an indication about the reliability of this rating scale as the second working day is a replication of the first one. Since there are no significant differences we can conclude that the Rating Scale Mental Effort provides us with reliable measurements. The

RSME scores of working day 1 correlate (product moment correlation) r= .81 (p=.001) with those of working day 2.
The increase of the RSME score accompanying the increase in information load is an indication that the RSME provides information on mental effort as 'controlled information processing': the higher the information load, the more mental effort is needed to process the information.

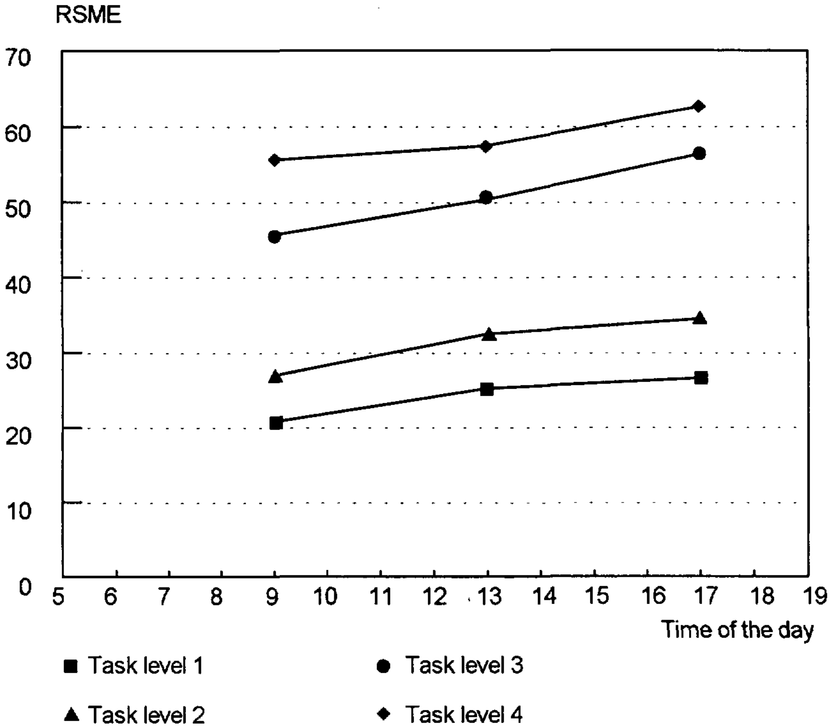

Figure 4.2b: *Laboratory task, working-day 2*

The differences between the various levels of information load were also reflected in the reaction times because a considerable increase in reaction time was seen between the simplest condition (level 1) and the dual task condition (level 4). In addition some effects of the increase of information load could be traced in the physiological parameters. Heart rate was found to be significantly faster in the dual task condition than in the simplest condition while heart rate variability was lowest in the dual task condition (see for an extensive description of reaction time and heart rate results: Aasman et al., 1987, 1988). These physiological reactions are presumed to accompany mental effort exertion (Mulder, 1980).
The findings with respect to the reaction times can be regarded as an indication of 'converging' evidence of the validity of the RSME while the

physiological results can be regarded as an indication of the 'concurrent validity' of the RSME.

Furthermore it appeared that the RSME score for each task level was systematically highest at the end of the working day (17.00 hours), while the information load of the experimental task was the same as at the start of the working day. By 17.00 hours the subjects had done a day's work which brought about a change in their psycho-physiological state. The bus drivers had been taxing their resources and so their performance potential was lower: they were tired. The increase in the RSME score can be regarded as an indication that the RSME also provides information on effort like 'executive resource control'. Since a bus driver's psycho-physiological condition was sub-optimal he had to compensate by investing extra effort in order to perform at an acceptable level.

RSME scores were also gathered while the drivers were performing their real task of driving a bus. Several times during their working day the bus drivers had to rate how much effort they were exerting. The numbers of passengers transported and the time left to spare upon arrival at the end of his route was also noted. These parameters were used as indices for the task load. Figure 4.3 presents the results of this effort rating.

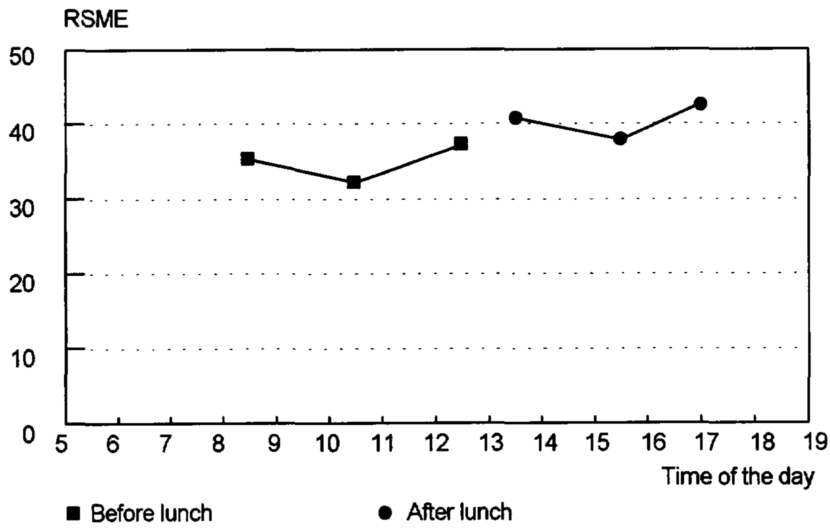

Figure 4.3a: *RSME during working-day 1*

As can be seen in Figure 4.3 the bus drivers reported that they have to invest more effort to carry out their task at the end of the day. Analysis of variance teaches us that there is no significant effect of the factor 'working day' (w1 vs w2) on the one hand but that on the other hand it appears that the factor 'time' (end of the day versus early morning) shows a significant effect ($F(5,23) = 11.6$; $p = .001$). No significant interaction effects were found. It turned out that an increase in number of passengers during the morning period correlated ($r_{pm} = +.30$) with an increase in effort ratings and a decrease in spare time at the end of the bus route correlated $r_{pm} = -.33$ with an increase in effort ratings (These task load indices correlated $r = -.31$ when scores of differences were calculated). During the afternoon period the correlations became respectively greater: $r = +.46$ and $r = -.64$ for passenger numbers and amount of time to spare. The correlation between both task load indices also increased ($r = .-53$) (see also Meijman, 1991 page 180). The dip in the curve of the effort ratings in the mid morning period (the differences between times 1, 2, and 3 was found to be significant) and to a lesser extent (not to a significant extent) in the afternoon period. This reflects the fact that around those periods (after the morning rush-hour and before the evening rush-hour) there are fewer passengers and there is less traffic. This makes the bus driver's task easier. Moreover we saw the same pattern on both working days. The scores of working day 1 correlate highly ($r = .71$; $p = .001$) with those of working day 2. This again is an indication that subjects give reliable estimates.

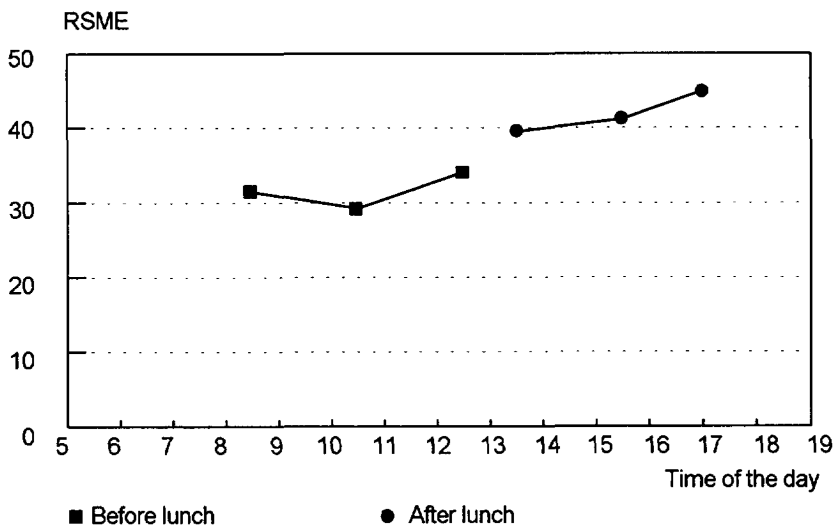

Figure 4.3b: *RSME during working-day 2*

4.7 Conclusion

In general rating scales appear to be very sensitive instruments. However, the various rating scales that are regularly employed in this field do not take into account the conceptual difference between workload and effort. This chapter describes the construction of a simple rating scale, i.e. the Rating Scale Mental Effort. This scale employs verbally labelled anchor points that refer to an underlying continuum of effort expenditure.

The first question, of course, is whether this scale really measures mental effort investment. In the previous chapter, on the theoretical aspects of mental effort, it has already been stated that the level of effort expenditure is determined by the objective task demands in relation to the 'performance potential' of the individual. Practically speaking this means that if it is assumed that an instrument measures mental effort expenditure, this instrument should reflect differences between various task load levels as well as changes in the performance potential of the individual.
In the first study in which this rating scale was used it appeared to differentiate systematically between various levels of information load and there are indications that changes in the performance potential of individuals are also reflected in the RSME score. This suggests that the Rating Scale Mental Effort does indeed measure effort. Furthermore it has been demonstrated that the RSME is easy to apply both in laboratory settings and in real work situations. Subjects give reliable estimates of the various task load levels when using the RSME and the instrument proved to be non-obtrusive.

In the next chapter an experimental study will be described that is directed towards assessing the construct validity of the Rating Scale Mental Effort.

Chapter

5 Validation of the Rating Scale Mental Effort[1]

5.1 Introduction

As has been pointed out in Chapter 3 mental effort is conceived as being related to the concept of attention which, on the one hand, is seen as a central regulatory process in information processing and, on the other hand, as the compensatory process that is needed to adjust the sub-optimal psycho-physiological state of the individual to the particular state that is required by the demands of the task. Recent models of mental effort (Sanders, 1983; Mulder, 1986) propose integrating both approaches to mental effort. Since both approaches refer to the energetic metaphor it was decided that effort should in the first place be viewed as 'energy' (Chapter 3). This immediately raises the issue of the measurability of such a concept. This issue also relates to the point about the construct validity of the Rating Scale Mental Effort.
In the preceding chapter we saw how the Rating Scale Mental Effort (RSME) differentiates between various levels of task load. It also appears that RSME scores are influenced by changes in the psycho-physiological state of the subject caused by prolonged, intensive work.

In this chapter the validity and usability of the RSME will be the central topic. I will therefore be discussing at length a laboratory study that was designed to examine the construct validity of this scale. Several other studies in which the RSME has been used will also be looked at so as to obtain further information on the usability of the RSME.

5.2 Outline of the experiment

5.2.1 General plan

The construct validity of the RSME will be examined by means of a laboratory experiment in which both aspects of mental effort: the 'state control' and the 'processing effort' approach, are manipulated. In order to be

1 The experiment described in this chapter has been carried out at the Institute for Experimental and Occupational psychology at the University of Groningen. This study was financial supported by a grant from the Dutch Organization for Advanced Research (NWO - PN 560-265-025).

able to study mental effort as a 'mechanism' that regulates the flow of attention that is needed for processing information the difficulty and complexity of the tasks to be carried out were manipulated. The subjects performed the same type of memory-search tasks as those described in the previous chapter. The demands of the task were manipulated by varying the memory load and the processing complexity of the information that had to be processed.

Studying mental effort as a compensatory process involves manipulating a person's psycho-physiological state. The psycho-physiological state of our subjects was manipulated by introducing two pre-treatment conditions for the subjects before the experiment sessions started.

5.2.2 Manipulation of the task-load and the processing complexity

For the experiment sessions in the laboratory subjects were seated in front of a visual display unit on which characters were presented to them. First of all a set of characters was presented which the subject had to memorize (the memory-set) after which a set of four characters (the display-set) was presented. This display set did (or did not) contain one of the characters of the memory set. Upon pushing one of two buttons the subject's reactions were registered accordingly as positive or negative. Four variants of the above mentioned task were used. The task load was manipulated by presenting two levels of memory load: one character in the memory set (task 1) and four characters in the memory set (task 4). Each task level consisted of 170 trials, which made the total duration about ten minutes.

The processing complexity was also manipulated in accordance with the Shiffrin & Shneider paradigm (1977) by introducing a 'consistent mapping' and a 'varied mapping' version of the task. In the consistent mapping condition the memory-set remained the same during all the trials. After intensive training this mapping condition results in 'automatic' processing, that is to say, fast and parallel information processing requiring hardly any attention (effort). This effect can be enhanced by using numbers as distracters in the display-set. The fact that these tasks are performed 'automatically' becomes apparent from the decrease in reaction times as the learning process progresses. After enough training has taken place the reaction times cease to decrease but instead remain at a constant level even when the number of letters in the memory-set increases. During this training period it is assumed that subjects establish a relationship between the letter(s) displayed in the memory-set and the 'target-letters'. This means that after a time the subjects just scan the pattern of the target-letters and try to match these with the pattern of the 'memory-set' thus saving themselves from carrying out exhaustive attention-demanding (thus effortful) search processes in the working memory. It is assumed that an increase in memory load in the consistent mapping condition will not require more effort.

In the varied mapping condition each display-set (containing the 'target-letters') is preceded by a new memory set. This means that the subjects do not have the opportunity to establish a relation between the letter(s) in the memory-set and the 'target-letters' to which they have to respond. Consequently it is assumed that for each trial extensive search processes in the working memory are necessary. This is called 'controlled' processing of information, which is slow, serial and effortful. In contrast to the consistent mapping condition subjects do not show a decrease in reaction times after the same extensive training period. In this condition an increase in memory load will always be accompanied by an increase in expenditure of mental effort.

At the end of each laboratory session a second task was administered, the so-called 'QRST-task'. This task based on the same principle as the preceding task was turned into a 'double-task'. Subjects were not only requested to respond when one or more of the letters presented belonged to the memory-set, but also to count how often a letter from the memory-set was presented in the display-set. For this task the memory-set always contained four letters. (In fact this is task level 4 from the experiment described in the previous chapter). It should be noted that there is no consistent mapping version of the QRST task. This task included 40 trials and therefore lasted about three minutes.
During one experiment session each task-level (memory load 1 and memory load 4) was presented twice with a rest period of approximately 5 minutes in between. In one session all the experiment tasks were presented in a consistent mapping version while in another session all the tasks were presented in a varied mapping version. The order of presentation of the task levels and of the consistent and varied mapping conditions were balanced in order to ensure that their sequence would not affect the final outcome. The duration of a complete laboratory session was approximately 60 minutes.

5.2.3 Manipulation of the psycho-physiological state

The 'state control' approach requires manipulating a subject's state. A well-known change, usually known as fatigue (see Meijman, 1991), leads to the state that can be observed after prolonged work, for instance at the end of a working-day.
Before the experimental task in the laboratory setting commenced the subject's psycho-physiological state was manipulated. Two pre-treatment conditions were distinguished: 'pre-treatment A' and 'pre-treatment B'. Pre-treatment A was meant to simulate a subjects's working day[2]. The subjects

2 The subjects who took part in this experiment were 32 male students (non-psychologists) who volunteered to participate. They varied in age from 18 to 29 years (average 22.9 years) and were paid Dfl. 10,--/hour.

arrived at the institute at 09.00 hours and worked on mentally demanding tasks until 16.00 hours. In the pre-treatment condition the subjects first had to summarize scientific articles. After that they had to do English listening comprehension tests. During the first part of the pre-treatment session subjects were instructed to finish their summaries within a certain time limit. These time limits were rather tight which meant that subjects were required to work under pressure. This pre-treatment condition was defined as '*mentally active* and *physically passive*'.

The pre-treatment ended at 16.00 hours at which time the experiment sessions in the laboratory started. These sessions took place in another room. During these sessions the standard tasks described above had to be performed.

The other pre-treatment condition was intended to simulate a day off. In this pre-treatment B situation the subjects were asked to stay at home and to abstain from heavy physical and mental activity such as doing sport or studying. In this experimental condition the subjects arrived at the institute one hour before the start of the laboratory experiment sessions.
All subjects travelled the distance of approximately eight kilometres to the institute by bicycle. This pre-treatment condition was described as '*mentally passive* and *physically active*'.

It was thought that because the subjects had been engaged in prolonged mentally demanding tasks during pre-treatment A this would induce mental fatigue and it was assumed that the accompanying psycho-physiological state would be less optimal for the experimental tasks that followed. Furthermore subjects were physically passive. As stated before, compensatory effort is needed whenever the subjects' actual state is sub-optimal alongside of the state that the demands of the task require. Consequently it was predicted that the subjects would have to invest more effort during the memory-search tasks after they had been exposed to pre-treatment A than after having been exposed to pre-treatment B.

5.2.4 Manipulation of subjects' amount of control over the task

According to some researchers (Lundberg et al., 1980; Frankenhauser, 1986, Karasek & Theorell, 1990) the degree to which people are able to control their work influences the extent to which stress-like reactions become apparent. In order to investigate whether the degree of control over the task also influences the amount of effort that has to be exerted a fourth factor was introduced. This factor resulted in two conditions the first was one *in which subjects had no control over the task speed. The speed at which the* task had to be executed was determined by the computer that controlled the experiment a (machine paced condition). In the second situation the subjects controlled the speed at which the task was executed. As soon as the subject responded (by pushing a button) the next trial would be initiated (this was the unpaced condition).

5.3 Design and methods

There were four experimental factors in this experiment.

a. One between-person factor: manipulation of control over the task, two levels (machine paced, unpaced).
b. Three within-person factors:
   1. manipulation of the psycho-physiological state, two levels (pre-treatment A, pre-treatment B);
   2. manipulation of information processing mode, two levels (consistent mapping, varied mapping);
   3. manipulation of memory-load, two levels (1 or 4 four characters in the memory set).

These factors were elaborated in an experimental design in the following way: four research-days were reserved for each subject. On two of the research days subjects underwent pre-treatment A and on two of the research days they underwent pre-treatment B. The experiment sessions in the laboratory took place during the second half of the afternoon (at about 16.00 hours) on each research day. Each experimental task was presented in a consistent mapped (CM) version and in a varied mapped (VM) version and each task came after a pre-treatment A or a pre-treatment B session. At the end of each laboratory session the QRST task was also administered.

The group of subjects was split into two halves (subjects were randomly assigned). The first group of 16 subjects were placed in the machine paced 'condition' and the second group of subjects did the unpaced experiment.

This resulted in four different experimental conditions for each subject in both the groups (machine paced and unpaced):

- Pre-treatment A and consistent mapping;
- Pre-treatment A and varied mapping;
- Pre-treatment B and consistent mapping;
- Pre-treatment B and varied mapping.

All the relevant factors like the order of presentation of the pre-treatment conditions, memory load-levels and the information processing mode were balanced between the subjects to make sure that the order of events would not affect the results.

The resulting experimental design is schematically presented in Figure 5.1.

machine paced (n=16)				unpaced (n=16)			
pre-treat A		pre-treat B		pre-treat A		pre-treat B	
CM	VM	CM	VM	CM	VM	CM	VM
Task: 1 4 QRST	Task: 1 4 QRST	Task: 1 4 QRST	Task: 1 4 QRST	Task: 1 4 QRST	Task: 1 4 QRST	Task: 1 4 QRST	Task: 1 4 QRST

Figure 5.1: *the experimental design.*

In the third chapter (section 3.7.1) three classes of measurement techniques were described. In this experiment indicators from all three categories were used:

a. physiological;
   - state indicators: heart rate (inter-beat interval time) and adrenaline and noradrenaline level;
   - information processing: heart rate variability index.
b. behavioural;
   - information processing: reaction time, percentage of correct responses.
c. subjective;
   - state indicators: feelings of activation and tension as measured by the GACL (Hellinga, 1985).

Furthermore the Rating Scale Mental Effort (RSME) was used.

The adrenaline level indicates the subject's general activation level while noradrenaline is primarily supposed to reflect the degree of physical activation (Lehmann, et al., 1982; Fibiger, et al., 1984). Together with the heart rate and the feelings of activation and tension these parameters were used to study the effects of the pre-treatment conditions (see Hockey, et al., 1986). Adrenaline and noradrenaline excretion was measured by analyzing urine samples. For the purposes of this analysis urine samples were collected during the research days at several predetermined times. From these urine samples it was possible to ascertain the excretion rate of (nor)adrenaline during the preceding period (Westerink, et al., 1982). At the same time subjects had to fill in a questionnaire (Groninger Adjective Checklist - Hellinga, 1985). This questionnaire was a translated and modified version of Thayer's (1978) 'activation adjective checklist'. It contained mood statements (I feel sleepy, ... angry, anxious, etc.) and consisted of two sub-scales: the degree to which the subject feels activated and the degree to which he feels tense.

While tasks were being performed in the laboratory people's heart rates were recorded. By means of spectral analysis[3] (cf. L. Mulder, 1988) the heart rate variability was calculated. Particularly the band width around 0.10 Hz (0.07 Hz. - 0.14 Hz.) is used as a physiological indicator that mental effort is being exerted (G. Mulder, 1980). It is believed that a decrease in heart rate variability reflects an increase in the degree of mental effort that is being exerted.
Reaction time and the heart rate variability index were used to study the manipulation of the task load (Hockey et al., 1986). Both kinds of indicators ('state', and 'processing') were used to study the convergent validity of the Rating Scale Mental Effort (RSME). Subjects had to give their evaluations immediately after each task level.

5.4 Hypotheses

This chapter deals primarily with the question of whether a subjective instrument, in this case the RSME, is capable of measuring mental effort. If this is so then the scores on this instrument should discriminate between the several experimental conditions created in this experiment. What is especially interesting in this respect is the distinction between both modalities of information processing. Subjective instruments are generally believed to reflect only overall changes in psycho-physiological state and not temporary changes in the processing of information (Hockey et al., 1986; Gopher & Donchin, 1986). By manipulating both the factors 'pre-treatment' and 'mode of information processing' in one experiment we should be able to gain a clearer insight into this topic.

More explicit hypotheses concerning the scores on the RSME may be formulated. The modality of information processing is believed to influence the level of effort that is required. Automatic processing is supposed to proceed autonomously and not to be attention-demanding. Therefore I first of all expect to find lower RSME scores in the consistent mapping conditions than in the varied mapping conditions. Secondly I expect that a more difficult task will require more effort than an easier task. In this experiment this means that the tasks with a memory load of one letter in the memory-set will produce lower scores on the RSME than the tasks with four letters in the memory-set. But this only holds for the varied mapping conditions. Further more it is predicted that an increase in task load (or rather in memory-load)

3 The program CARSPAN, developed at the Institute of Experimental Psychology, was used to calculate HRV from interbeat intervals.

under the consistent mapping conditions will not result in an increase in the RSME scores. On the other hand it is expected that under the varied mapping conditions an increase in memory load will result in an increase in the RSME score.
Finally, it is thought that sub-optimal condition subjects (those who had followed pre-treatment A) will have to exert more effort to perform their tasks than the optimal condition subjects (those who had completed pre-treatment B) which will make their RSME scores higher. The subjects have to adjust their state at that moment to the state required by the task performance. With respect to the factor 'modality of information processing' it is expected that only the controlled processing mode will be affected by changes in state. This mode is the attention-demanding processing mode and so it is thought that extra effort might be required to sustain the performance level.

5.5 Results

5.5.1 Pre-treatment

To find out whether the pre-treatment had had the desired the effects we have to look at the relevant parameters. It is important to know whether the various pre-treatment conditions had created different psycho-physiological states at the start of the laboratory sessions. Here I will only present the salient points regarding the changes in psycho-physiological state; for a detailed description of these results see Zijlstra & Meijman (1989).

neuroendocrine reactions

Figure 5.2 presents the curve of the adrenaline excretion rate during the research days.
Urine samples were collected at 14.00 hours and 16.00 hours (just before the laboratory sessions started) and at 17.30 (after the laboratory sessions had finished). Figure 5.2. shows that a clear difference between the two pre-treatment conditions is reflected in the adrenaline excretion levels recorded at the start of the laboratory sessions (16.00 hours). The adrenaline excretion rate is supposed to reflect the general level of activation (physical and mental). Subjects in the pre-treatment A condition had at that moment much lower excretion rates than the pre-treatment B condition subjects. The activation level of subjects in pre-treatment A decreased during the English listening comprehension tests. During the laboratory sessions the change in state had to be forced: they had to make themselves active again as was reflected in the rate of increase in excretion between 16.00 hours and 17.30 hours. In the pre-treatment B condition the subjects had been at home and had cycled to the laboratory at about 15.00 hours. Figure 5.3 illustrates that subjects in the pre-treatment B situation had a higher level of physical activity, as indicated by the noradrenaline curves.

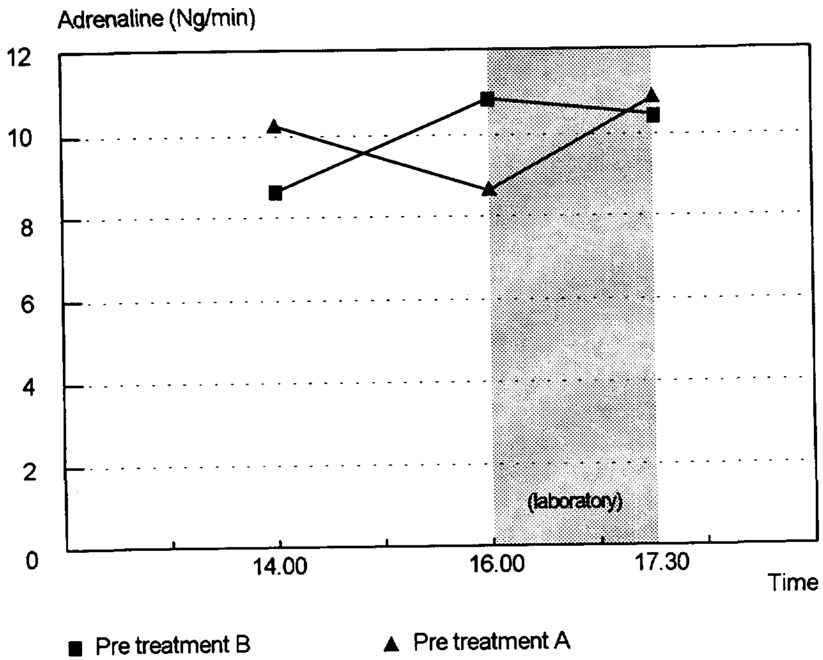

Figure 5.2: *Adrenaline excretion*

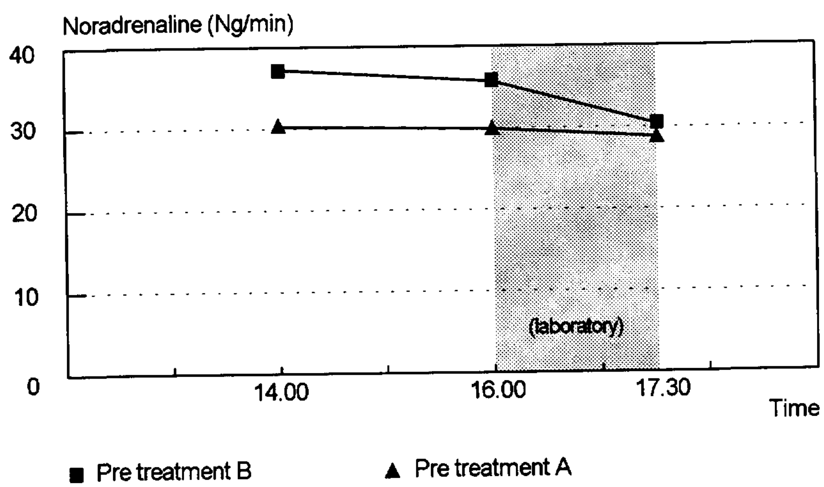

Figure 5.3: *Noradrenaline excretion*

heart rate and subjective parameters

The results on the catecholamines are confirmed by the results pertaining to the heart reactions. A significant decrease in heart rate (longer inter beat interval times) i.e. about 975 msecs was seen in pre-treatment A conditions compared with 925 msecs in the pre-treatment B conditions ($F(1,26)=15$; $p=.001$). The subjects also reported that they felt significantly less active in the 'A condition' ($F(1,30)=18.4$; $p=.001$), as gauged with the Groninger Adjective Checklist.

In conclusion it might be stated that the prescribed pre-treatment tasks created different psycho-physiological states at the onset of the laboratory sessions as could be measured by the physiological and subjective parameters.

5.5.2 Laboratory session

Before discussing the further results of this experiment it should be stated that all balanced independent variables have been examined. None of these variables have proved to have significant effects. For this reason they have been left out of further discussion.
Furthermore it must be noted that the earlier mentioned QRST task is not comparable with Tasks 1 and 4. The QRST task is a double task with less trials that does not have a consistent mapping version. In the first instance therefore this task is left out of the analysis.

Performance

Figure 5.4 presents the Reaction Time results derived from the experimental task during the laboratory session.
Figure 5.4[a] shows that there is a significant increase in Reaction Time between Task 1 and Task 4 in the Varied Mapping condition while the increase in the Consistent Mapping condition is only minor (not significant). This demonstrates that the introduction of a consistent mapping condition (i.e. automatic processing of information) was successful. Moreover it appears that the pre-treatment conditions did not affect the Reaction Times. Figure 5.4[b] shows the results of the 'unpaced' condition. In this condition the subjects were able to determine the task speeds themselves. This figure has the same pattern as the 'machine paced' condition but on closer inspection we can see that in the Consistent Mapping condition the Reaction Times are lower while for Task 4 in the Varied Mapping condition the Reaction Times are higher. This means that the 'easy' tasks are performed faster while subjects took more time to do the 'difficult' tasks. Table 5.1 presents the results of the Analysis of Variance (repeated measurement design with 'pacing' as a 'between subject' factor). This table confirms that the above mentioned effects are found to be significant.

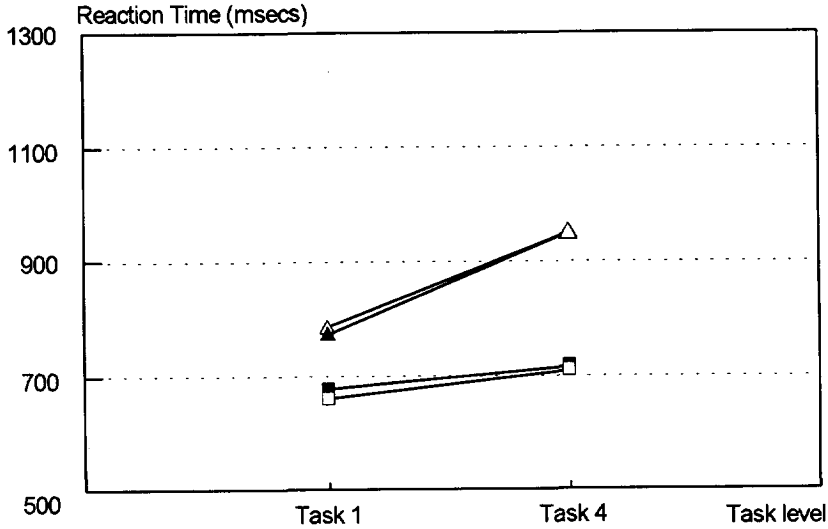

Figure 5.4a: ■ Pre B CM □ Pre A CM

▲ Pre B VM △ Pre A VM

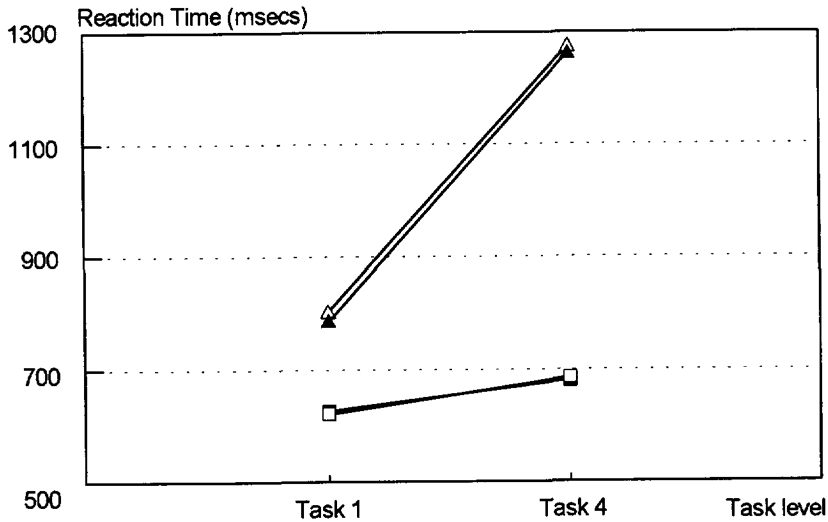

Figure 5.4b: ■ Pre B CM □ Pre A CM

▲ Pre B VM △ Pre A VM

Figure 5.4: *Reaction time for both tasks (Task 1 and Task 4) broken down into pre-treatment conditions (A and B) and information processing mode (Consistent and Varied Mapping, CM and VM respectively). 5.4a is machine paced condition, 5.4b is unpaced condition.*

Table 5.1: *Analyses of Variance for Reaction Times (2 subjects with missing values).*

factor	F	df	p
Constant (=model)	2932	(1,29)	<<.01*
Pacing	5.3	"	.03*
Pre-treatment	.4	"	.55
Pacing by Pre-treatment	.33	"	.57
Mapping	387.7	"	<<.01*
Pacing by Mapping	57.0	"	<<.01*
Memory load	411.7	"	<<.01*
Pacing by Memory load	75.1	"	<<.01*
pre-tr. by Mapping	1.9	"	.25
Pacing by pre-tr. by mapping	.02	"	.88
pre-tr. by memory load	.36	"	.55
Pacing by pre-tr. by memload	.01	"	.92
Mapping by memory load	252.7	"	<<.01*
Pacing by mapping by memload	69.4	"	<<.01*
Pre-tr. by mapping by memload	1.6	"	.22
Pac. by pre-tr. by mapping by memload	.02	"	.88

The significant results are marked with a *

Legenda

'Pacing' is the factor machine paced vs self paced;
this is the between-subjects factor.
'Pre-treatment' is pre-tr. A vs pre-tr. B
'Mapping' is consistent mapping vs varied mapping
'memory load' is memload 1 vs memload 4

In order to obtain an indication of the quality of the performance we should look at the percentage of correct answers given during the laboratory task. These results are presented in Figure 5.5.

This figure demonstrates that there is no effect from the factor 'task-level' in the Consistent Mapping condition, but in the Varied Mapping condition there were significantly more errors in Task 4. Moreover we can see that in the 'unpaced' condition the subjects made less errors while working on Task 4 than when 'machine paced' because they took more time and so made less errors. The quality of the performance is only slightly influenced by the pre-treatment condition (F(1,29)=13.1 p=.04) and then only in the Varied Mapping condition.

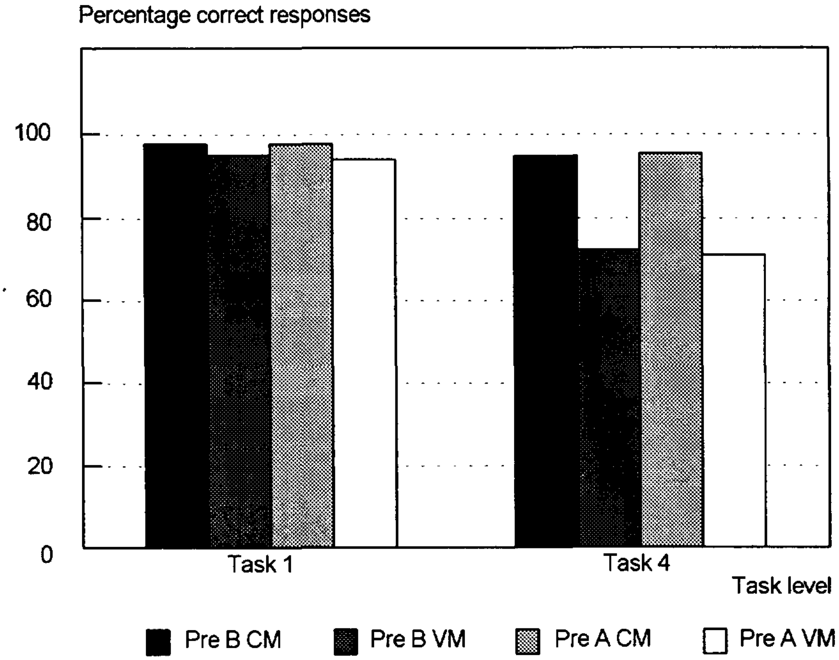

Figure 5.5a: *Quality of Performance, machine paced condition*

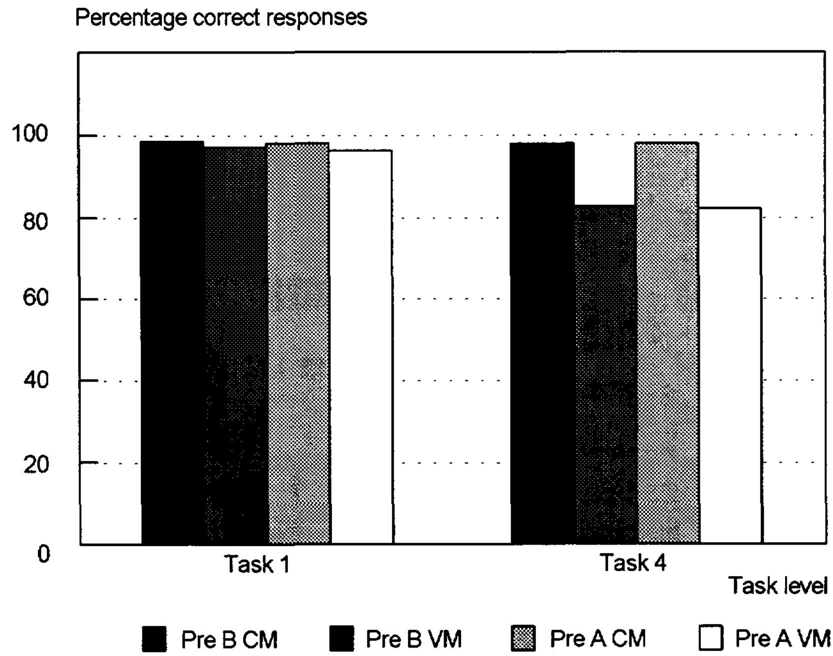

Figure 5.5b: *Quality of Performance, unpaced condition*

The performance results show that the manipulation of the experimental factors 'memory load', and 'processing complexity' was effective. An increase in memory load resulted in faster Reaction Times but only for the Varied Mapping condition. It also became clear that the change in psycho-physiological state did not affect performance in terms of Reaction Times; the subjects maintain their performance level. Moreover these results demonstrate that when subjects have control over the speed of the task they may change their strategy by taking more time for complex tasks in order to achieve a higher level of accuracy.

Heart rate variability

Figures 5.6[a] en 5.6[b] present the results of the heart rate variability index.

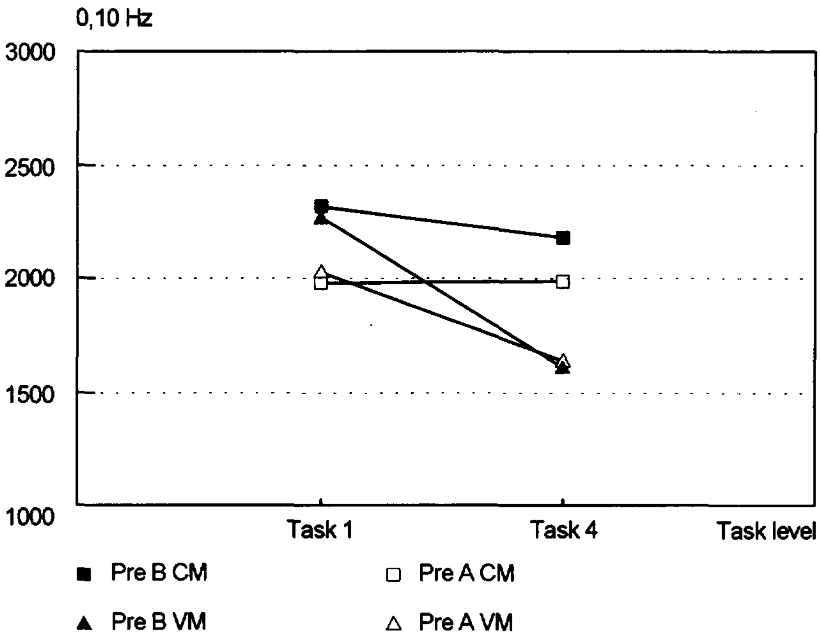

Figure 5.6a: *Heart rate variability in machine paced condition*

Figure 5.6[a] (machine paced condition) shows that a decrease in heart rate variability can only be observed in the Varied Mapping condition for Task 4. In Figure 5.6[b] (unpaced condition) it is again shown that Task 4 in the Varied Mapping condition results in a decrease in heart rate variability. Moreover we can see that Task 4 in the Consistent Mapping condition, following a pre-treatment B condition (simulating a free day), shows an increase in heart rate variability. Or perhaps it is better to say that conversely subjects participating in Task 1 under the same conditions showed a remarkable decrease in heart rate variability, especially when we take into account the fact that the other three conditions have more or less the same values for this task. When both figures (a and b) are compared we can see that the values of the heart rate variability in the 'unpaced' condition are lower than in the 'machine paced' condition. We have already seen that the Reaction Times for Task 1 in the 'unpaced' condition are also lower. The implication of these findings will be discussed later.

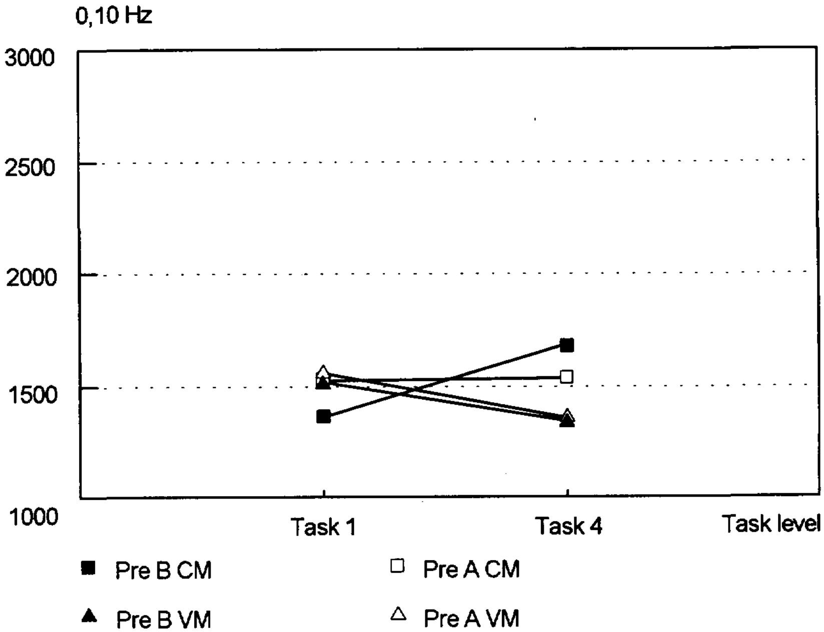

Figure 5.6b: *Heart rate variability in unpaced condition*

Table 5.2 presents the results of the analyses of variance of the heart rate variability (HRV) parameter.

Table 5.2: *Analyses of Variance: HRV, repeated measurement design, full model (n=30: two subjects with missing data)*

factor	F	df	p
Constant (=model)	96.5	(1,29)	<<.01*
Pacing	2.7	"	.11
Pre-treatment	.1	"	.76
Pacing by Pre-treatment	1.52	"	.23
Mapping	2.7	"	.17
Pacing by Mapping	.69	"	.41
Memory load	.84	"	.37
Pacing by Memory load	1.38	"	.25
pre-tr. by Mapping	.60	"	.45
Pacing by pre-tr. by mapping	.00	"	.99
pre-tr. by memory load	.03	"	.86
Pacing by pre-tr. by memload	2.07	"	.16
Mapping by memory load	10.76	"	<<.01*
Pacing by mapping by memload	.13	"	.72
Pre-tr. by mapping by memload	.36	"	.55
Pac. by pre-tr. by mapping by memload	.06	"	.81

The significant results are marked with an *

Although there seems to be some influence from the factors 'pacing' and 'pre-treatment' in the 'machine paced' condition, analysis of variance results showed only one significant effect: a two-way interaction effect with the factors 'processing complexity' by 'memory load' (F(1,26)=10.8 p=.003).

It has already been mentioned in the description of the results of the pre-treatment situations that the mean heart rate inter beat interval was significantly influenced by the factor 'pre-treatment'. Analysis of variance also showed significant effects for the factor 'processing complexity' (F(1,26)=5.2 p=.03) and a two-way interaction for 'processing complexity' by 'task load' (F(1,26)=6.8 p=.02). Table 5.3 presents the results of analyses of variance of the heart rate (HR).

Table 5.3: *Analyses of Variance: Heart rate, repeated measurement design, full model (n=30: two subjects with missing data)*

factor	F	df	p
Constant (=model)	1455.32	(1,29)	<<.01*
Pacing	.17	"	.68
Pre-treatment	15.0	"	<<.01*
Pacing by Pre-treatment	.01	"	.91
Mapping	5.2	"	.03*
Pacing by Mapping	.11	"	.75
Memory load	2.99	"	.10
Pacing by Memory load	1.70	"	.20
pre-tr. by Mapping	.13	"	.72
Pacing by pre-tr. by mapping	.02	"	.90
pre-tr. by memory load	1.39	"	.25
Pacing by pre-tr. by memload	.07	"	.80
Mapping by memory load	6.81	"	.02*
Pacing by mapping by memload	.01	"	.94
Pre-tr. by mapping by memload	.27	"	.61
Pac. by pre-tr. by mapping by memload	5.12	"	.03*

The significant results are marked with an *

Another aspect is that the neuroendocrine reactions (Fig. 5.2) showed that the laboratory sessions required activation patterns for the subjects that were opposite to those of the pre-treatment conditions. Those who had been physically active (pre-treatment B) had to deactivate, while those who had been quite passive in the pre-treatment condition (A) had to become active again. This is reflected in the respective decreases and increases in the adrenaline excretion rates during the laboratory session (period 16.00 hours - 17.30 hours).

These physiological results demonstrate again that the manipulation of both factors: 'processing complexity' and 'memory load' was effective, although no main effects of both factors were found for the heart rate variability. It is clear that an increase in task load resulted in a decrease in heart rate variability (i.e. more exertion of mental effort), at least for the Varied Mapping condition. This supports the hypothesis that an increase in task load would not result in an increase in mental effort in the Consistent Mapping condition.
The results also demonstrate that the factor 'pre-treatment' has an effect on heart rate, but not on heart rate variability.

Rating Scale Mental Effort

Figure 5.7[a] and [b] presents the results of the scores obtained with the Rating Scale Mental Effort.

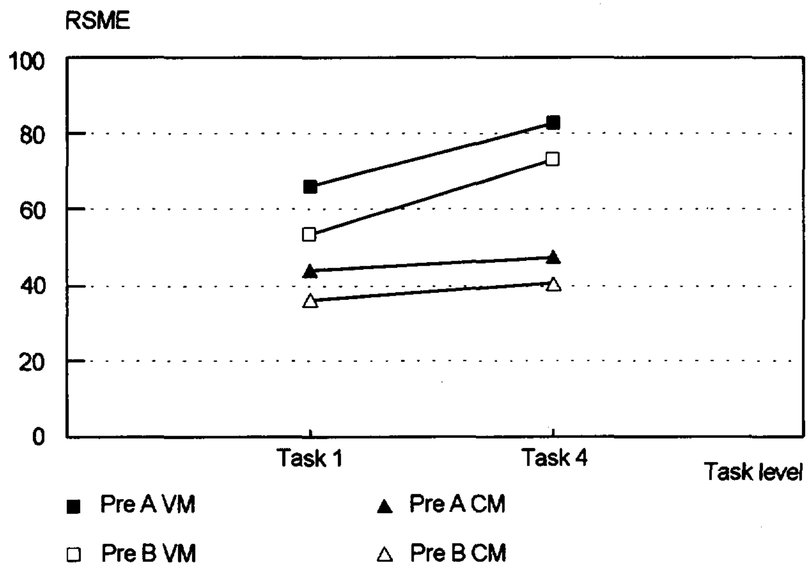

Figure 5.7a: *RSME score in machine paced condition*

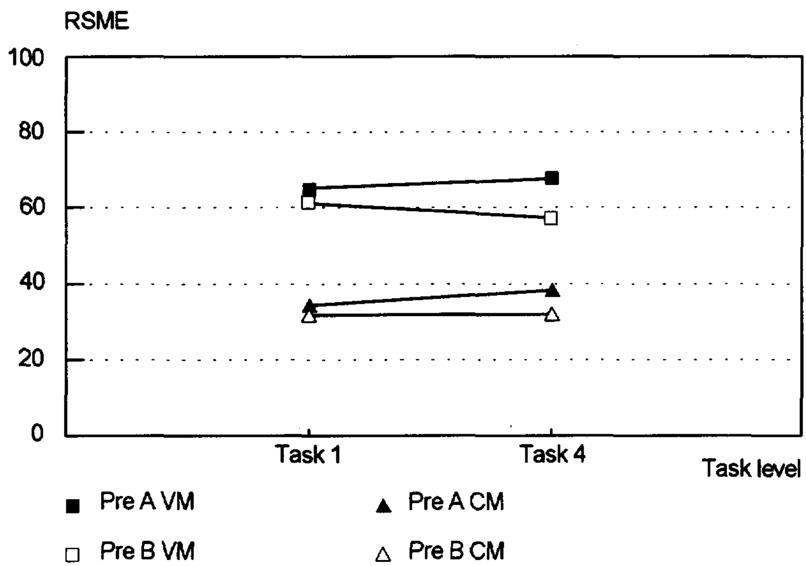

Figure 5.7b: *RSME score in unpaced condition*

Figure 5.7[a] shows that in the 'machine paced' condition an increase in memory load (task 1 versus Task 4) is accompanied by an increase in the RSME score, especially in the Varied Mapping condition. Furthermore the scores in the Varied Mapping condition appear to be systematically higher than the scores in the Consistent Mapping condition indicating that 'controlled' processing requires more effort than 'automatic' processing of information.
Moreover we can see that the RSME scores are systematically higher after the subjects have been exposed to a pre-treatment A condition (simulating a working day).

From Figure 5.7[b] it becomes clear that in the 'unpaced' condition the increase in RSME scores that should have accompanied the increase in memory load disappeared. The effects of the factors 'processing complexity' and 'pre-treatment' remain visible, although the latter effect is more apparent for Task 4.

Table 5.4 presents the results of the analysis of variance of the Rating Scale Mental Effort.

Table 5.4: *Analyses of Variance table: repeated measurement design complete model (n=31: one subject with missing data)*

factor	**F**	**df**	**p**
constant (=model)	567.3	(1,30)	<<.01 *
Pacing	2.2	"	.15
Pre-treatment	22.9	"	<<.01 *
Pacing by pre-treatment	.7	"	.41
Mapping	161.4	"	<<.01 *
Pacing by Mapping	.4	"	.53
Memory load	7.6	"	.01 *
Pacing by memory load	9.7	"	<.01 *
Pre-tr. by mapping	.5	"	.48
Pacing by pre-tr. by mapping	.08	"	.78
Pre-tr. by memory load	.5	"	.47
Pacing by pre-tr. by memload	3.9	"	.06
Mapping by memory load	2.4	"	.13
Pacing by Mapping by memload	9.1	"	<.01 *
Pre-tr. by mapping by memload	.1	"	.78
Pacing by pre-tr. by mapping by memload	.8	"	.37

The significant results are marked with an *

This table confirms that there are significant effects for the factors 'pre-treatment', 'processing complexity', 'memory load', and for a two-way interaction 'pacing by memory load' and a three-way interaction 'pacing by processing complexity by memory load'.

The RSME scores in the 'machine paced' condition resemble the results of the Reaction Times in this condition but there is one difference: the Reaction Times are not influenced by the factor 'pre-treatment' while the RSME scores are significantly influenced.
On the other hand in the 'unpaced' condition the RSME scores are remarkably different from the Reaction Times in this condition. In the 'unpaced' condition the subjects take considerably more time to do the difficult task (Task 4 in the Varied Mapping condition); they also had better results in terms of percentage of correct answers but they reported no increase in effort as measured by the Rating Scale Mental Effort. When the 'machine paced' and the 'unpaced' conditions are analyzed separately it appears that in the 'machine paced' condition a significant effect of the factor 'memory load' is found ($F(1,15)=41.8$ $p<<.001$), while this effect is absent in the 'unpaced' ondition ($F(1,15)=.06$ $p=.80$).
Another interesting finding is that the RSME scores in the Consistent Mapping condition appeared to be influenced by the factor pre-treatment, although it had been expected that 'automatic' processing would not be affected by changes in the psycho-physiological state.

In conclusion we might say that the Rating Scale Mental Effort is sensitive to changes in the psycho-physiological state of the subject, increases in memory load and to differences in processing complexity.

The QRST task

The QRST task was presented at the end of each experiment session. As has already been stated this task is a 'double-task' and so it is therefore presumed to be more difficult than task 4. Furthermore there is no differentiation between Consistent Mapping or Varied Mapping. The factor 'processing complexity' is only relevant for the tasks (1 and 4) in the preceding period. Since one experimental session had only Consistent Mapping conditions and the other session had only Varied Mapping conditions this preceding period might be regarded as some kind of pre-treatment preparation for the QRST task.
A second distinction is that the QRST task had less trials (40) than the other tasks (170) and was therefore much shorter (duration about three minutes). Figure 5.8[a] and [b] presents the results of the RSME scores for the QRST task.

Figure 5.8[a] presents the results of the RSME scores for the QRST task in the 'machine paced' condition. It is clear that the points indicating the values for the QRST task occur in pairs. It is worth noting that the discriminating factor appears to be the factor 'processing complexity'. The factor 'pre-treatment' does not result in a significant effect in the 'machine paced' condition while the factor 'processing complexity' shows a marginally significant effect $F(1,14)=4.5$ $p=.053$).

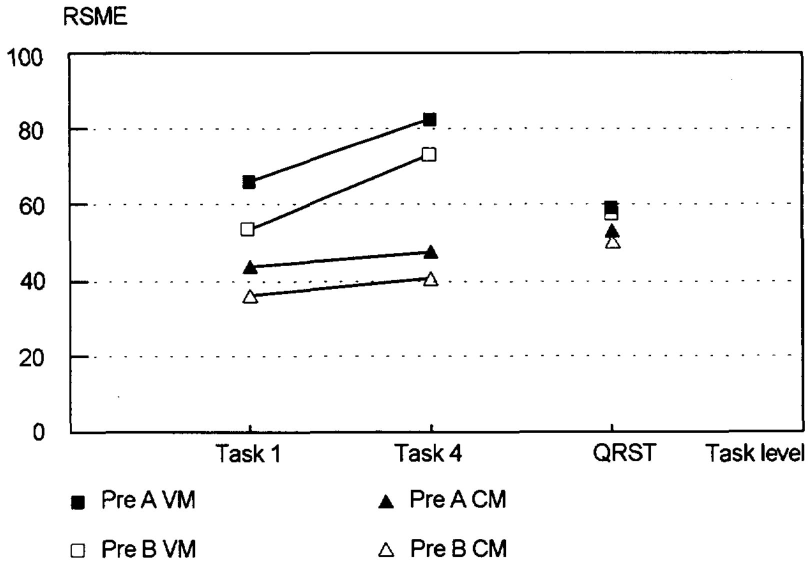

Figure 5.8a: *RSME score for QRST-task in machine paced condition*

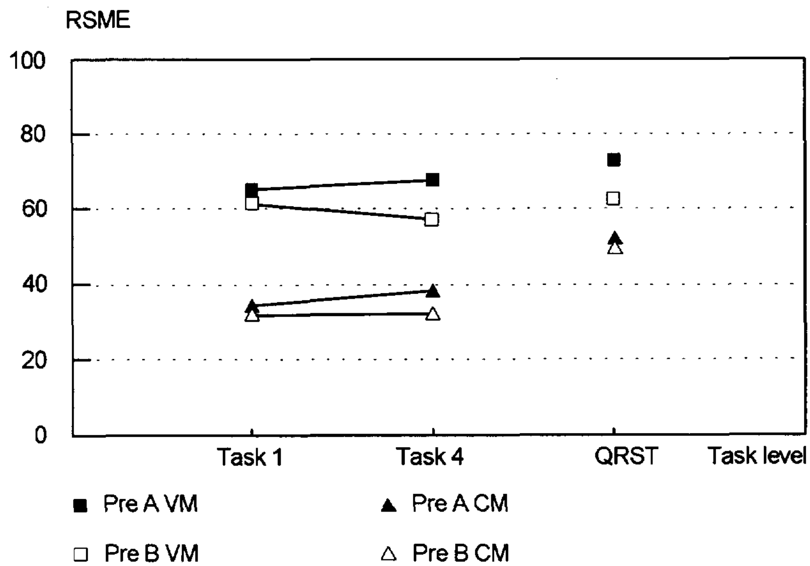

Figure 5.8b: *RSME score for QRST-task in unpaced condition*

In the 'unpaced' condition we find the same pattern (Fig. 5.8[b]). It is remarkable that the original pre-treatment factor did not result in a significant effect, while the factor 'processing complexity' did once again produce marginally significant effects (F(1,14)=4,6 p=.051). The Varied Mapping condition appeals enormously to the information processing system, while the Consistent Mapping condition hardly does at all. It is very likely that 1 hour of intensively performing tasks in the Varied Mapping condition makes a person more tired than performing tasks in the Consistent Mapping condition. Moreover the information processing capacity is very specifically addressed in the Varied Mapping condition, while the original pre-treatment was of a more general nature and was meant to bring about a change in the subject's general psycho-physiological state.
This might explain why the original pre-treatment factor did not show a significant effect while the factor 'processing complexity' did.
Another aspect is that the RSME scores on the QRST task are lower than the values of Task 4 in the Varied Mapping condition. While the QRST task is assumed to be more difficult than Task 4 the QRST task is a 'double-task'. An explanation can be found in the fact that the QRST task took about 3 minutes, while Task 4 took approximately 10 minutes. This indicates that the duration of the task (or one should say 'time on task') may also influence the subjective ratings.

5.6 Discussion

From the summary of the results it is clear that the various pre-treatments resulted in distinguishable psycho-physiological states at the beginning of the actual experimental session in the laboratory. After the pre-treatment A condition, in which a working day was simulated and the subjects were at a rather low level of activation by the end of the day, they were required to activate themselves again in order to carry out the experimental tasks in the laboratory. Figure 5.2 illustrates that there was an increase in the adrenaline excretion rate in the period between 16.00 and 17.30 hours. The results of the mood state measurements confirmed that subjects felt less activated and more tense after pre-treatment A. Meijman (1991) calls such a psycho-physiological state fatigue and according to him this indicates resistance to any further activity because of the higher 'costs' that accompany further activity.

In this respect it is interesting to note that in this experiment subjects report that they have to invest more effort to carry out the experimental tasks in the laboratory session after pre-treatment A has been administered. This means that there are indeed additional 'costs' involved. On the other hand it appears that the results of work behaviour, in terms of reaction times and errors, are not influenced by the pre-treatment conditions.

This implies that comparing both parameters might give an indication of the efficiency of work behaviour during the experimental sessions.

5.6.1 RSME in relation to performance and heart rate variability

RSME scores related to performance

Comparing the 'machine paced' condition with the 'unpaced' condition leads to some interesting findings. In particular it is interesting to note that in the 'unpaced' condition the RSME scores appear not to be influenced by the factor task load while this factor has a significant influence in the 'machine paced' condition. On the other hand it appears that the reaction times show a far greater increase as a result of the increase in task load in the 'unpaced' condition.

In Fig. 5.9 the RSME scores for both, Task 1 and Task 4, are plotted against the reaction times for the 'machine paced' condition.

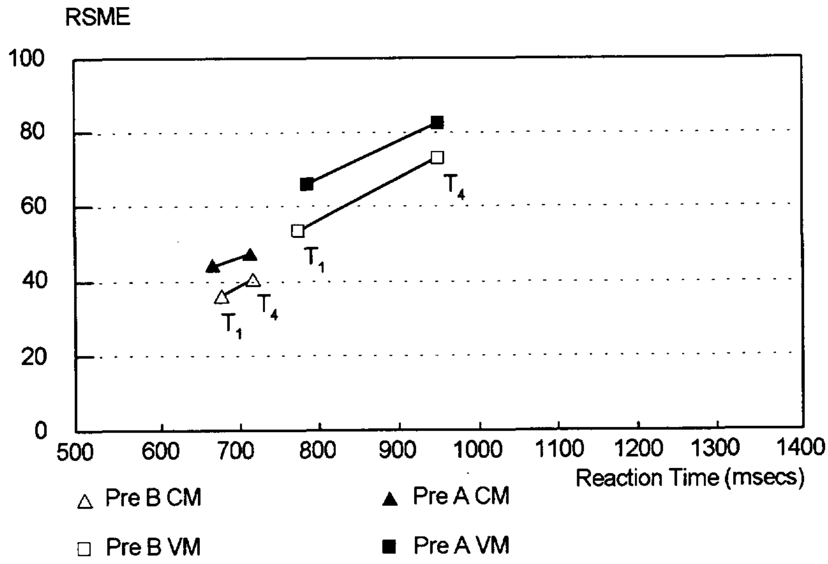

Figure 5.9: *RSME score (y-axis) and Reaction Time (x-axis) in machine paced condition*

This figure clearly demonstrates that in the 'consistent mapping' condition there is hardly any increase in subjective effort and time required for the task as a result of an increase in task load. At the same time we see in the 'varied mapping' condition a significant increase in the time required to perform the task and also in the reported level of effort exertion due to an increase in task load. This finding supports the assumption that the information processing mode was manipulated by introducing a consistent and varied mapping condition. This figure also shows that the pre-treatment conditions did have an effect on the subjective experiences of the subjects but did not influence their performance.

In Figure 5.10 the RSME scores for both tasks are plotted against the Reaction times for the 'unpaced' condition.

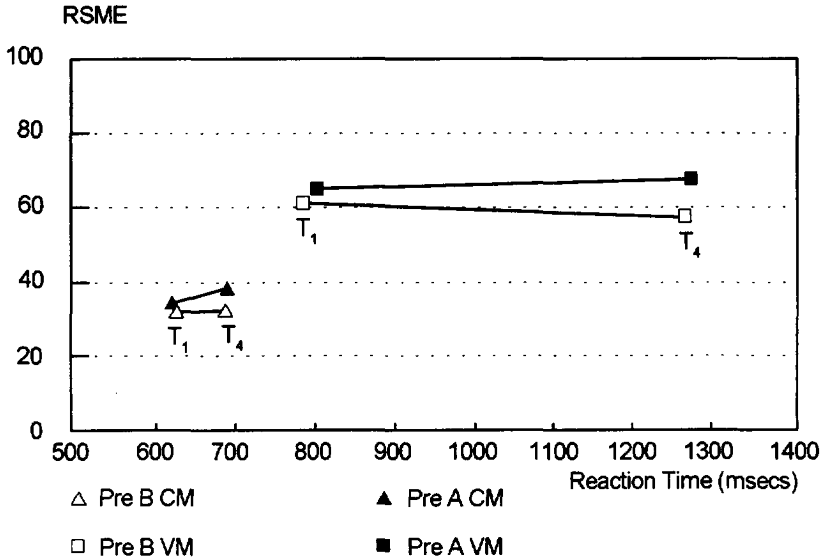

Figure 5.10: *RSME score (y-axis) and Reaction Time (x-axis) in unpaced condition*

The main difference between Fig. 5.9 and Fig. 5.10 is in the 'varied mapping' condition. The reported increase in effort exertion had disappeared and the amount of time required to perform the task was much longer than in the 'machine paced' condition. Though the task had become more difficult the subjects did not report having invested more effort. The heart rate variability also showed no significant effect as a result of increase in task load in this condition.

In this condition the subjects controlled the execution speed of the task themselves, new stimuli appeared after the subject had responded to the previous stimulus. This means that the subject takes more time to react when the task becomes more difficult. This might be interpreted as saying that the subjects used a different strategy in this condition. This strategy not only required less effort but it also resulted in a qualitatively better performance. As may be remembered from Figure 5.5 the subjects made about 10 % less errors in the 'unpaced' condition.

In the 'machine paced' condition the time allowed for response is fixed and probably not enough for Task 4. Subjects therefore have to respond very quickly with the risk that they will make more errors. Apparently they employ a more risk taking strategy when they are working under pressure.
The same phenomenon applies to the QRST task. The reaction times are much larger in the 'unpaced' condition (about 1100 msecs. versus about 900 msecs. in the 'machine paced' condition) but the subjects also give approximately 10 % more correct answers.

In conclusion one might say that when subjects have control over the task they can choose their own strategy: by taking more time they can improve the quality of their performance and reduce the amount of effort that is needed. This demonstrates that people are capable of regulating their effort expenditure.
Furthermore it is clear that the RSME scale is sensitive to variations in a person's psycho-physiological state and in task load. In this experiment task load was varied by having different memory load levels in combination with the factor processing complexity.

RSME-scores and Heart Rate Variability

Since both the Rating Scale Mental Effort and the heart rate variability index (0.10 Hz component) are believed to be indicators of mental effort, it is interesting to compare the measurement results of both parameters.
Looking at the plots of the heart rate variability values (Fig 5.6) and the RSME scores (Fig 5.7) some remarkable differences emerge, in particular in the 'unpaced' condition. The 'consistent mapping' condition after pre-treatment B shows a deviant pattern. Task 1 shows a decrease in variability in the heart rate (indicating an increase in effort exertion) while Task 4 shows an increase in variability. In this condition Task 1 may be regarded as the most easy task of the entire experiment so it is unlikely that subjects invested a lot of effort in this task because of its complexity. This conclusion is supported by the results of the reaction times. Therefore this result could be an artefact.

Although it is clear that the scores on the RSME show significant effects as a result of the factors 'pre-treatment', 'processing mode' and 'memory load' and that there is a two-way interaction 'pacing' by 'memory load', the heart rate variability scores are only significantly influenced by the two-way interaction 'processing mode' by 'memory load'. Moreover no significant correlation could be calculated between the RSME scores and the heart rate variability values. This seems to contradict the suggestion that both parameters are indicators of mental effort.
The explanation could be that the Rating Scale Mental Effort is just more sensitive to the changes that have been brought about. Since the respective

factors have significant effects on other parameters (catecholamines, heart rate, reaction time, performance quality) which support what we postulated and heart rate variability only shows an effect in task 4 of the varied mapping conditions (the most difficult conditions) the conclusion that the RSME is a more sensitive instrument than the heart rate variability index seems plausible. This backs up findings reported by Casali et al. (1983) and Gopher et al. (1984) (see Chapter 4).
However, the findings of Jorna (1985), who examined divers in underwater situations suggest another explanation. Jorna reported that while divers are performing cognitive tasks under water there are no effects on heart rate variability as the results of differences in task load and the variance in the heart rate variability spectrum are very low in this condition. However when the divers performed the same tasks out of the water, not wearing their diving-suits, the general level of variance in heart rate variability increased and significant effects of differences in task load were found. This suggests that heart rate variability may be related to the general level of arousal and may thus be sensitive to specific changes in the psycho-physiological state.

This suggestion is supported by the findings of Veldman (1992). Veldman studied the effect of noise on people in experimental conditions. In a laboratory setting subjects performed the same kind of experimental tasks (with slight modifications) as those described in this and the previous chapter. As dependent measures he used performance indicators (reaction times), physiological responses (heart rate and blood pressure) and subjective reports (including the RSME). Veldman reports a discrepancy between the amount of effort reported subjectively and the cardiovascular indices. The Rating Scale Mental Effort discriminates between the various task load levels presented to the subjects. The increase in mental load, indicated by increases in reaction times and the number of errors made, is also experienced as an increase by the subjects. The cardiovascular indices appear not to be influenced by variations in task load. On the other hand it appeared that when the tasks have to be performed in noisy conditions subjects do not report more subjective effort. However, the cardiovascular indices (heart rate and blood pressure) increased because of the noise conditions. Noise conditions may well raise the general arousal level of subjects as the cardiovascular system would tend to indicate. However, this does not mean that the information processing function is affected.

In our experiment the heart rate variability measurement results showed a general decrease in variability where the 'unpaced' condition was compared to the 'machine paced' condition. When we assume that the 'unpaced' condition requires a higher level of arousal, because subjects have more control over the task these results do correspond with Jorna's and Veldman's findings. This may also explain why in the earlier reported finding subjects show a decrease in heart rate variability in Task 1 in consistent mapping

conditions (the unpaced condition), which was originally regarded as an artefact. It had already been observed that reaction times were remarkably small in this condition (Fig. 5.4[b]). Together with the decrease in heart rate variability in this condition this finding may be interpreted as an indication of an increase in the arousal level, because subjects are 'excited' about the 'game' and try to react as quickly as possible with this very simple task.
This general decrease in variability may also camouflage the effects of the factors 'memory load' and 'processing complexity' in the 'unpaced' condition, and thus prevent these factors from having significant effects in the general model of analysis of variance.

This leads to the conclusion that the heart rate variability parameter should primarily be regarded as an indicator of the degree to which people feel 'aroused' or 'tense'. A low level of variability may reflect a high level of arousal: this may mask task load effects.
This also makes it clear that the heart rate variability parameter cannot be used as a measure of the concurrent validity of the RSME.

5.6.2 Theoretical implications

In order to be able to draw conclusions about the construct validity of the RSME we first need some indications about the validity of the integrated models of effort (Sanders, 1983; Mulder, 1986) which were taken as the basis for our experiment.
Table 5.5 presents an overview of the most significant results and is helpful to formulate a conclusion.

Table 5.5: *Effects of factors on various parameters.*

parameters	**Factors**		
	Pre-treatment	processing complexity	task load
Adrenaline	+	-	-
Noradrenaline	+	-	-
Reaction time	-	+	+
HRV	-	+	+
RSME	+	+	+

As has already been mentioned both the factors 'pre-treatment' and 'mode of information processing' had significant effects on various parameters indicating that the respective manipulations had been successful. The

reaction times are influenced by the factors 'task load' and 'processing complexity' but not by the factor 'pre-treatment'. Heart rate variability is influenced by 'task load' and 'processing complexity', but not by the factor 'pre-treatment'. The RSME scores are influenced by 'task load', 'processing complexity' and by 'pre-treatment' and all the effects were as had been predicted in the hypotheses.
Moreover it appeared that only the subjective ratings of effort (RSME scores) were sensitive to both aspects of effort that have been mentioned: effort as 'state regulation' and effort as 'executive resource control'. This may lead to the conclusion that from a theoretical point of view the Rating Scale Mental Effort indeed measures the concept of mental effort, as conceptualized in recent models (cf. Mulder, 1986).

Figure 5.7 shows that in the 'consistent mapping' condition (automatic processing) an effect of the 'pre-treatment' factor was found as well. This contradicted our expectations in so far we had assumed that both the aspects of effort would be independent of each other. In such a case we would expect that with automatic processing changes in task load-level and/or in psycho-physiological state would not have a significant effect on the effort-level.
This finding indicates that both aspects of effort cannot be completely separated from each other. When no attention is demanded for the processing of information it still may be that the subject has to change his state. Moreover it may be clear that even in the situation of Consistent Mapping some activity is required on the part of the subject. Although it is assumed that in the automatic processing mode some specific processing resources are hardly addressed, it is plausible to assume that the general resources have been addressed (multiple resources approach, Navon and Gopher, 1984). The subjects at least have to stay awake; they have to respond to the tasks presented etc. Some general level of activity is required: even during very simple tasks subjects have to maintain a certain 'state' of activation. After pre-treatment A condition the activation level was lowered (see also section 5.5.1), consequently the subject had to adjust his state to the level of activation required for the experimental task. This means that also in the consistent mapping condition subjects have to invest effort. The 'state' of the central processor is influenced by the general psycho-physiological state of the subject.

These findings may be regarded as empirical support for the conceptualization of effort as formulated in the integrated model of mental effort (Mulder, 1986), as described in Chapter 3.

5.6.3 Conclusions regarding the RSME

The results of the experiment concerning the Rating Scale Mental Effort confirm the results of earlier studies (see Chapter 4) and correspond with predictions that were formulated on theoretical grounds. Assuming that these models of mental effort are valid we may accept the conclusion that the Rating Scale Mental Effort does indeed measure the construct 'effort'. Contrary to what is usually said about subjective methods, i.e. that they only give global indications of people's psycho-physiological state (Hockey et al., 1986), it appears that the RSME scores not only reflect changes in states but also differences in modalities of information processing. What can particularly be demonstrated with the Rating Scale Mental Effort is the difference between consistent mapping conditions and varied mapping conditions and the increase in task load. This is not so remarkable when one thinks that the difference between consistent mapping and varied mapping is also externally observable in the task-demands. Subjects notice of course that they have to respond to the same letter(s) each time although they are not aware of the theoretical notions behind these differences. This raises the question: what do subjects refer to when they respond to the RSME, to an externally observable change in the task demands or, to some internal cues or reference-points related to the information processing system?

The results of this experiment are insufficient to give a complete answer to this question. However, when the results of the 'unpaced' condition are taken into account, it is plausible to think that it was not only externally observable changes which resulted in distinguishable effort-scores. The factor 'pacing' did not in itself lead to effects that were significant; only when combined with the factor 'task load' were effects found that were significant. The subjects in the 'unpaced' condition were not the same as in the 'machine paced' condition (pacing was a between-subjects factor) yet the effort-scores were comparable. In the 'unpaced' condition the increase in task load did not result in an increase in effort-score which would have been logical if only externally observable cues had accounted for increases in the effort-scores. The sensitivity to changes in psycho-physiological state also suggests that people use some internal references for the availability of resources when they rate their effort exertion.
To some extent this question may be regarded as purely academic. The converging evidence on performance parameters and physiological indicators on the validity of the scale and the fact that the scale-scores conform with theoretical models are, from a pragmatic point of view, justification enough for employing the scale.

Various analogies can be thought of for mental effort; the energy metaphor (mentioned in Chapter 3) is just one of them but it is a useful one. An extra complication is that one should take into account that the size of the 'fuel-reservoir' (i.e. performance potential) may vary from time to time.

In Chapter 3 it was stated that the amount of effort that is required depends on the task demands on the one hand and the (remaining) performance potential of the worker on the other hand. In the 'machine paced' condition 'time constraints' are part of the task demands; this aspect is absent in the 'unpaced' condition. The (remaining) performance potential of the subjects may be assumed to be comparable in both conditions. Yet the subjects in the machine paced condition report that they have to invest more effort, particularly with the more difficult tasks. It would seem that the time constraints rather than the amount of information needing to be processed constitutes the real difficulty surrounding the demands of these tasks (cf. Kahneman, 1973).

This brings us to another conclusion with regard to the Rating Scale Mental Effort. The fact that subjects do not rate Task 4 in the 'unpaced' condition as effortful; as the task with an equal information load in the 'machine paced' condition may be interpreted as an indication that the scores of the Rating Scale Mental Effort can be conceived as expressions of *subjective, or psychological costs* that are inherently associated with the execution of the task.

5.6.4 Psychological costs

This experiment clearly demonstrates the importance of the factor 'time' in relation to work behaviour. The aspect of 'time-constraints' has already been referred to above. The results of the QRST task in comparison with Task 4 (Varied Mapping condition) make clear that 'time-on-task' also influences the psychological cost level. On the one hand it is evident that the more time the task takes, the longer the period will be during which effort has to be exerted thus leading to an increase in costs. However, on the other hand it has also become clear that subjects may take more time to complete the task in order to reduce their level of effort expenditure, in other words, they may make a trade-off between 'time' and 'effort'. Of course it is also possible to take more time in order to increase the level of performance with the same (or less) amount of invested effort. If workers have 'decision latitude' they can choose which strategy they will follow: an 'effort-consuming' strategy or a 'time-consuming' strategy. Examples of both strategies can be found in physical and mental work. For instance, when one has to transport a lot of sand, one has the choice between taking as much sand in the wheelbarrow as possible (effort consuming, because of the heavy load) or walking the route a few more times with a wheelbarrow that is only half-full (more time-consuming but requires less effort). Similar examples may be found in mental work, for instance with calculating: one may calculate something in one's head (more complex, but quick) or one can work it out on paper and then count up the units (more time-consuming, but less complex). These examples illustrate that the relation between time and effort cannot simply be expressed as a linear equation. Therefore both things, time and effort, should preferably be envisaged as two dimensions of the concept of psychological costs.

There is another aspect relevant to the time factor as well. As time passes a worker's psycho-physiological state may change. The worker's psycho-physiological state may influence his decision to choose a particular strategy. For instance, when a worker is tired; when he does not have much energy left, he will be 'forced' to choose a more time-consuming strategy. However, in general a person's choice of a particular strategy will depend on individual preference, knowledge and various task constraints that are relevant, like the amount of time available.
The amount of time that is available is sometimes regarded as one of the most important characteristics of the task because this is what determines whether there is time-pressure or not. According to Kahneman (1973) time-pressure determines whether a task is difficult or not. Again this illustrates the importance of the factor 'time' in the estimation of psychological costs.

This suggests that people estimate their 'costs' according to two criteria: 'time' and 'effort'. This conclusion is relevant to the study that will be described in the next chapter. When people choose a particular strategy they apparently take into account their estimation of the time that is available, the amount of time that will be required and the complexity of the task in terms of required effort.
Although these dimensions are clearly not independent, the complexity of their inter-relatedness justifies treating them as seperate dimensions. The 'time-dimension' can be conceived as being related to making optimal or efficient choices between various behaviour alternatives (or strategies). This dimension can be labelled the 'strategy dimension'.
The 'effort-dimension' can be conceived as being primarily related to the complexity or difficulty of the task demands in relation to the actual performance potential. This dimension may be labelled the 'processing dimension' (or 'energy dimension').

5.7 Psychometric aspects of the RSME

In Chapter 4 several criteria were mentioned with which instruments should comply. These criteria concern the traditional aspects of reliability and various aspects of validity and some pragmatic aspects as well like intrusiveness and applicability.
These aspects will be discussed in relation to the RSME in this section.

5.7.1 Reliability

An instrument's reliability is its capacity to produce the same results when a given measurement procedure is repeated after a certain time or under different conditions (in theory on test construction or measurement theory also referred to as 'test-retest reliability' or 'reproducability'). Where the RSME is concerned one has to allow for the fact that the construct measured is not very stable. Effort expenditure is related to a person's state and this may change during the course of time. Consequently reliability indices may be somewhat lower than those of instruments that are presumed to measure 'traits'.

In the study described in Chapter 4 the second working day may be regarded as a replication of the first working day. The RSME scores of the laboratory study done on the first working day correlated r = .81 with those of the second working day. The RSME scores gathered from the real work situation of the bus drivers on the first working day correlated r = .71 with those of the second working day. In real life situations it is not possible to control all conditions which is why a somewhat lower correlation between the scores of both working days could be expected. Although not extremely high these results are satisfying.

The experiment described in this chapter provides an opportunity for calculating the reliability of the RSME. Since each task-level was administered twice; before and after the break, we may use the correlation of the corresponding RSME scores as an indication of the reliability of the RSME. The RSME ratings of the task levels correlated (product moment correlation) r_{pm} = .88 for Task 4 in the pre-treatment A with Varied Mapping task condition and r_{pm} = .58 in the pre-treatment A with Consistent Mapping task condition (for all correlations 1-tailed $p < .001$) and the average correlation for RSME (including all task levels and conditions is r = .78 ($p < .001$).
The same procedure was followed for calculating the correlations for the heart rate variability scores (mid frequency band: 0.10 Hz). These correlations ranged from r_{pm} = .78 for task level 1 in the pre-treatment B with Consistent Mapping condition and r_{pm} = .72 for Task 4 in the pre-treatment B with Varied Mapping Condition (all correlations 1-tailed $p < .001$), and the average correlation (for all task levels and conditions) was r = .76 ($p < .001$).

The above mentioned correlations were not very high even though the experimental conditions were completely controlled and did not change. However, one should remember that in the second half of the experiment session the subjects had already been performing the tasks for some time while in the first half of the experiment session they had just started. Their performance potential may therefore have deteriorated a little which could lead to structurally higher RSME scores in the second half of the experimental session. This means that in correlating of r = .78 one may be underestimating the actual reliability of the RSME.

The results regarding the RSME as reported by Veldman (1992) confirm the above mentioned findings on the reliability of the RSME scores. Veldman also reports significant increases in RSME scores as a result of an increase in task load (Veldman, 1992; page 58-59).
The conclusion of this section is that the RSME is a reliable instrument.

5.7.2 Validity

As has already been mentioned in Chapter 4 various aspects of validity can usually be distinguished. As for the validity of the RSME we are mainly interested in seeing whether the RSME does indeed provide us with information about the construct of effort. The 'construct validity' of the instrument is therefore the main topic of this section.

The amount of effort that has to be invested depends on:

- the task demands;
- the available 'performance potential';
- the duration of the activity (time-on-task), an aspect that was highlighted in the preceding experiment.

The studies described in this chapter and the last chapter demonstrate that RSME scores adequately reflect variations in task load in both the experimental tasks done in the laboratory and in tasks of daily life. Furthermore it appeared that the RSME distinguished between the worker's various levels of performance potential and was influenced by fluctuations in task duration.

Additional information can be gathered from other experiments where the RSME has been applied. One such experiment had to do with the tasks of nautical officers responsible for navigating a ship. Their tasks were simulated on a ship's-simulator (see also Perdok, 1984). There are usually several persons present on the bridge of a sea-going vessel at any one time and each of these people has his own specific task be it connected with navigation, communication or controlling the ship's engine-room (speed regulation). When a ship is equipped with advanced technological instruments like sophisticated automatic pilot systems it is possible to make do with fewer crew members on the bridge. This simulation study was about 'one-man manning' situations on the bridges of sea-going vessels. The subjects were given various course permutations along which they were expected to sail while remaining responsible for the navigation, communication (coast-ship) and engine-room (speed control). These appeared to be the most relevant task-aspects. The scenarios variations were: differences in weather conditions and speed and in the courses of oncoming shipping.

The subjects of this study were eight nautical officers (with a mean age of about 26 years) and each had at least three years of nautical experience. Half of this group had recently finished a training with the new 'one-man manning' equipment, a so-called 'integrated training', hereafter referred to as the Seaman Integrated Training (SIT). The others had been trained in the traditional way, that is to say, in the navigational aspects of their work. Their training is hereafter referred to as the Seaman Not-adequate Training (SNT) because it did not adequately prepare them for one-man-manning situations.

While the subjects executed their task they filled in the RSME and after completing the task they filled in the 'Schaal Ervaren Belasting' (Scale Experienced Load - SEB; Meijman, 1991). This SEB consists of a questionnaire with items pertaining to symptoms which indicate various degrees of fatigue. With the RSME the subjects had to rate the set of tasks as a whole and separately as well. Their performance was also registered. Several methods were used for determining the quality of the performance (Perdok, 1984). Amongst others things the passing distance between the ship being monitored and other ships served as a measure of performance quality. In navigation circles a passing distance of less than one nautical mile is regarded as dangerous.

During the experiment the researchers discovered that the subjects had different strategies for avoiding collisions than they had expected them to have. So it was that the theoretical gradations of difficulty that had been manipulated in the various scenarios lost their validity. The average scores for the various scenarios are presented in Table 5.6.

Table 5.6.: *Comparison of RSME scores of two groups of nautical officers with respect to the various tasks and the set of tasks as a whole.*

VARIABLES	Average score SIT-group	Average score SNT-group	F	sign. (=.05)
$RSME_{Tasks}$	31	48	5.2	.03
$RSME_{Navigation}$	28	44	4.8	.04
$RSME_{Engine\text{-}room}$	27	39	2.9	.09
$RSME_{Communication}$	28	36	1.9	.17
Miles (passing by)	2.3	1.8	4.4	.05
SEB	8	11	5.5	.03

The SNT group of subjects reported having put significantly more effort into the tasks as a whole and into the task of Navigation. They had higher SEB scores indicating that they had endured heavier workloads and their performance was worse. This means that their qualification discrepancy was not compensated even though they exerted more effort. The difference with respect to the Communication and Engine-room tasks was not significant although the SNT group had higher RSME scores. It may be noted that it is not remarkable that there is no significant difference with respect to

Communication as this can generally easily be time-shared with other tasks. Moreover it appeared that the RSME score correlated with the performance-indicator 'passing-distance' ($r = -.57$; $p<.05$). This means that higher effort-scores are reported when other ships are passed at shorter distances.

This study demonstrates that where one group is more qualified for a task than another group the difference in effort investment between the two groups can be measured with the RSME.

In previous experiments we have mainly dealt with situations where the psycho-physiological state of subjects has been manipulated. The differences in performance potential that these situations created were firstly of a temporary nature and secondly, they negatively affected the performance potential because subjects were more or less fatigued. Consequently, by investing more effort subjects could see to it that their performance did not worsen.

In this study we have been dealing with a more stable aspect of performance potential: the level of qualification. Moreover the performance potential of one group was 'up-graded' by means of training so that these individuals became better suited to the task. This resulted in a lower level of exertion, in better performance and in a lighter workload as indicated by the SEB. This conformed with the theoretical concept of effort.

Another aspect that is of relevance to the validity of the RSME concerns the relation between the RSME and the SEB. In the study described above we have seen that one group of subjects reported that they felt more fatigued after completing the task (as expressed by means of the SEB) than the other group whilst they reported having invested more effort in the executing of the task. The SEB-scores correlate $r = +.55$ with the RSME-scores as reported by Meijman (1991, page 144).

Similar results are reported by Meijman for other studies. One of these studies was a replication of the bus driver's study described in Chapter 4. In this study the bus drivers were examined in three different conditions and they did an early morning shift, a normal shift and a late shift.

The RSME scores for morning, normal and late shifts respectively were: .71; .56; .74 after 4 hours of work, and .70; .77; .66 at the end of the working day and they correlated very positively with the SEB-scores (see Meijman, 1991; page 175).

Another study, relevant to the present study, concerned the workload of driving licence examiners. These examiners were observed in three different working situations: taking 9 examinations per day, taking 10 examinations per day and taking 11 examinations per day (Meijman et al. 1985). At several different points during the day they were asked to rate their effort investment and just before lunch time and at the end of the working day they filled in the SEB. Certain physiological indicators of individual subject's states were also measured: their adrenaline excretion levels, their feelings of activation and degree of irritation. The results of the RSME scores are presented in Figure 5.11.

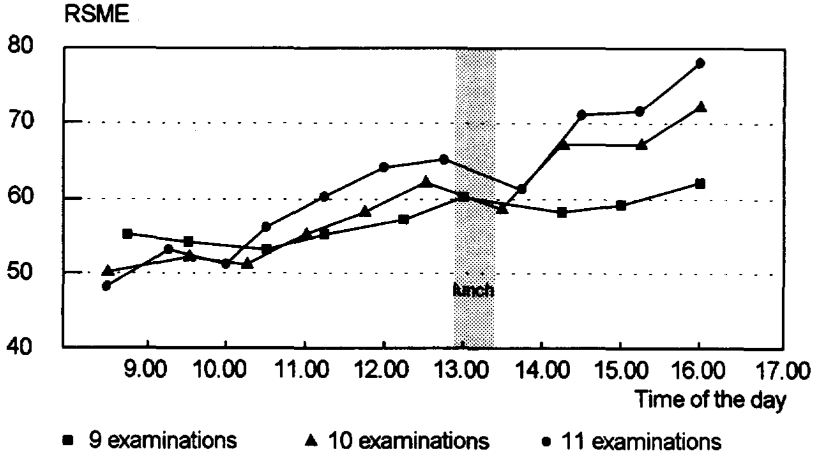

Figure 5.11: *RSME scores of examiners (see text)*

Figure 5.11 shows that at the start of the working day and just after lunch time the three scenarios did not differ. There were however differences at the end of the morning and at the end of the working day. It appeared that statistically significant differences could only be found in the last examinations of each day. The '9 scenario' had significantly lower scores than the '11 scenario'. Subjects also appeared to be more active and less irritated at the end of the day in 'scenario 9' than in the other scenarios as confirmed by the 'state-parameters'.
The results of this study are interesting because they prove that the effects of a higher workload do not show immediately but become apparent after a few hours. Apparently a short rest (a lunch break) gives the subjects enough time to recover and after a few hours we see again that the subjects who are exposed to a higher workload report having to exert more effort in order to perform their task.

These results also conform with the theoretical conception of effort. The positive correlations between the RSME scores and the SEB scores of this study are also reported by Meijman (1991; page 175-177) in his study on the usability of the SEB.

The findings of Wiethoff et al. (1988) published in a study into the effects of sleep deprivation are also interesting with respect to RSME. They report that RSME adequately discriminates between various levels of task load. They

also found that subjects who had been deprived of sleep for one night had to put more effort into the task than control group subjects, especially when the task was very simple. When the task was rather difficult it appeared that the rise in RSME scores was less. According to Wiethoff et al. subjects found it harder to stay awake when they were working on the easy task than when they were working with a difficult task and therefore they had to put in more effort to compensate for their sub-optimal state.
These findings can be added to results of the other studies that have been mentioned.

In conclusion we may state that the construct validity of the RSME seems to be satisfactory, the RSME provides information about the theoretical construct of effort as described in Chapter 4. The instrument appears to be sensitive to changes in task load, to changes in performance potential as a result of prolonged exposure to workload (changes in psycho-physiological state) and to differences in performance potential as a result of variations in qualification level.
With respect to the diagnostic capacity of the RSME it may be noted that the RSME only provides us with one score. The RSME score is one-dimensional which means that from this score one cannot deduce whether the fluctuations in RSME scores are caused by variations in task load or by variations in the subject's state. For diagnostic purposes the RSME should be used in combination with workload or state parameters.

From a pragmatic point of view the RSME appeared to a very useful instrument as well. Administering the scale does not interfere with task execution and it takes the subjects only a few seconds to rate their effort expenditure. Therefore the RSME proves to be a very suitable instrument for field and laboratory studies.

5.8 General Conclusion

With respect to the initial questions of the study described in this chapter we can conclude that the integrated models of effort are empirically supported by the results of this experiment. Both aspects of effort: effort as a state control mechanism and effort as a central executive resource control have been demonstrated.
Furthermore it has been demonstrated that the RSME measures the construct of effort.
The research findings described in this chapter clearly demonstrate that people are able to rate their effort investment according to the various demands that tasks (situations) may put upon them. It may also have become clear that effort is related to both the task demands and the performance potential a person has to meet these demands.

This supports our initial assumption that people use an overall rating of their psychological 'costs' that integrates an estimation of the task demands: actual performance potential and 'time-on-task'. The scores of the RSME may be interpreted as the 'costs' incurred upon an individual when a particular task is being carried out under certain circumstances. The energetics metaphor appears to adequately the concept of mental effort in this respect, as long as the aspect of the varying size of the 'fuel-reservoir' is taken into account.

Another important finding of this study relates to the influence of the factor 'control over the task'. When subjects have the opportunity to exert control over the situation (in this experiment: speed of task-execution), it appears that strategies become very important. By using more time the subjects were able to perform better (with less errors), without having to increase their effort-level. More generally speaking, it means that degrees of freedom do indeed prove to be very important. It allows the operator to select his own strategy for performing his task. This not only makes it possible for him to improve his performance but it also allows the operator to regulate his level of effort expenditure. In other words, it permits the operator to regulate his 'costs' in relation to task-execution. The operator can decide himself how much effort (costs) or time he will have to exert to realize his performance (output). The importance of 'degrees of freedom' has already been emphasized: the freedom that enables the operator to, more or less, select efficient strategies for his work (cf. Teiger, 1978; Sperandio, 1978; van Aalst, et al., 1986; Karasek & Theorell, 1990).

Clearly the strategy in the 'unpaced' condition can be regarded as more efficient in terms of effort investment: the subjects performed better (made less errors) and reported that they had invested less effort, though they did take more time. This situation demonstrates that a trade-off is made between decreasing effort investment and increasing time investment. Apparently two dimensions have to be taken into account when efficiency of work behaviour is the subject of study.
At the same time we can see that subjects performed less efficiently after the pre-treatment A condition: the performance level remained the same but subjects reported that they had invested more effort. This may be regarded as an example of what is known as a loss in 'processing efficiency' while no change is apparent in 'strategic efficiency'.

Chapter

6 Evaluation of cognitive tools[1]

6.1 Introduction

Having dealt with the construction and validation of an instrument for measuring mental effort I will now turn to applying this instrument in an interface evaluation study. This instrument will be used to measure the 'psychological costs' of work behaviour.

In this chapter a study will be described in which the interfaces of two different word processors are compared and evaluated on the grounds of efficiency. The approach to interface evaluation that will be applied has its roots in the Action Facilitation Approach where efficiency is a core concept. A prerequisite for using this approach is that one has to have an adequate and valid operationalization of the concept of 'efficiency in work behaviour'.

The study presented in this chapter poses two questions. The first is: can we measure (the improvement of) efficiency in actual work behaviour (i.e. with more realsitic tasks)? This question concerns the operationalization and the measurability of the concept of efficiency.
The second question relates to the applicability of a methodology for evaluating Man-Machine Systems (or rather the interfaces of those systems) on the basis of the concept of efficiency. This second question directly addresses the validity of the Action Facilitation Approach.

6.2 General outline of the study

The general idea behind this study is to check whether empirical support can be found for the validity of the Action Facilitation Approach. This will be the case when:

1. experts are able to differentiate between two word processors when they evaluate those two word processors according to AFA criteria;
2. the word processor that is found to correspond best with the Action Facilitation Approach principles proves also to be the one that allows people to work most efficiently.

1 The study described in this chapter was carried out in the Laboratory for Work Psychology at Delft University of Technology with the assistance of Mario van Dulmen.

The two word processors that were selected for this study will from now on be referred to as WP-A and WP-B. It is evident that it would have been more appropriate to compare a word processor that was designed according to AFA guidelines with a word processor that was designed from a different perspective but since such a word processor does not exist I selected two that differ in several ways but in respects that are relevant to Action Facilitation (see Chapter 2).
From the available range of word processors I selected one that was command-driven (WP-A) and one that was menu-driven (WP-B). The word processors did not differ with respect to their functionality, both were from the same 'generation'. However, they differed with respect to their structure and in the way in which they were operated. The menu-driven word processor is very rigid in its structure; it does not offer information about how to proceed or how to return to the main menu when one is editing a text, etc. For all text manipulations (printing, archiving, etc) it is necessary to return to the main menu first. A particular function key has to be used for this purpose but relevant information such as what function key must be used, is not presented on the screen. An external aid (like a stencil) has to be used or else all the relevant function keys should be learnt by heart. The program does not allow for parallel processing, like editing one text while printing another. In terms of Action Facilitation one could say that the rigidness of this menu-driven word program means that it deviates to a large extent from the work procedures that are normal for most workers.
The command-driven word processor (WP-A) offers some information about the commands to be used on the screen. This overview may help the user to orientate to the program and thus also to the task. Such information is lacking in WP-B. Furthermore WP-A has a more flexible structure than WP-B because one can proceed from one operation to another without having to return to some kind of main menu. This aspect of orientation is regarded as very essential from the point of view of Action Facilitation.
WP-A is thought to comply most with the design principles formulated within the Action Facilitation Approach.

These word processors were evaluated in two independent procedures. First an 'expert' assessment was given. Several experts (scientist who are well acquainted with Action Theoretical concepts and with the Action Facilitation Approach) were asked to evaluate both word processors according to the dimensions formulated within the Action Facilitation Approach (as described in Chapter 2). These dimensions have been operationalized in a check-list (see appendix).
Since both word processors were selected according to critical features of the Action Facilitation Approach it was expected that the evaluation of the experts might result in a clear distinction between both word processors. This procedure should actually be regarded as a check for the selection process.

Secondly, both word processors were experimentally evaluated. A group of skilled subjects used them to carry out some standardized word processing tasks. During task execution the amount of effort investment was measured by means of the Rating Scale Mental Effort and time parameters were also registered. These parameters were used to measure differences in efficiency.

6.3 First study: the Expert evaluation of word processors

Eight experts were asked to evaluate both the word processors according to 'Action Facilitation Approach' dimensions. The experts were selected on the basis of their knowledge of 'Action Theory' and the 'Action Facilitation Approach' and because they are familiar with the field of interface evaluation.

6.3.1 Design and methods

The design guidelines formulated within the AFA (cf. Chapter 2) have been operationalized in distinct questions (Arensman, 1990; see also appendix A).

The AFA guidelines are listed below:
- Try to support the process of action preparation, i.e. orient to the task and form an action plan by offering adequate information about the system.
- Try to achieve an uninterrupted execution of action plans by presenting adequate feedback about progress and results of activities.
- Make it possible for changes to be made in action plans and action execution.
- Support, if necessary, parallel execution of various activities.
- Offer means to support the supervisory process, especially with regard to anticipating future actions.
- Bear in mind that people have limited capacities with respect to cognitive, sensory and motorial mechanisms.
- Account for the fact that people strive towards lowering the level of regulating their actions.
- Try to accommodate to the user's working-style and working-methods and other relevant differences between users (skills, knowledge, etc.).

As an illustration:
the third design principle '**Make changes in action program and action execution possible**' has been operationalized in three questions:
- Are commands reversible, for instance by means of an 'UNDO' function?
- Are 'short cuts' possible for experienced users?
- Is there an 'escape' possibility in every situation within the program?

The items refer to the presence/absence of certain interface features. These questions were clustered in each dimension and put together in a check-list. The questions have to be answered with 'yes' or 'no'. The questions have been formulated in such a way that a positive answer ('yes') always indicates the presence of a feature that is supposed to support action facilitation. Therefore the total number of 'yes' answers could be summarized as being equal to the total score per dimension. The total score for each word processor is the unweighed sum of the scores of all the dimensions.

A second check-list was used consisting of the five 'usability' principles of the DIN-norm 66234/8 (Deutsche Institut für Normung). These principles contain recommendations for designers of interfaces. These principles are the result of an extensive survey by Dzida et al. (1978) in which experienced users of computer systems were asked to indicate critical aspects of an interface. This resulted in five principles which can be seen as the minimum requirements for man-computer dialogue designs (Dzida, 1985). These requirements refer to certain relevant work-psychological principles though they are not specific to Action Theory and the Action Facilitation Approach. This is what made them suitable for our purpose of obtaining a general opinion from experts concerning these word processors. These results can be regarded as a check for the results of the AFA check-list (i.e. a kind of counter expert opinion).

This DIN-norm consists of the following five principles (DIN 66238 part 8, 1986):

1. **Suitability for the task**: A dialogue is fit for the task it has to perform if it supports the user in the job he is actually doing, without an unnecessary additional strain being placed on him by the system itself.
2. **Self-descriptiveness**: A dialogue is self-descriptive if, on request, the user can obtain an explanation of the purpose and capabilities of the dialogue system, and if each step of the dialogue is immediately comprehensible or if the user can ask for information explaining the respective dialogue step.
3. **Controllability**: A dialogue is said to be controllable if the user can influence the speed of operation as well as the selection of the tools or type and scope of inputs and outputs.
   The user must be able to adapt the dialogue speed to his own working speed. For example, work must not be governed by a work-cycle rhythm, the user must not be put under pressure by having to fear that displayed material will disappear from the screen. The user's inputs should not be delayed by having to wait unnecessarily for the output of data from previous dialogue steps.
4. **Correspondence to user expectations**: A dialogue corresponds to user expectations if the system's dialogue behaviour is based on user's experience with work processes as well as experiences formed in the

course of using the system, using the user manual and doing user training.

Dialogue behaviour within a dialogue system should be consistent. Inconsistent dialogue behaviour forces the user to adapt to changing operating conditions, impairs his learning process and subjects him to undue stress and strain.

5. **Error tolerance**: A dialogue is error-tolerant if the intended result is obtained despite recognizably faulty input without or with only minimal correction effort. The user must be made to understand that an error has been made so that it can be corrected.

   User inputs must not result in undefined system statuses or system breakdowns. It may be advisable to correct uniquely correctable errors automatically and to continue with processing. Here it is generally meaningful to explain to the user in a manner comprehensible to him how a correction is performed. It must be possible to switch off automatic correction.

   If an error can be corrected by the system in various ways, the alternatives should be presented to the user for selection without excluding the possibility of new input.

These five principles were presented to the experts in the form of five statements concerning the word processor at hand. The experts were requested to answer on a five-point scale whether they agreed or disagreed with this statement.

Example (design principle 2):
"The goal of the dialogue and the possibilities of each step in the dialogue are made clear enough"

not at all |——|——|——|——| completely
1 2 3 4 5

Procedure

All eight experts worked with both word processors (four were asked to work first with WP-A and then with WP-B and the others were asked to work in the reverse order to rule out the possibility of results being affected by the order in which these word processors happened to be used). They then had to answer the questions in the check-lists after accomplishing a standard word processing task. Within each group a few experts were asked to use the DIN check-list first and the other experts were requested to start with the AFA check-list.

The tasks that are used in evaluation studies should be similar to the 'real' tasks of that particular work domain. By this I mean that these tasks should

allow the following aspects: 'action orientation', 'action preparation', 'action execution', 'monitoring' and 'supervision' to be present. The tasks that have been developed for this study are examples of everyday word processing tasks.
Two texts, comparable in length and format, that dealt with the same topic (two chapters of a research report) were selected. The experts were asked to edit these texts and make lay-out adjustments according to instructions on a paper concept of the text. Eight representative word processing operations, like 'moving lines', 'underlining text', 'bold printing', 'deleting lines', 'inserting text', 'indenting lines', 'starting printing' and 'storing the text' were selected. The length of the task was such that, under normal circumstances, it could be finished in about half an hour.

The procedure regarding the selection of word processors has already been described in the previous section. WP-A was the command-driven program which is expected to be favourably judged by the experts.

Subjects
As has already been mentioned, eight work psychologists who are very well acquainted with the field of interface evaluation and the theoretical notions of Action Theory and Action Facilitation, agreed to participate in this study. The nature of their work requires that they are also skilled in word processing.

6.3.2 The expert's evaluation results

Figure 6.1 presents the results of the expert's evaluations for both word processors as assessed according to the Action Facilitation check-list.
This figure clearly demonstrates that seven experts gave WP-A higher appraisal than WP-B when they evaluated both word processors with the AFA check-list.
Variance analysis (MANOVA; SPSSPC) confirmed that WP-A received statistically significant better ratings ($F(1,7)= 240$; $p=.001$).

Figure 6.2 presents the results of the evaluation according to the 'DIN check-list'.
These check-list results were also been tested according to the same procedure as the AFA check-list.
It appeared that no statistically significant differences were found when both word processors were compared according to the DIN check-list ($F(1,7)=1.45$; $p=.268$).

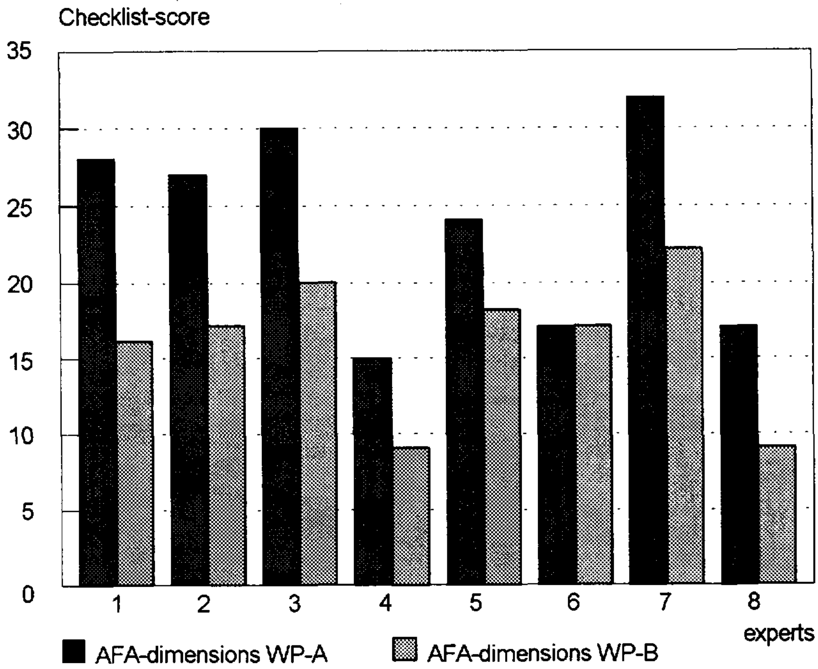

Figure 6.1: *Expert evaluation of Word Processors with AFA-checklist*

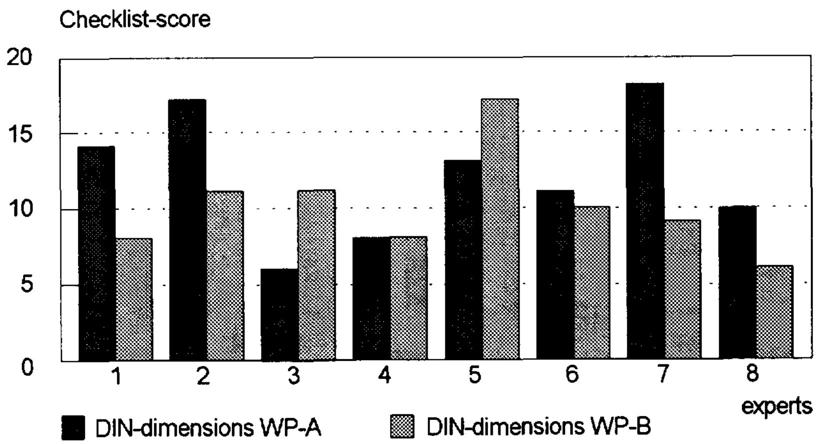

Figure 6.2: *Expert evaluation of Word Processors with DIN-checlist*

6.3.3 Discussion of the first study

The rating difference that emerged when the AFA check-list was used was as predicted. WP-A rated favourably in the eyes of seven of the eight experts: the result was quite unambiguous. However, the DIN check-list ratings did not reveal a similarly clear result. Although five out of eight experts gave higher ratings for WP-A than for WP-B the difference in terms of the scores was smaller.

One explanation can be found in the fact that with the AFA check-list higher scores can be obtained because the check-list consists of more items than the DIN check-list or that, in other words, the variance in the DIN check-list scores is restricted.

An alternative explanation could be that the DIN check-list differentiates less clearly between both word processors. In this respect it is worth noting that, helped by the AFA check-lists, the experts' judgement was much more consistent than the judgement made with the DIN check-list. By comparing the variances explained by the differences between both systems and the various raters we can get some idea of the inter rater reliability. For this we use the formula for intra-class correlation (Winer, 1991):

$$[\, r = 1 - MS_{w.\,people}/MS_{b.\,people} \,]$$ [2]

In the case of the AFA check-list the variance attributable to differences between the word processors is much larger than the variance that is attributable to differences between the experts (r = 1 - 6.5/26.7 = .76). In the case of the DIN check-list the opposite was true (4.81 versus .01).

This means that in this study the experts found it difficult to evaluate both word processors with the help of the DIN check-list. It is a general complaint among experts in the field that the DIN norms, as formulated in the preceding section, are very general and vague. This may also partly explain why we did not find significant differences with respect to the DIN rating of both systems. The vagueness of the DIN dimensions opens the way to various interpretations and prevents experts from reaching a consensus. This makes the DIN check-list unsuitable as an evaluation instrument. On the other hand it makes it clear that the Action Facilitation Approach may be viewed as a coherent and manageable frame of reference.

6.4 Second study: experimental evaluation of word processors

This second study again consists of two parts: one part that is devoted to exploring the measurability of efficiency in work behaviour and another part that is devoted to the comparison and evaluation of both word processors. However both parts are combined in one experimental set-up which is why the design and methods will be outlined together in one paragraph.

2 $MS_{w.\,people}$ is the Mean Squares (variance) caused by the various raters, and $MS_{b.\,people}$ is the variance that can be attributed to both systems.

6.4.1 General outline of the experiment

As has already been stated (in Chapter 2), efficiency increase can be considered to be the result of a learning process leading to the acquiring of skills. This process may include a conscious search for alternative ways (strategies) to execute the task but it may also involve processes of a less conscious nature (e.g. Leplat, 1989).

In order to be able to measure actual work behaviour efficiency it is necessary to provide subjects with the opportunity to improve their efficiency. This means that subjects need time to learn and to try alternative strategies.

In order to meet this requirement we created conditions that comprised two experimental sessions. During these experimental sessions subjects executed word processing tasks. Before the start of the first session and in between the sessions the subjects had the opportunity to practise and to learn. These training sessions were 45 minutes long at the very most. However, subjects decided themselves when they felt that they were 'ready' for the experimental session. In general the subjects decided to practice for 25 to 35 minutes.

The first half of the group of subjects worked twice with the same word processor while the second half of the group of subjects worked successively with both word processors.

It is assumed that in the condition where subjects worked twice with the same word processor efficiency would improve because in this condition subjects were given the opportunity to learn and/or look for alternative ways to execute the task (i.e. adapt their action plan).

The remaining situations where subjects worked with both word processors were meant to provide a comparative study of both word processors.

6.4.2 Design and Methods

Design

Inter-individual differences constitute a well-known problem in interface evaluation studies (cf. Egan, 1988; Mayer, 1988). Within-subject designs are generally advised in these situations (Winer, et al., 1991). However, when learning or problem-solving behaviour is (or may be) involved there is always a risk of having transfer effects (cf. Drury, 1990). What is learned in solving the first problem is likely to affect performance in the next. So transfer effects are likely to occur in learning to use word processors, as will be seen to be the case in this study (cf. Singley & Anderson, 1985). The solution suggested by Drury (1990) was: train subjects in all conditions to ensure that plateaus of performance are reached in each condition.

The learning/training process required to achieve this level can be of interest in itself. If subjects have had the same amount of practice with the various word processors then differences in levels of acquired skill (efficiency) may

indicate how easy it is to operate these word processors. When the word processor order of use is counterbalanced the eventual transfer effects become visible.
Moreover a within-subjects design allows small differences to be detected even though the sample size might be limited. Another possibility would be to select and match subjects. However, this is very difficult to realize in practice.

To avoid this problem the decision was made to use a within-subject design in this study and a between-subject design as well.
I compared two word processors, and, as mentioned above, there were two groups of subjects who were exposed to different experimental conditions. The schematic representation of the resulting design is presented below.

	1st session	2nd session
condition AA	wp-a	wp-a
condition BB	wp-b	wp-b
condition AB	wp-a	wp-b
condition BA	wp-b	wp-a

This design offers various possibilities. First of all it should be clear that in conditions AA and BB the idea was to explore the measurability of improvement of efficiency in work behaviour while in conditions AB and BA the intention was to execute a comparative evaluation of both word processors. Of course conditions AA and BB can also be contrasted in order to test differences between WP-A and WP-B. Contrasting conditions AA and BB would in fact mean setting up a between-subjects design.

In the previous chapter it was concluded that people estimate their 'costs' according to the two dimensions: 'time' and 'effort'. Therefore we will indicate in Figure 6.3 the 'costs' according to those two dimensions.

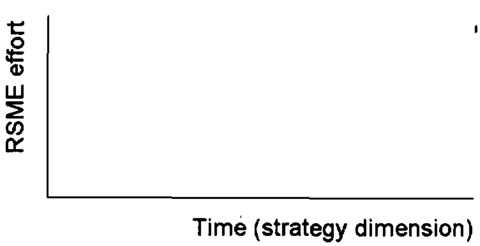

Figure 6.3: *Two dimensions of efficiency.*

The X-axis represents the 'strategy dimension' and the Y-axis represents the 'effort dimension'.
With respect to *strategic* efficiency one can imagine that the 'longer the route through the program', as indicated by the amount of time that is needed to execute an action or task, the less efficient that interface (or software-program) will be.
Regarding *processing* efficiency we could say that the more mental effort has to be exerted while using the interface, the less efficient that interface will be. I will use the score on the Rating Scale Mental Effort (RSME) to indicate this 'costs' aspect.

In the previous chapter I also mentioned that people may change their strategy which can lead to a trade-off taking place between 'time' and 'effort'. In such situations it is difficult to decide whether there will be an increase or a decrease in efficiency because 'time' and 'effort' are not easily expressed in comparable units. A possible solution in these situations may lie in transferring the respective scores into 'z-scores'.

Methods
It was our intention to evaluate the word processors under circumstances that - as much as possible - resemble normal working conditions. Hence the reason that an office environment was simulated in the laboratory. This 'office' was furnished with the usual office equipment: desk, telephone, personal computers, etc.
Facilities were developed for taping and registering the entire 'dialogue' between the person and the computer, i.e. all the keystrokes and all the information presented on the video screen. All keystrokes were stored in a file together with the time when they were made as recorded on the system's clock (in milli seconds).
Furthermore a video camera was installed for observing the subjects from the experimenter's room while they were working.

Procedure
Each subject came to the laboratory for half a day. After an introductory session during which the equipment was demonstrated and the goal of the experiment was explained, the subjects were informed and instructed on the functioning of the word processor(s) they were about to work with. Immediately after these instructions the subjects had the opportunity to practice so that they could familiarize themselves with the word processor. Subjects were allowed to ask questions about the program. During this practice period the subjects were requested to take special note of a list of regular word processing operations from which the operations during the experimental session would be selected. The remainder of the experimental session was taken up with doing the experimental tasks.
After a short break the procedure of instructing, practising and experimentation was repeated.

During the experiment sessions, when the measuring was being done, the person leading the experiment disappeared from the laboratory.

Subjects

As has been pointed out in Chapter 2 the efficiency level is closely related to the acquired skill level. Persons who are very skilled in a task usually perform it very efficiently. Consequently for evaluation purposes one should select subjects with comparable skill levels. This skill-level should not only be operationalized in months or years of computer-experience as is often done (i.e. novices versus experienced users). In our view subjects in evaluation studies should be experienced and skilled professionals in the field for which the system is developed. They are the ones who know best how to perform the task. Of course they should be instructed and trained to familiarize themselves with the system that is to be evaluated. Ideally the subjects should be experienced users of the system but with comparative evaluation this may not always be possible. One can not expect subjects to have a lot of experience with two or more comparable systems.

The subjects in this experiment were 28 professional secretaries (aged between 21 and 45 years), that is to say, there were seven subjects per condition[3]. They were all employees of Delft University of Technology who worked during their regular working hours in our 'office'. Each subject participated voluntarily and with the permission of their superiors. The subjects did not receive extra payment for their participation, just their salary.

We chose secretaries as subjects in this experiment because they are very skilled in word processing tasks. Since they were all employed at Delft University of Technology they all used the same type of word processor in their daily work which was not of the same type as either one of the word processors used in our evaluation study. Each subject had at least one year's extensive experience with word processors. These were very important aspects because they made our subjects comparable in the relevant aspects such as level of qualification, experience with word processors, etc.

Task

The subjects performed the same kind of word processing tasks as those executed in the 'expert evaluation' study according to a fixed number of prescribed operations. These tasks comprised eight representative word processing operations like: moving lines, deleting lines, archiving, etc. Both tasks could be completed within approximately half an hour. Each subject

3 Originally we strived to have at least 10 subjects per condition. However, several 'candidates' were forced, for various reasons, to let us down. We were unable to find new subjects at short notice.

was given the opportunity to finish the task. Differences in the quality of performance were hardly possible with this task. The order of presentation of both tasks was balanced.

As explained in Chapter 2, efficiency is the ratio between 'costs' and 'benefits'. In this experiment we will classify 'the resulting output' as the 'benefits' of work behaviour, or in other words the 'completion of the task'. Since all subjects were given the opportunity to complete the task, we can assert that the output of work behaviour is comparable for all subjects. Quality standards were irrelevant which meant that the 'benefits' were 'standardized' when subjects had completed the task so we only had to compare the 'psychological costs' (how long it took the subjects to complete the task and how much effort was required) of each subject to get an estimate of their (relative) efficiency during the execution of the task.
The task in this experiment was one with a clear starting and finishing point. However, it should be acknowledged that in daily practice tasks usually have a more continuous nature. In such situations one should establish a representative time-frame that could serve as a basis for efficiency measurement.

Variables
Keystrokes and time-parameters were registered during task execution. With the help of these parameters it was possible to calculate the exact length of each experimental session, i.e. the task (TASK-TIME). Moreover it was possible to extract some additional information from these parameters, like the amount of time subjects needed to accomplish the prescribed operations (EXECUTION-TIME). The EXECUTION-TIME is the net amount of time that subjects need for the various operations. It is the amount of time required for each prescribed operation with respect to the text excluding the time that subjects need to read the document, look for the next operation and print the corrected text. When the subjects started printing out the text the task was regarded as completed and the session was over.
The time taken up between TASK-TIME and EXECUTION-TIME is essentially the amount of time that subjects need for orientation and supervisory activities, i.e. scanning the document (how many and what kind of operations are required, where is the first, second, etc. operation) and for proceeding from one operation to the next and to different places in the text. This comes very close to what we would like to call task 'preparation', or Action Theory terms 'action preparation'. Therefore this difference between TASK-TIME and EXECUTION-TIME is called 'time for preparation' (TASK-TIME - EXECUTION-TIME = PREPTIME).

Immediately after each session was finished, subjects estimated how much effort they had invested in the task by means of the Rating Scale Mental Effort. This is an indication of the effort dimension.

Furthermore at the end of each session the subjects who had worked with both word processors were asked to evaluate their experiences on a six-point scale ranging from 'very pleasant to work with' to 'very unpleasant to work with'.

6.4.3 Hypotheses

As has been stated above the conditions AA and BB were created for measuring *improvement* in efficiency. We assumed that in each condition subjects would work more efficiently in the second session. This would be reflected in the reduction of psychological costs. Therefore we anticipated that subjects would need less time to complete their tasks in the second session and/or would indicate that executing the task required less effort in the second session. This should be evident from the shift in a subject's 'position' within one or both dimensions (i.e. TASK-TIME and RSME-score) when plotted in a figure as shown in Figure 6.3.

The conditions AB and BA were designed for comparing both word processors. Since WP-A received the best ratings according to Action Facilitation norms it was predicted that this word processor would be the most efficient one. It was thought that this should become apparent from the lesser amount of time required to complete the task (TASK-TIME) and/or the lower effort ratings when subjects work with WP-A compared to when they work with WP-B. It is assumed that WP-A offers somewhat more facilities for orientation, therefore one would also expect that less time is needed for 'preparation' compared to WP-B.

First the results of the part of the study on the measurability of the efficiency concept will be discussed.

6.5 Results of efficiency measurement

The conditions AA and BB were primarily meant for exploring the measurability of the efficiency concept. The results of these conditions are presented in Figure 6.4 (a & b).

Certain equipment was rendered useless due to technical failure which means that one subject's results for condition BB are missing. For organizational and practical reasons (i.e. because of time constraints) it was not possible to invite extra subjects to take part in the experiment.

In Figure 6.4 the score of each subject on the RSME (y axis) is plotted against the time needed to accomplish the task (TASK-TIME on the x axis). The y axis represents the effort dimension and the x axis represents the strategy dimension.

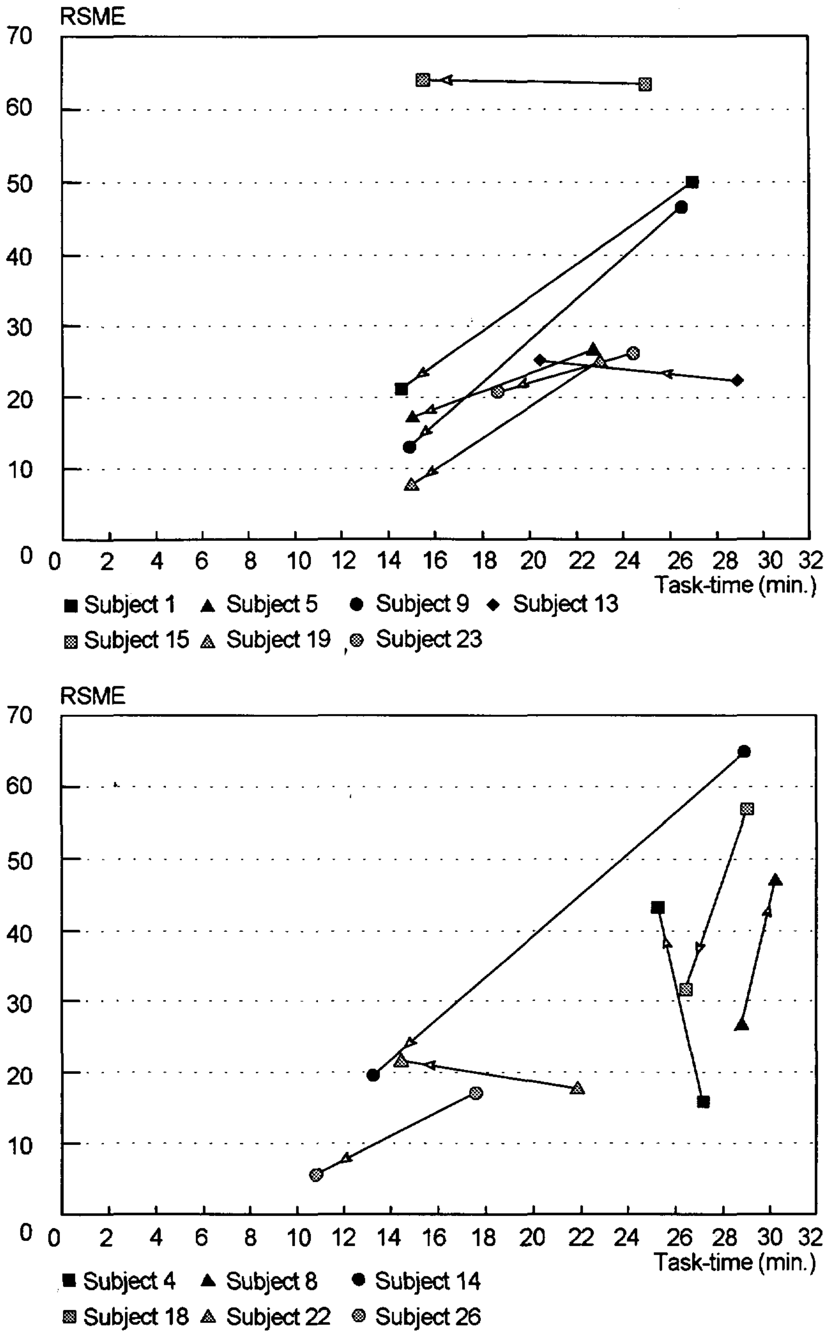

Figure 6.4 (a & b): *Mental effort (y-axis) is plotted against duration of the task (x-axis) for each individual in both sessions. The arrow points towards the second session.*

Figure 6.4[a] shows the results of the experimental condition in which subjects worked twice with WP-A. It becomes clear that all subjects in this condition needed less time to complete their task in the second session and that most subjects (nrs. 1, 5, 9 and 19) also reported having invested less effort during the second experimental session. This meant that there was a shift towards more efficiency in the second session. Subjects 13, 15 and (to a lesser extent) 23 represented the 'strategic efficiency' dimension and subjects 1, 5, 9 and 19 represented a combination of both 'processing efficiency' and 'strategic efficiency'.
The differences in results between both sessions have been statistically tested with the module 'T-TEST' in SPSS/PC.
These results are presented in Table 6.1. Since our hypothesis was that the scores in the second session would be lower we used a 1-tailed probability test.

Table 6.1: *Results of T-Test for condition AA.*

Variables	Mean & S.E. Session 1	Mean & S.E. Session 2	t-value (& df)	1-tail prob (α =.05)
TASK-TIME	25.3 (.85)	16.3 (.88)	7.4 (12)	.000
EXECUTION-TIME	11.4 (1.1)	6.5 (.72)	3.8 (12)	.002
PREPTIME	13.9 (.5)	9.8 (.3)	7.7 (12)	.000
RSME	37.1 (6.1)	24.5 (8.3)	1.3 (11)	.117

It appears from this table that subjects needed significantly less time to complete their task during the second session. The reduction in the RSME score, however, appears not to be significant. As can be seen in Figure 6.4[a] there were only three subjects who reported having invested considerably less effort in the second session, the other subjects reported investing a little less or almost the same amount of effort in the second session. Moreover we can see that subjects needed significantly less time for 'preparation' and 'execution' in the second session. Since the tasks are comparable with regard to length and complexity and the system is the same in the second session, the differences that were found can be attributed to improvement in efficiency.

Figure 6.4[b] presents the result of condition BB. This is the condition in which subjects worked twice with WP-B. Five out of the six subjects in this condition performed faster in the second session but two of them (subjects 4 and 8) reported a substantial increase in effort in the second session. Subjects 14 and 18 needed more time in the second session but on the other

hand these two subjects reported a substantial reduction in effort investment in the second session.
However, none of the parameters appear to show significant results (see Table 6.2).

Table 6.2: *Results of T-Test for condition BB.*

Variables	Mean & S.E. Session 1	Mean & S.E. Session 2	t-value (& df)	1-tail prob. (α =.05)
TASK-TIME	25.5 (2.0)	20.0 (3.3)	1.4 (10)	.091
EXECUTION-TIME	8.3 (1.4)	6.3 (1.1)	1.1 (10)	.140
PREPTIME	17.3 (3.0)	13.7 (3.3)	.8 (10)	.223
RSME	33.8 (8.8)	28.7 (6.7)	.5 (10)	.328

The results of this condition do not quite match the results of condition AA. Figure 6.4[b] has already made clear that the pattern of change in condition BB is different from in condition AA. At a glance one might say that only subjects 14, 22 and 26 performed more efficiently in the second session in condition BB while all subjects in condition AA performed more efficiently during the second session.

Since the results of this condition are rather unclear it might be sensible to take a look at the results on an individual level. It appears from the log file - where the results of each subject are stored - that subjects 14 and 18 did not execute three of the required operations in the first session. Subject 14 completed all operations in the second session, subject 18 performed only slightly better in the second session, as she only skipped one operation. This operation was probably too difficult for her. This explains why she needed more time for the second session; it also indicates that she had not as yet, since the initial training period, mastered the word processor.

Subject 4 executed the task faster in the second session but needed to invest more effort. Subject 8 took more time and invested more effort during the second session. This suggests that they tried harder to complete the task. Subject 4 apparently changed her strategy, an illustration of the earlier mentioned possibility of making a trade-off between saving time and increasing in effort investment.
A gain in strategic efficiency is compensated by a loss in processing efficiency. This means that we have to find a way to compare such diverse quantities as the RSME score and TASK-TIME. The solution may lie in comparing the z-scores of both parameters and calculating a ratio between

these z-scores (z RSME/z TASK-TIME). This results in: -.97 and .85 for subject 4 in the first second sessions respectively. The increase in effort is relatively smaller than the decrease in TASK-TIME because the absolute value of the ratio decreases (i.e. the decrease in the denominator of the ratio is greater than in the counter). This could be interpreted as a relative improvement in efficiency.
This brings the total number of subjects who improved their efficiency in this condition up to five.

6.5.1 Discussion of the efficiency measurement

The results of conditions AA and BB clearly demonstrate that improvement in efficiency in work behaviour can be measured. This improvement becomes apparent from the reduction in the amount of time needed to complete the task which possibly (but not necessarily) occurs in combination with a reduction in effort investment. The results, as presented in Figure 6.4 (a & b) show a clear tendency to lean in the in the predicted direction. These figures also show that there were considerable differences in the individual score patterns, especially in condition BB. Consequently the standard deviations (and standard errors) were great and since the groups were small this might explain why we did not find statistically significant differences at group level.
The results of condition AA were more pronounced than those of condition BB. When subjects worked twice with WP-A they needed significantly less time to complete their task the second time round. When subjects worked twice with WP-B it could be seen that some worked more efficiently. It was also apparent that some subjects worked faster and invested more effort. This may be viewed as changing strategy, putting more effort into the task in order finish a little earlier or, to put it another way, as trying harder. In these cases we saw that an increase in strategic efficiency is compensated by a loss in processing efficiency. This situation demonstrates that people sometimes make a trade-off between two aspects of costs and therefore it proved to be useful to make a distinction between both dimensions of psychological costs.

Other factors were, that one subject did not manage to complete her task in condition BB while two subjects did not complete their tasks in the first session. This obviously affects these people's efficiency measurement scores. Strictly speaking there is no efficiency result when the task is not completed because if it is not completed, i.e. if there are no benefits, the ratio between costs and benefits (i.e. efficiency) does not exist.

If we take this latter observation into account we might conclude from the results of this part of the experiment that we have, by and large, been able to demonstrate changes in efficiency in work behaviour in a direction predicted on the basis of our theoretical concepts. This can be viewed as supportive evidence for the (construct) validity of the concepts and operationalizations of strategic and processing efficiency.

The fact that the results of condition AA are more clearly defined (resulting in significant differences) than those of condition BB can be regarded as supportive evidence for the validity of the AFA. It should be remembered that WP-A was evaluated as conforming most to the AFA guide-lines and therefore as being more directed to efficiency improvement than WP-B. More evidence is expected to emerge from the direct comparison between both word processors (the second part of the experiment).

6.6 Results of experimental comparison of both word processors regarding efficiency

This section deals with the question of whether it is possible to measure differences in efficiency between both word processors by directly comparing them. Figure 6.5 (a and b) presents the results of conditions AB and BA with respect to the amount of time subjects needed to complete the task and the amount of effort they reported having invested.

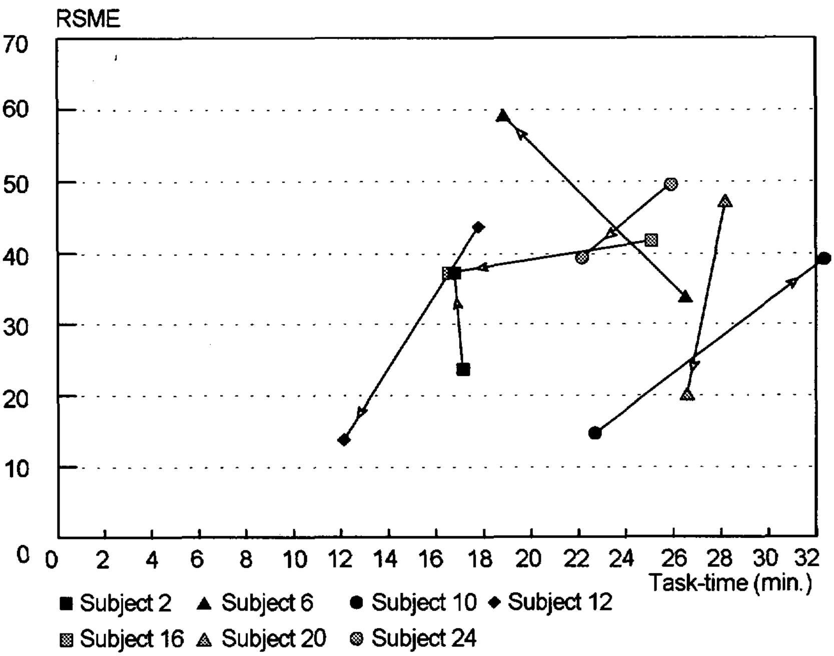

Figure 6.5a: *Condition AB. Arrow points towards second session.*

Figure 6.5[a] shows that, except for subject 10, all subjects in condition AB performed faster during the second session when they worked with WP-B. However, subjects 2 and 6 reported having invested more effort in their second session.

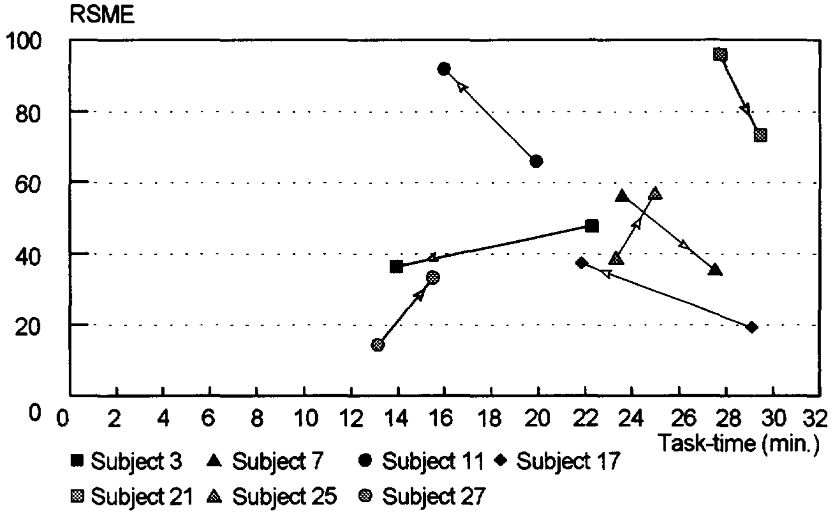

Figure 6.5b: Condition BA. Arrow points towards second session.

Figure 6.5[b] completes the picture because this figure presents the results of the condition where subjects worked with WP-B during the first session. As can be seen in this figure there were three persons (subjects 3, 11 and 17) who needed less time in the second session (while working with WP-A). The other subjects (7 and 21) took more time and two subjects (25 and 27) needed more time and put more effort into working with WP-A.

In order to test whether the order in which the word processors (A-B versus B-A) were worked with influenced the results an analysis of variance (MANOVA) for the conditions AB and BA was made with the factors 'word processor' (WP-A versus WP-B), 'session' (first versus second) both taken as 'within-subjects' factors, and 'condition' (AB versus BA) being taken as a 'between-subjects' factor.
No significant effects were found. However, it appeared that a marginally significant effect ($F(1,26)= 3.8$; $p= .06$) of the variable RSME (subjective effort-score) could be noted as a result of the factor 'condition'. Nevertheless I think we should take this 'marginal' significance seriously since the subject groups were rather small. It proved that the combination B-A was evaluated as being more effortful than the combination A-B (as can also be seen by comparing the level of the RSME scores in Figures 6.5 (a & b)). Working with WP-A after having worked with WP-B was rated as being more effortful than

doing it the other way round. Apparently the work procedures that are required in order to work with WP-B interfere more with those of WP-A than the other way round.

Since it seemed that the order of presentation of both the word processors did have an effect it is not very relevant to compare the first and second sessions of the two conditions. Instead we will concentrate on the first sessions in both cases (i.e. make a between-subject comparison). There appeared to be no significant differences, but, as said before, both the groups were very small. Therefore the first sessions of conditions AA and BB have also been included in the analysis.
A T-Test was executed in which the first sessions of all four conditions were included in order to test whether the results obtained while working with WP-A differed significantly from the results obtained while working with WP-B. The results are presented in Table 6.3

Table 6.3: *T-Test WP-A vs WP-B for all four conditions (first sessions).*

Variables	Mean & S.E. WP-A	Mean & S.E. WP-B	t-value (& df=25)	1-tail prob. (α =.05)
TASK-TIME	24.3 (.9)	24.0 (1.4)	.18	.428
EXECUTION-TIME	11.6 (.95)	8.8 (.89)	2.12	.022
PREPTIME	12.7 (.77)	15.2 (1.6)	-1.45	.085
RSME	36.6 (3.7)	41.5 (7.0)	.63	.268

As the table above shows, there is a significant difference with respect to the variable EXECUTION-TIME. When subjects worked with WP-B in the first session they were able to execute their operations faster.
When we take a closer look at the results of the conditions AB and BA, we can see that subjects 3, 10, 11 and 17 work faster with WP-A, while subjects 2, 6, 12, 16, 20, 24 and 25 worked faster with WP-B (7 versus 4).
The subjects 2, 6, 7, 10 and 21 reported that they had invested less effort while working with WP-A, while eight subjects (11, 12, 16, 17, 20, 24, 25 and 27) also reported having invested less effort while working with WP-B.
The results of this condition seem to be rather ambiguous. Some subjects reported putting less effort into working with WP-A while others reported having put less effort into working with WP-B. This can be seen as an illustration of the differences in individual preferences and strategies between the subjects. Some managed better or preferred to work with WP-A, while others have a preference for WP-B. However, it should be noted that these results had been influenced by the order of presentation of the word processors.

In this respect it is useful to look at the subjective evaluations of both word processors.

6.6.1 Subjective evaluation of the word processors

The subjects in conditions AB and BA were asked to rate their preferences with respect to the word processor(s) on a six-point scale. The word processors could be evaluated with terms like 'very pleasant to work with' (1) and 'very unpleasant to work with' (6). The higher the score, the more negative the rating.

The mean score for WP-A was 4.4 (S.E.= .3) and the mean score for WP-B was 3.7 (S.E.= .6).

Since these ratings have just an ordinal measurement level a non parametric test was used (Mann-Whitney). It appeared that no statistically significant differences could be found to justify preferring WP-A above WP-B or vice versa (see Table 6.4).

Table 6.4: *Results of Wilcoxon (Mann-Whitney) test concerning preferences for WP-A or WP-B (two subjects with missing data).*

Mean Rank	Cases	
13.54	12	wp-A
11.46	12	wp-B
	24	Total
U = 59.5	W = 162.5	p(1-tail)=.239 (α =.05)

It is interesting to note that the subject's general opinions about both word processors were rather negative. However, this is not startling, if one considers that all the subjects were used to working with later 'generation' word processors which were therefore, in many respects, superior to the ones being used in this experiment. The subjective opinions concerning both word processors are also very divergent. The ratings vary from 1 to 6 for WP-B and from 3 to 6 for WP-A.

It would be interesting to know how the individual ratings relate to the individual experimental results. This might tell us whether efficiency really is a contributory factor to individual preference.

For this purpose we selected the subjects who worked more efficiently with WP-A or, as the case might be, with WP-B. Their preferences with respect to the word processors were also noted.

Table 6.5: *Efficiency related to individual preferences.*

	faster	less effort	preference
works more efficiently with WP-A:			
	faster	less effort	preference
subject		2	[4]
"	3	3	*
"		6	-[5]
"		7	-
"	10	10	A
"	11		-
"	17		B
"		21	A
works more efficiently with WP-B:			
	faster	less effort	preference
subject	2		*
"	6		-
"	7		-
"		11	-
"	12	12	B
"	16	16	B
"		17	B
"	20	20	B
"	24	24	B
"	25	25	A
"	27	27	B

It is clear that subjects' expressions about preferences match up with their behaviour. In general one can say that subjects give the highest ratings to the word processor that enables them to work most efficiently. Only subject 25 worked more efficiently with WP-B but claimed to prefer working with WP-A. For subjects 2, 6, 17 and 21 the results were not so clear.
These findings suggest that individual preference is related to efficiency.

6.6.2 Discussion of the comparative evaluation study

As one might expect the sequence in which the word processors were worked with had an effect on the comparative evaluation results. It appeared to be most difficult to work first with WP-B and then with WP-A. We should bear in mind that WP-B was the menu-driven word processor which meant that subjects had to adapt to certain strict working procedures (returning to

4 preference rating is missing

5 indicates a tie

the main menu). Apparently it is more difficult to handle a situation in which the work procedures are less strict (i.e. the worker has more control) when one has got used to other strict work procedures such as those of the menu-driven word processor.
Therefore it may be concluded that there were transfer effects from the first session to the second session in the conditions AB and BA. The training periods and the intervals between both sessions proved to have been insufficient. Probably it would have been better if both sessions had been spread over a longer period of time. However, this might have induced other uncontrollable factors like a change in the physical state of the subjects, motivational problems, etc. Furthermore how long this intermediate period, which prevents transfer effects, should last is unknown. These factors constitute a serious complication for employing within-subjects designs in evaluation studies like this one.

Since I was aware of the risks of a within-subjects design the experiment was designed in such a way that between-subject comparisons were also possible. By comparing the first sessions of all four conditions we gained a between-subject design. It turned out that subjects were able to complete their task more rapidly when they worked with WP-B while no effect was perceptible in the area of the effort-dimension. These results suggest that WP-B is more efficient to work with than WP-A. This contradicts our hypothesis that WP-A would prove more efficient to work with because it corresponds more closely with AFA design guide-lines. However, we should bear in mind that subjects did not complete their task faster but were only able to execute the required operations more quickly. This is not unusual for a menu-driven program as long as one knows how these operations have to be executed. If one knows which item to select from a menu the whole thing can be done quickly but it may take some time before one knows which item to select.

Finally we should not forget that our subjects used both word processors for the first time and had relatively little time to get to know these systems. From literature (cf. Mayer, 1988) we know that novices profit more from a clear menu structure than from any form of dialogue mainly because only one way is usually presented to carry out the task which is easier to remember than various alternatives. However when novices become more experienced the earlier advantages of the menu-structure turn into serious drawbacks. From this point of view it is useful to also take into consideration the results of the other two conditions. The results of these conditions confirmed that subjects were more likely to *improve* their efficiency while working with WP-A than while working with WP-B.
One possible explanation could be, that the rigidness of the structure of WP-B holds subjects back from adopting different strategies or from enhancing their strategy so as to improve their efficiency. WP-A is by contrast more

flexible and provides more opportunity for the worker to be in control which makes the *potential* for efficiency improvement greater. One could also say that there is more to learn from working with WP-A. Learning has, to some degree, to do with gaining more experience which on the one hand means skill acquisition and on the other hand means finding the best way to carry out a task. Both aspects relate to improving efficiency (cf. Chapter 2).

Furthermore in a comparative evaluation the fact that two subjects were unable to complete their task while working with WP-B (condition BB) should also be taken into consideration. It may be seen as an indication that WP-B is rather difficult to master for some people. This may be because WP-B is less transparent because of it's rigid menu-structure. For some subjects this is apparently a problem while for others it is not. This illustrates again the differences in preferences and working-styles which became visible in the subjective evaluation of both word processors. However, it is important to note that almost all subjects had a preference for the word processor that allowed them to work more efficiently whether this efficiency improvement showed up on the time-dimension axis or on the effort-dimension axis.

Additional information can be obtained by comparing the results of the conditions AA and BB. As Tables 6.1 and 6.2 show there was hardly any difference in TASK-TIME where the first sessions in both conditions were concerned. On the other hand the EXECUTION-TIME was less in condition BB in the first session. Consequently the PREPTIME was greater in this condition. This would indicate that subjects in condition BB take more time to prepare their actions. This interpretation seems valid because, as has been stated before, this word processor offers less help in orientation than WP-A.
When tested (T-Test) the difference in PREPTIME between the first sessions in both conditions was not found to be significant (t-value -1.18; 1-tail $p(\alpha=.05)=.131$). The mean PREPTIME for WP-A was 13.9 minutes with a standard deviation of 1.2 min. and the mean PREPTIME for WP-B was 17.3 minutes with a standard deviation of 7.3 min. In other words: the variance between subjects in condition BB (WP-B) is much larger than in condition AA (WP-A). These large differences between the subjects in condition BB might be attributed to various factors such as differences in the skills, knowledge or strategies used by the subjects. However, since we selected subjects on the basis of their skills and because in condition BB the subjects received equal opportunities for familiarizing themselves with the word processor, it is not very likely that these factors will explain a lot of these variances. Therefore it is more likely that differences in work styles and/or the strategies used by the subjects are responsible for the variance.
It should be noted that in condition AA the PREPTIME required for the second session was significantly less than for the first session (t-value 7.7; 1-tail $p(\alpha=.05)=.000$) but there was no significant difference between both

sessions in condition BB (t-value = .8; 1-tail p(α =.05) = .223). Although the mean PREPTIME decreased from 17.3 min. to 13.7 min., there are large standard deviations in both sessions (7.3 and 8.1 respectively). The standard deviation is even a little larger in the second session while in condition AA the standard deviation diminished in the second session.
This finding indicates that subjects who worked with WP-A know how to proceed when they have to do the same kind of task again. This creates a more efficient strategy. The explanation can be found in the fact that the help information on the screen with WP-A is sufficient for designing an adequate 'action program'. At the start of the second session they have their 'action program' ready and available. This action plan has been executed before and has led to the desired results. Such an action plan can be executed more resolutely a second time. This might be regarded as a stage that comes before lowering the regulation level of an action. The task still requires the *subject's attention but there is a reduction in the amount of time necessary* for 'orientation' (PREPTIME) in the second session. This interpretation may also explain why effort reduction is absent in those cases.

Furthermore this finding supports our earlier assumption that WP-B seems to be more difficult to master. Alongside of WP-A this word processor (WP-B) offers less help information. Consequently subjects find it difficult to orient themselves to the tool and they fail to design an adequate action program. This means that during the experimental session some subjects are still trying to find out how the word processor has to be operated. At the start of the second session they have not yet all found an adequate action program and so some subjects were still modifying their action program during the second session. One might call this a 'trial and error' strategy. The fact that there are such large standard deviations with respect to EXECUTION-TIME *supports this interpretation. Some subjects are lucky and* succeed rather quickly but for other it takes longer to succeed.
An example of such a difference in strategy is, that some subjects proceeded from one operation to the next by scrolling and scanning the screen (by trial and error), while others looked for a characteristic word in the text and proceeded to use the 'search-key'. The first strategy is more time consuming and requires more attention (effort) than the second strategy. The latter strategy is therefore more efficient.

6.7 Conclusion

In view of the above mentioned facts, we can formulate the following conclusions.
First of all it appeared that the check-list that was derived from the Action Facilitation design guide-lines had a satisfactory inter rater reliability. This may be regarded as an indication that the Action Facilitation Approach offers

a coherent and manageable frame of reference for evaluation purposes. The DIN check-list appeared not to differentiate clearly between both word processors. The inter rater reliability of the DIN check-list proved to be very low. This makes the DIN check-list unsuitable as an evaluation instrument. Secondly, the experts' evaluation indicated that both word processors differed in the relevant Action Facilitation Approach dimensions. Word processor A was found to comply more with the AFA guide-lines than word processor B.

Furthermore in this study I have been able to demonstrate improvement of efficiency in work behaviour. This occurred when subjects worked with the same word processor for two consecutive sessions. When they worked with WP-A there was clearly an increase in efficiency and when WP-B was used there was a less marked increase in efficiency. These results suggest that WP-A has more potential for efficiency *improvement* than WP-B because the results show a more profound shift towards more efficient work behaviour in condition AA.
Consequently it may be concluded that when the AFA check-list is used to evaluate a system it appears to be more directed at evaluating the *potential* for efficiency improvement than at evaluating the actual differences in efficiency between several systems. As the AF approach is actually a set of design recommendations aimed at improving efficiency this conclusion supports the conclusion on the coherence of the Action Facilitation Approach.

The results of conditions AB and BA made it clear that a direct comparison with a 'within-subject design', like the comparison made between conditions AB and BA, requires an experimental set-up that is different from the one that was used in this study. The chance that various newly learned work procedures and acquired knowledge will interfere appears to be rather great. This means that in order to evaluate systems on their efficiency one needs to have subjects who have gained a lot of experience with those systems and who have passed on from the learning phase. However it would be quite difficult to use within-subject designs in these circumstances because the training periods needed to get subjects up to the required level of experience with the various systems would be too long.
This study has made it clear that efficiency improvement can be quite indicative. It would indeed be worth investigating the differences in efficiency between novices and experienced users of the various systems in order to get an idea of how far efficiency in work behaviour can be improved. Such a study would probably also give a deeper insight into differences in workstyles and strategies between novices and experts. These aspects seem to be critical in system design. In this study it appeared that subjects preferred the systems that enabled them to work efficiently. Subjects expressed a preference for the system that enabled them to complete their task faster

and/or with less effort. It is interesting to note that some subjects focused on the time dimension while others focused on the effort dimension. They also differed in their preferences for various systems which might indicate that while some work-styles and strategies are effective with one system they are not effective with another.

6.7.1 Theoretical Implications

The results of this study show that the idea of striving towards (improvement of) efficiency can be applied to system design. In this study I have been able to demonstrate that efficiency in work behaviour can be measured and that it is sensible to make a distinction between the two psychological cost dimensions. Clearly efficiency is not a one-dimensional concept and there are trade-off mechanisms at work. Moreover it seems as if the operationalizations of the psychological costs of work behaviour are valid. The experiment results correspond with the predictions that were formulated from theoretical notions. We have found improvement in efficiency in the aspects where we expected to find improvement.

This finding might also be regarded as empirical support for the theoretical notions behind the experimental set-up; in this case Action Theory and the Action Facilitation Approach method. The results of this study suggest that the Action Facilitation Approach rates the potential for improvement in efficiency of a system rather than the direct efficiency level achieved at an early stage.

It is likely that people will increase their efficiency during the course of time with all systems, no matter how complex they are (cf. specialists). However, increasing efficiency at an early stage is more interesting from a design point of view because it demonstrates that people have to invest little time or effort in acquiring the skill to operate that system. This makes it more attractive for various groups of users to work with such a system.

6.7.2 Practical Implications

It is evident that regarding the results of this study a first suggestion to designers would be to use the design guide-lines that have been formulated from the Action Facilitation perspective. This will lead to the development of systems that enables people to easily increase their level of efficiency while using that system.

Since most subjects expressed a preference for the system that allowed them to work more efficiently, it may be concluded that efficiency (improvement) is a relevant criterion for evaluating work tools.

A relevant finding is that individual differences with respect to preferences, working-methods and work-styles seem to be very important. Not only can this be seen as a possible explanation for not finding statistically significant results in this study but it may have far-reaching practical consequences that are relevant for designers as well. In general two conclusions are possible.

One conclusion could be that it is not rewarding to put a lot of effort into making systems 'user-friendly' because individuals differ too much with respect to preferences and work-styles. It would mean that one actually has to develop 'custom made' systems. It is much cheaper to develop and introduce a robust system and put a lot of effort and time into training and teaching people how to operate the system.
Such a solution may be acceptable with a select group of users but would be very expensive for widespread systems (such as word processors).

The second conclusion is that system designers should invest time and effort in making their systems flexible enough for various working-methods and work-styles. This implies that designers should first study how tasks in a particular domain are executed by novices and experienced persons (experts) alike and establish which are the most frequently used strategies and working-styles. Designers of interfaces should design their products in such a way that various strategies can indeed be realized. The Swiss psychologist Ulich came up with an approach for task design that he called 'differential dynamic design' (Ulich, 1978; 1991; Ulich et al., 1980) which may also be applied to system design. The essence of this approach is that workers should be able to modify the system according to their own preferences. This can be done by offering various alternatives with respect to task structures or interfaces or dialogue-structures and by letting the workers choose themselves which alternative they prefer at a particular moment. Furthermore they should have the opportunity to change their opinion at a particular moment. This approach has been successfully applied to task design and could be applied to interface design as well, for instance, by developing software-systems that are adaptable and easy to programme or by offering various dialogue-systems (Ulich 1991). From a work psychological point of view this would be the preferable solution because it would mean that the worker is really in control and consequently this might improve the quality of the work and an individual's well-being.

Chapter

7 Concluding remarks

7.1 Introduction

In this chapter we will return again to the objectives of this study. In summarizing the main conclusions of the various studies described in this book I shall endeavour to formulate an answer to the question of whether or not the goals set have actually been achieved.

7.2 The general aim of this study

This study focused on the improvement of modern tools. Although it has been acknowledged that Taylor's work is a good starting point for such a study, it has been argued here that modern tools need a different approach. Taylor's approach was based on 'time and motion' studies. This involved observing people while they are working. However, as cognitive operations are not observable a different approach was therefore required. Secondly, Taylor assumed that there is 'one best way' to carry out a task which led him to prescribe which tool the worker should use and which working method. Modern work psychological insight acknowledges that there is no 'one-best-way' to do a task. Differences in working styles[1] and strategies may lead to different working methods. This should be taken into account when designing tasks and tools. A prerequisite is that the worker is in control. Having control, in this case pertains to the presence of 'decision latitude' (degrees of freedom), which means to say that the worker may make his own decisions on matters such as working method, strategy, etc. My assumption is that such an approach will be more efficient in the longer term because it enables workers to regulate their effort expenditure. This points to a third departure from Taylor's approach. Taylor focused on productivity while I have focused on 'human costs'.

1 It should be noted that working-style and strategy are not the same. Working-style can be regarded as certain heuristics that have evolved into a personal style (trait), while strategy refers to conscious choices in favour of a particular working-method.

The contribution of this study to the improvement, i.e. (re)design of modern tools, should be two-fold:

a. to develop a methodology (i.e. instrument and procedure) for the evaluation of interfaces where psychological efficiency is the main criterion;
b. to attempt to validate the design guide-lines which have been formulated within the Action Facilitation Approach.

These design guide-lines have been derived from a theory on human work behaviour (Action Theory) that puts the goal-directedness of human work behaviour in a central position. Action Theory is based on the assumption that people develop their own plan of action when they have to carry out a task. An implication of this assumption is that human beings are in control.

7.3 The sub-goals of this study

7.3.1 Measuring psychological efficiency

The evaluation methodology that has been developed in this study focuses on measuring improvement in efficiency. The concepts of efficiency that have been employed in this methodology have been described in Chapters 2 and 5. It has been argued that when human beings strive towards efficiency in their behaviour they estimate their costs according to the two criteria of time and effort. The concept of efficiency that has been applied incorporates this idea. All this resulted in the concepts of 'strategic efficiency' and 'processing efficiency' (Chapter 5). Strategic efficiency has been conceptualized as being primarily related to the choice of behaviour alternatives or as being related to the strategy that is applied. Processing efficiency has been conceptualized as being related to the amount of energy that is required to execute a particular behaviour alternative (or strategy). The respective concepts have been operationalized in terms of the 'time required to complete the task' and 'mental effort'.

Furthermore it has been argued that the effort concept is a better indicator of the psychological costs than the workload concept. Workload relates to the complexity of the demands of the task, while effort reflects the correspondence between the worker's psycho-physiological state at a particular time and the complexity of the task demands.

From Chapter 4 it can be concluded that most instruments that are used for evaluating tasks or Man-Machine systems focus on measuring the workload and do not take the differential aspect of the psycho-physiological state into account. This has resulted in the development of the Rating Scale Mental Effort.

The experimental study described in Chapter 5 has made clear that the Rating Scale Mental Effort (RSME) seems to be a valid and reliable indicator of the amount of mental effort that a person exerts when carrying out a task.

However, it has also been stated that although there is converging evidence with regard to the validity of the scale, some questions still remain not fully answered; in particular the question as to what subjects are referring to when they respond to the RSME. Whether the theories and models on that have been used are valid, is also still not completely clear. This means that the claim of validity of the RSME should be accepted with these marginal notes. It has been concluded that the score on the RSME may be regarded as an adequate estimation of the mental 'costs' that are associated with task execution. We have been able to provide an answer to the second sub-goal that was formulated in Chapter 1: to develop a procedure or instrument to measure the 'mental costs' of task execution.
The RSME score was used as an operationalization of 'processing efficiency', while time-parameters were used as indicators of 'strategic efficiency'. These concepts and operationalizations were applied in the experiments described in Chapter 6. These experiments were directed at examining: a) the validity of the efficiency concepts and b) the applicability of these concepts as evaluation criteria for system design.

The results of the experiments outlined in Chapter 6 proved that when people work twice with the same system they either complete their task in less time or they invest less effort the second time round. This illustrates that people indeed reduce their 'costs' when they are working for some period on a task or with the same system. When they maintain their output at minimally the same level they display an improvement in efficiency. As has already been mentioned in Chapter 2, this is the core of the Action Facilitation Approach and the essence of acquiring skills (cf. Leplat, 1989). It should be noted that the output level is usually prescribed by the organization. This means that in most cases workers are not allowed to decide for themselves whether they reduce this level. Therefore I have focused primarily on the psychological costs of work behaviour.
It also appeared that a reduction in the time needed to complete the task is not always accompanied by a reduction in the RSME score. This is an indication that the perception of the length of the task (time-on-task) was not decisive when the subjects evaluated the effort invested. In condition BB there were even subjects who were able to reduce the time needed to complete the task but at the expense of a higher effort expenditure. These results support the idea that people estimate their psychological costs in two dimensions and that they make trade-offs between these two dimensions. Furthermore this phenomenon may also be regarded as a contribution to the discussion in Chapter 5 on the question what do subjects refer to when they give their effort rating. Apparently this process does not only involve externally observable cues like, duration of the task.

The results of the first experiment presented in Chapter 6 demonstrate that, in situations where improvement of efficiency was predicted on theoretical

grounds (conditions AA and BB), an improvement in efficiency could indeed be measured. This can be regarded as supportive evidence for the statement that the efficiency concepts (and their operationalizations) are valid and applicable.
In particular it has proven to be useful in making a distinction between the two dimensions of 'cost': time and effort. The results of the experiments outlined in Chapter 6 indicate that people apparently differ with respect to how they estimate their costs. This is illustrated by the fact that some people prefer to work with a tool that allows them to work faster while others seem to prefer a tool that demands less effort (cf. section 6.4.7).
In Chapter 2 it was indicated that people may have different and changing, preferences with respect to their work strategy. This also seems to be true of working with tools. Some people prefer to take more time while others prefer to put more effort into a task to finish it earlier. This indicates that people make a trade-off between the two dimensions of 'costs'. A change of strategy may derive from fatigue or changes in circumstances, like the introduction of time pressure. In situations where people are under pressure time-wise (feeling rushed) they may be prepared to take risks. The example of the welder in Chapter 2 illustrates this very well. At yet other times people may want to take their time and choose an easier way in order to avoid high mental costs. The results of the experiments in Chapter 5 (particularly the unpaced condition) illustrate this.

Furthermore the results of the experimental comparison of both the word processors made it clear that subjects 'prefer to work with', the word processor that allowed them to work most efficiently. This supports the initial assumption that people strive to achieve efficiency in work behaviour and it illustrates that it is relevant to evaluate work tools on the basis of the degree to which they facilitate this endeavour.

7.3.2 The evaluation methodology

The results of the experimental comparison of both word processors (Chapter 6) made clear that, although efficiency improvement is measurable with our instruments and can therefore be used as an evaluation criterion, a specific methodological design is required for comparing and evaluating various design alternatives. The (dis)advantages of various research designs and methodologies in human factor research have been discussed at length (cf. Karat, 1988; Landauer, 1988; Drury, 1990). Experimental laboratory research like that used in this study, is generally regarded as a very effective but expensive method. Hence the reason that such research is only carried out when fundamental questions have to be answered and not when superficial aspects have to be evaluated (Karat, 1992).
In our study it also appeared that directly comparing of both word processors within one experimental session has the methodological risk of 'transfer

effects'. This means that when subjects have been working for some time with one of the systems the knowledge about how to operate that system interferes with their attempts to understand how to work with the other system.
The results of the experiments described in Chapter 6 suggest that an alternative solution can be found if the various design alternatives are compared on the grounds of efficiency *improvement* rather than on the level of efficiency per se. Such an approach does not have the methodological and practical disadvantages that were mentioned in the direct comparison between both word processors (transfer effects) because subjects do not have to be experts in operating the system (which would be very difficult with a newly designed product) or work with both systems within a relatively short time.
From a work psychological point of view there are also some arguments in favour of such an approach. It has been argued before that 'degrees of freedom', or rather 'being in control' is important for the worker from a work psychological point of view. If efficiency is to be improved there should be opportunities for making (strategical) choices between various behaviour alternatives during the work process. Increased efficiency may therefore be seen as one of the indicators that the worker is in control of his work situation.

This inference is supported by the results of the experiment described in Chapter 5. The results of that experiment demonstrated that 'control' over the speed at which the task is executed is essential in conjunction with efficient performance. The amount of control a person has over a task determines to what extent he will be able to regulate his effort expenditure and choose his own strategies. In the 'unpaced' condition where subjects controlled the speed at which the tasks were presented to them they preferred to trade off time for effort when the task became more difficult and in this way they achieved better results in terms of percentage of correct answers.
Furthermore the study in Chapter 6 revealed that there are individual differences with respect to work-styles and preferences. This can be seen as an argument that workers should be in control of their tools if they are to have their own working style and the autonomy to decide which working-methods they would employ at a particular moment. The latter aspect is important because due to fatigue or other things people may sometimes want to vary their working-method. In regulating their effort expenditure people tend to change their strategy once in a while. For instance they may trade off effort for time when they feel that they are getting tired.

To summarize the results so far we may conclude that regarding the first two sub-goals of this study, improvement in efficiency is a good criterion for evaluating design alternatives. The evaluating can be effected by utilizing the efficiency concepts developed and operationalized in this study. The

Rating Scale Mental Effort proved to be a reliable and valid instrument for measuring psychological costs. Other advantages are that it is cheap and easy to apply all of which makes it very suitable for evaluating design-alternatives.

7.3.3 Action Facilitation Approach validated ?

The results of the expert evaluation (first study in Chapter 6) demonstrated that the Action Facilitation Approach is a coherent and manageable frame of reference for evaluating interfaces. Furthermore the experimental evaluation of both word processors (second study in Chapter 6) suggests that these guide-lines are valid. The word processor that was rated by the experts to be most in compliance with the AFA guide-lines proved to be the system that allowed for greatest improvement in efficiency.

However, we should note that the improvement in efficiency that was demonstrated occurred during the subjects' first 'learning phase'. when they had just started learning how to use the word processors. In this initial learning period features like 'possibilities for orientation' which was one of the respects in which the word processors differed from each other proved very important. This means that we do not know for sure whether the results of this experiment would have revealed that experienced users as well would work more efficiently with WP A.

Skilled users are not likely to need a lot of time for orientation to get used to the system they are using, they are more likely to be in need of other features like those referred to in the Action Facilitation Approach, for instance features that allow them to change their plan of action and to change how they execute their actions or, alternatively, features that support the supervision process. These are the features that would allow skilled users to work efficiently.

The finding that WP A (the word processor that was found to comply most with AFA guide-lines) provides most space for improvement in efficiency can be seen as important supportive evidence for the validity of the Action Facilitation Approach.

I think it is fair to state that at the moment there can be no reason for claiming that the results of this experiment should not be generalized. It is not possible to indicate from our experiment which particular features of WP A were responsible for the improvement in efficiency. To establish which of the above-mentioned features are determinant further research would have to be carried out with specific groups of users (novices and skilled users alike). In such a study it would be necessary to control the presence of particular features.

Therefore the conclusion that the results of the study described in Chapter 6 can be regarded as supportive evidence for the validity of Action Theory and the Action Facilitation Approach is justifiable.

7.4 Action Facilitation Approach in perspective

Design and evaluation are two subsequent phases in the design process (cf. Eekels, 1984) so they are therefore linked together. This means that design principles are often also used as criteria in the evaluation of a product. From the point of view of consistency in design methodology it is also advisable to adopt this policy. The design process usually starts with an extensive analysis of the task for which the tool is to be used so that the requirements of the product can be specified. Therefore methods of task analysis take a prominent place in interface evaluation.

Evaluation methods cover a wide range of techniques (cf. Gould, 1988). In general a distinction can be made between two broad categories (cf. Karat, 1988): those that are based on observations and measurements of users working with a real (or prototype) system and those which are based on an analysis of the task to be performed.

Evaluation methods which are based on task analysis assume that understanding the task elements provides a good enough basis for making design decisions. Formal methods in this category are also associated with theories of human cognition. These methods are characterized by the assumption that it is possible to put forth a rule-based design guide for general purpose systems on the basis of cognitive principles (cf. Chapter 2). An influential design approach in this tradition is the GOMS approach of Card, Moran and Newell (1983) together with the related production system analysis (Kieras and Polson, 1985, Polson, 1987). The essence of the approach adopted by Card et al. may be formulated thus: 'understanding the requirements of the task is the key to understanding the behaviour'. Therefore the design process should begin with an attempt to understand what it is that the user must do to be able to use the system to accomplish the goal he has in mind. Such analysis is called 'task analysis'. The task is then broken down into a series of cognitive and motor components. The GOMS framework provided Card et al. with the basis for predicting user performance (i.e. which methods users will follow when carrying out a particular action) and how long a specific action will take. Comparisons between design alternatives can then be made by going on the performance of 'ideal' (or expert) operators (i.e. problem solving behaviour is non-existent or minimal and no errors are made).

It is assumed that this 'expert' behaviour will not require any problem solving behaviour. It is presumed that the user will simply retrieve the correct plan from memory or make no mistakes when following instructions but herein lies also also its weakness. According to Card et al. the GOMS model is directed at the skilled and routinized behaviour of 'experts'. The model is based on the ideal that the optimal action sequence for performing a given task has been established by experts. It implicitly rests on the underlying assumption (the old Tayloristic notion) that there is 'one-best-way' to carry out a task upon which the GOMS-model is built. In reality different experts may have different action sequences for the same task, they may also use different strategies at different times. Expert behaviour would be better characterized as the ability to apply the most appropriate task strategy in a given situation (Saltzman et al., 1987; Leplat, 1989). This is made possible by another quality peculiar to experts: their ability to anticipate better than others the tasks that are to come. These aspects are not taken into account in the GOMS model. Moreover the model cannot predict the behaviour of beginners nor anticipate their learning process. The model has mainly been applied to the simple operations of text correction which can be easily broken down into well-separated 'unit-tasks'. Therefore the main criticism with respect to the GOMS model is that it does not account for actual work behaviour. Real life tasks require various combinations of problem solving behaviour and routinized procedures.

Other criticism (cf. Olson, 1987; Olson & Olson, 1990) relates to the fact that the GOMS approach ignores the aspects of user fatigue and mental workload. These aspects are important determinants in the adoption of different processing and work strategies.

The criticism that 'learning' or 'problem solving' behaviour is not taken into account has led to new developments with respect to the GOMS approach (e.g. Polson and Kieras, 1985). This study employs a model of the user that is based on production systems. A production system is composed of a number of 'rules' or 'productions' that represent knowledge. Each 'rule' is formed from a condition and an action which is accepted if the condition is found to be true. The system carries out its activities by going through a series of test-action cycles in which the environment (including working memory) is checked to see if any 'rule' conditions are met. If this is the case then the action associated with the relevant 'rule' is carried out.

Predictions of ease of learning are made by counting the 'rules' contained in the model. Performance predictions are obtained by counting 'production' cycles when tasks are being carried out.

This approach is also based on an extensive analysis of the task that has to be carried out. The underlying assumption is that it must be possible to consider all possible user responses to given situations and to base design decisions upon these responses. In my opinion this is an illusion: error behaviour, for instance, cannot be adequately allowed for in such an approach because especially when correcting errors people can apply a large variety of strategies. Therefore I think we must rely on 'user based evaluations'.

The first mentioned category of evaluation methods focuses on work behaviour: the aim is to try to establish the user's reactions and responses to the task situation (of which the system is a part). The method employed in this study exemplifies this approach. As was indicated in Chapter 2 Action Theory and the Action Facilitation Approach are theoretical frameworks that are mainly concerned with the optimization of work behaviour. It was from this perspective that criteria were formulated for task and system design (cf. Chapter 2). An essential assumption in Action Theory about the goal-directedness of behaviour is that people formulate their own goals and consequently decide upon their own behaviour. This assumption is reflected in the guide-lines that have been formulated in the Action Facilitation Approach and it should enable the worker to have control over his task situation. The essence of these design guide-lines is that designers should try not to design work processes in too much detail because this could lead to prescribing work behaviour and reducing a worker's 'decision latitude'. The results of the study in the previous chapter can, despite the small statistical evidence, be regarded as supportive evidence that following the Action Facilitation Approach design guide-lines may lead to improved efficiency in work behaviour.

The fundamental difference with respect to 'task based' evaluation methods is that those methods take the task as a starting point, while the Action Facilitation Approach (and Action Theory) considers the task to be just one of the determinants of work behaviour. This means that in order to optimize work behaviour the task itself might have to be changed as well. Focusing on work behaviour, understanding how it is regulated and how it can be improved, does not mean having to predict all the possible reactions of a worker to a (task) situation beforehand. Work behaviour itself is the object of the study and most important of all, work behaviour is not viewed as purely a reaction to the work situation. It is assumed that a worker also creates and influences the situation in which he works. For instance, by choosing different working-methods or strategies a worker may be able to change the task and its demands (cf. Sperandio, 1978; Teiger, 1978). Subsequently this may also change the requirements of the tools that are needed to carry out that task.

This means that designers should be aware of the fact that when they design tools (systems and/or interfaces) they are actually trying to design work behaviour. It has been pointed out throughout this study that work behaviour is dynamic, which means that static design solutions can never be optimal. Furthermore it was found that people differ in their preferences, working styles and strategies. Designers should take this into account and allow for various interaction styles to be applied by the users of their systems. The 'Differential Dynamic Design' approach of Ulich (cf. Chapter 6) takes this perspective into account. Applying the AFA guide-lines to system and task design may also contribute to optimizing work behaviour.

The extent to which it is made impossible for people to work efficiently is very important, not only with respect to the work process, but also with respect to the worker. If disruptions or hindrances at work prohibit efficient behaviour

this can lead to frustration and may induce stress in the work situation (cf. Semmer, 1984; Schönpflug, 1985; 1986[b]).
The value of the Action Facilitation Approach should be considered in this light. The guide-lines formulated within the AFA are directed towards removing or avoiding 'obstacles' in the design of systems that prevent workers from working efficiently.
As Roe and Meijer (1990) have already pointed out this notion of Action Facilitation can also be extended to task design. This is important for preventing stress reactions and for increasing the well-being of the individual worker.

7.5 Suggestions for further research

Although the theme of this study: improving work tools, is not new, the suggested approach - taking efficiency in work behaviour as a criterion - constitutes a departure from the traditional approach to system design. The results of the study discussed in Chapter 6 seem to confirm that the suggested way of facilitating actions is valuable. However, further research will be needed if this approach is to be developed and examined in full.
With respect to the Action Facilitation Approach it would be worth studying the value of the design guide-lines by comparing a system that has been designed according those specific guide-lines with another system. The study described in this book may give some clues but it cannot of course provide a definite answer to the question of the validity of the guide-lines mainly because, as yet, no system has been built according to these design guide-lines. This would be the first requirement and it would also give some indication of the problems designers might encounter in translating these guide-lines in design solutions. In this respect some experience was gained from a course in which the AFA was presented to a group of designers of application software (Arnold, & Boogert, 1987). It appeared that the design guide-lines, though they were evaluated rather globally, were considered to be very valuable and useful.
With respect to system design it would be valuable to do further research into the differences in work styles between novices and experienced users (experts) and to find out which factors contribute most to their (improvement in) efficiency. This would help us to understand which features of the system could help to improve to efficiency improvement and to find out what are the main obstacles that stand in the way of efficiency improvement.

The experiments on the validation of the Rating Scale Mental Effort revealed that it may be possible to 'measure' psychological costs but showed that the underlying process is still not clear. We still do not know precisely what people are referring to when they estimate their costs. We do not know whether they are referring to their own experiences related to some internal set point, and if they are, how such a process works, or whether they estimate their costs according to external observable cues. Further research into the validity of the models and theories of mental effort would also have to be undertaken.

7.6 Conclusion

In this study I have tried to present a scientific contribution to the (re)design of modern tools. The aim has been to argue that 'work behaviour based' methods for evaluation and design have some advantages over 'task based' methods. Work behaviour oriented design principles are more likely to result in systems that people can operate efficiently. Task based methods may easily lead to prescribing working methods.
The results of this study indicate that the Action Facilitation Approach may be regarded as useful in this respect although further research is needed to examine the full scope of its value and applicability.

Two lines of research have been presented in this study: an experimental psychology approach and a work psychology approach. The experimental psychology part of the study contains the development and experimental validation of the Rating Scale Mental Effort. This part of the study had a more scientific explanatory orientation.
The evaluation study in Chapter 6 was generally designed to view things from a work psychology angle. In this chapter the objective of the study (i.e. to evaluate tools) was placed in a technical context and again an experimental approach was chosen. Through its very set-up this experiment enabled us to examine (to an extent) the validity of a theoretical model of work behaviour: Action Theory and Action Facilitation Approach.
Although this part of the study can be characterized as being primarily of a technological rather than of a scientific explanatory nature, it illustrates that an experimental approach within a work psychological context is rather fruitful. Therefore the study described in the second part of this book is an example of what may be called 'experimental work psychology'. Simulating real work situations in a laboratory setting offers a good opportunity for studying a broad range of applied and fundamental questions in the domain of work psychology.

As to the overall goal of this study, a contribution that the improvement of modern tools, it has been demonstrated that the work psychological approach to system design, as operationalized in the Action Facilitation Approach, is a useful approach. The Action Facilitation Approach focuses on removing obstacles that hamper efficient working behaviour, this not only leads to the improvement of tools but it may also may increase the well-being of the individual worker.

Summary

The study described in this book aims at presenting a scientific contribution to the (re)design of modern tools, i.e. personal computers. Frederic Winslow Taylor may be rightly viewed as the first person to make the improvement of tools the subject of a systematic study. Taylor observed people while they were working and using tools. He attempted to find the optimal way to carry out various tasks by scrupulously studying the movements that the workers made and by noting how long these movements took (time and motion studies).
In *Chapter 1* of this book it is argued that although Taylor's work may be seen as a basis for this study his approach is no longer valid. Taylor focused on overt behaviour and neglected relevant (non overt) psychological processes. These processes are particularly relevant with modern (cognitive) tools like computers. Another argument is that, according to modern work psychological insights, there is no 'one-best-way' to carry out a task. Today it is assumed that workers should have control over their work situation, in particular where the choice of working methods and working strategies is concerned.
The suggestion is that we should adopt Taylor's idea of taking efficiency as a criterion when evaluating tools but emphasize the aspect of the (psychological) costs involved when people are at work or using tools. In other words it is assumed that a particular tool is better than an alternative one if it allows the worker to work more efficiently, i.e. to perform at least as well but at lower (psychological) cost. This notion is also at the heart of the Action Facilitation Approach (AFA), an approach that has been recently developed for improving interface designs.
The study in this book is designed in such a way that it can be regarded as an initial attempt to validate the Action Facilitation Approach.

In *Chapter 2* some general principles of behaviour economics are described and the concepts of psychological costs and psychological efficiency are developed. Psychological efficiency relates to the amount of effort that, as far as an individual can see, has to be invested in order to accomplish a certain goal.
The next step is to check whether these concepts can be incorporated into an existing body of theory on work behaviour. Hacker's Action Theory was chosen because at present it is considered to be the most complete theory on human work behaviour. This theory presents a set of views on the organization of human work behaviour and on the psychological processes that regulate this behaviour. Hacker assumes that work behaviour is goal-directed behaviour which results from a person's active and conscious decisions to act. This would imply that a task first has to be internalized and redefined as an individual and personal task. Such a personal task provides the motivation that constitutes the basis upon which a person decides to act. This implies that at the same time there is no 'one-best-way' to carry out a task and that the principle of 'equi-finality' is the leading principle.

The earlier mentioned Action Facilitation Approach has its roots in Action Theory. The AFA consists of 1) a set of principles of human action, derived from Action Theory, that serve as a model for the human worker, 2) a set of guide-lines based on these principles of human action that indicate dimensions of support and 3) a set of interface design recommendations that serve to operationalize each of the support dimensions. It is hypothesized that implementing these design guide-lines will lead to 'Action Facilitation'. Action Facilitation refers to supporting the worker in carrying out his task and it may be operationalized to improve or maintain performance at a lower level of cost to the individual.

Chapter 3 is devoted to certain theoretical viewpoints on mental effort. First of all it is argued that mental effort is a better indication of the psychological costs than the concept of workload. Effort takes the more dynamic aspects of costs into account while workload is considered to be a rather static concept. Workload is exclusively related to the demands of the task while effort also accounts for the performance potential of the worker. Performance potential in particular is susceptible to change. Another aspect is that effort investment is considered to be the result of a conscious decision to invest effort in order to carry out a particular task. This implies motivational aspects as well.
Two approaches to the concept of mental effort are discussed. The first approach stems from cognitive psychological research on the structure of human information processing. It is assumed that there is a central mechanism that controls the processing of information by focusing attention. The implication is that the more attention is needed to process the information the more effort is needed. So, mental effort is considered to be directly related to attention demanding information processing.
The second approach to the concept of mental effort focuses on the psycho-physiological state of the worker. The psycho-physiological state refers (amongst other things) to the degree to which a person is activated (cf. the difference between sleeping and being awake). Fatigue can be described as a particular psycho-physiological state, another such example is: one's state after consuming alcohol. These states influence the processing of information as one might imagine. Therefore one can also refer to these as 'cognitive states'. If there is a discrepancy between the actual state of the worker and the state that is required by the demands of the task the worker has to invest effort in order to compensate for the deficit. More recent models of mental effort try to integrate both approaches to mental effort. Such an integrated model is being used in the following research that aims at measuring the concept of mental effort.

Several methods and instruments for measuring mental workload and mental effort are discussed in *Chapter 4*. The conclusion is that all the instruments that have been mentioned actually focus on the (static) concept of workload. The construction and development of a new rating scale to measure mental effort is subsequently described. To construct this rating scale the 'magnitude estimation' method has been used. This method is based on estimating the ratio between a

particular stimulus (or item) and a given standard stimulus in order to determine the scale values of the stimuli or items.
The resulting scale (the Rating Scale Mental Effort - RSME) was subsequently used in a research project in which the workload of bus drivers was studied. This study provided initial information about the validity and reliability of the scale. The scale proved to be very sensitive to changes in task demands and changes in the psycho-physiological states (performance potential) of the bus drivers. This can be seen as an indication that the RSME indeed measures effort, as was envisaged in the integrated models of mental effort. Furthermore the RSME appears to provide very reliable estimations of differences in task demands both in laboratory research situations and in field studies.

The validity of the Rating Scale Mental Effort is explicitly examined in the laboratory study described in *Chapter 5*. For this purpose the performance potential of the subjects and the demands of the task were manipulated. The mode of information processing and the degree to which subjects had control over their task situation was manipulated. 'Mode of information processing' refers to the distinction between controlled and automatic processing of information. This distinction has been made because controlled processing is assumed to be effort demanding and automatic processing is assumed to require hardly any effort. An additional argument is that, according to various researchers, subjective methods like the RSME would not be sensitive to such a distinction.
The results of the experiment clearly show that the RSME scores discriminate -in line with the expectations formulated in hypotheses - between the various experimental conditions that are distinguished. The RSME scores are higher when subjects are required to carry out a particular task after they have been exposed for a long period to a similar task demand than when they are placed in situations where they have just started to work on that particular task. This indicates that subjects have to compensate in order to adjust their actual state (performance potential) to the state that is required by the task demands. Furthermore it appeared that the duration of a task (time-on-task) also influences the amount of effort that is reportedly expended.
It appeared that the degree to which subjects could control the pacing of the task influenced their level of effort investment. When they had the opportunity to control the speed of presentation of the stimuli subjects took longer to respond to the more difficult tasks, but they did not report increasing the amount of effort investment. This contrasted with situations in which subjects did not control the speed of presentation of the stimuli. In these situations reaction times and levels of effort investment increased when tasks became more difficult.
These results show that depending on the degree of control they have over their task situation people employ various strategies in order to regulate their effort investment (or rather their psychological costs). Making a trade- off between time and effort can be regarded as such a strategy. People can decide to work a little slower in order to keep their effort expenditure within (what are for them) acceptable limits. Time can also be regarded as an aspect of the costs. Conse-

quently it may be postulated that the concept of psychological costs has (at least) two dimensions: time and effort.
Further evidence for the validity and reliability of the RSME is found in other studies in which the RSME has been used. Several of these studies are briefly described. The conclusion of this chapter is that, according to the integrated model of mental effort, the RSME is a valid and reliable instrument that seems to be very sensitive to changes in task demands and the performance potential of the worker. It also appeared that the RSME was able to discriminate between changes in the mode of information processing (controlled versus automatic processing). It has to be said that this asset of sensitivity does not really benefit the diagnostic characteristics of this instrument.

In *Chapter 6* a study is described in which the RSME is applied in an interface evaluation study. This study aims at collecting empirical evidence for the validity of the Action Facilitation Approach.
For this purpose two word processors, which were presumed to differ in certain relevant AFA respects were selected. The appropriateness of the choice was checked and confirmed by means of expert evaluation of both the word processors. Subsequently both word processors were further evaluated in an experimental study. The hypothesis was that the word processor that was judged to comply most with the AFA dimensions would be the best for enabling subjects to carry out their task efficiently. Efficiency concerns the ratio between benefits and costs. If a task is adequately executed and results in the required output this is termed the 'benefits' of work behaviour. Psychological cost is conceived as consisting of the two (not completely independent) dimensions (cf. Chapters 2 and 5) of time and effort.
In the experiments both a within-subjects design and a between-subjects design was applied. Several methodological reasons like, the risk of transfer effects and small groups of subjects have been mentioned to justify such an approach.
The results do indeed show that transfer effects very probably influenced the measurements in the conditions where subjects worked consecutively with both word processors (the between-subjects design). It was concluded that a between-subjects design should not be used in a comparative evaluation of various interfaces. On the other hand it appeared from this comparison that subjects expressed their preferences for the interface (system) that allowed them to work more efficiently, i.e. to complete their task quicker or with less effort or both. Furthermore there are indications that individual preference influences the choice of a particular work strategy. It seems that some strategies are more successful with one interface while other strategies seem to thrive more with another interface. The results of the between-subjects design showed that the word processor that was evaluated as complying most with the AFA dimensions enabled subjects to reduce their psychological costs significantly (i.e. finish the task quicker or with less effort or both) than the subjects who worked with the other word processor. This is an indication that the first word processor has more potential for efficiency improvement. This is regarded as supportive evidence with respect to the validity of the Action Facilitation Approach.

For designers of systems this study first of all demonstrates that efficiency can be used as an evaluation criterium for system design and secondly, that applying the AFA may lead to developing systems that will enable people to work more efficiently.
Another relevant conclusion is that individual differences with respect to work styles and working methods appear to be very important. This indicates what kind of urgency there is for new flexible systems which each individual can easily adapt to fit his own purposes. This helps the worker to regulate the level of his psychological costs. The 'Differential Dynamic Design' approach could be useful in this respect.

Chapter 7 contains the conclusions that can be drawn from the studies described in the previous chapters. The most important conclusions are that the RSME appears to be a reliable and valid instrument for measuring (an aspect of) the psychological costs and that there are indications that applying the AFA guide-lines indeed results in developing systems that enable workers to work more efficiently. Additionally the AFA is placed in a wider perspective. For these purposes it was compared to other approaches in order to develop design guide-lines, such as the Goals, Operators, Methods, and Selection rules (GOMS) approach. A distinction is made between so-called 'task oriented approaches' and so-called 'behaviour oriented approaches'. The first mentioned approach assumes that when the task is known, by means of detailed analysis of the task, design guide-lines can be formulated. It is argued that such an approach is in essence the same as Taylor's approach which was criticized in Chapter 1. It implicitly assumes that there is 'one-best-way' to carry out a task.
The second approach focuses on work behaviour, or rather, the actual usage of the system under (if possible) real conditions. This approach acknowledges the principle of equi-finality, and also that prescribing the working methods reduces a worker's amount of freedom. In the long run this may appear to be less efficient because workers may be less able to regulate their psychological costs.
The value of the Action Facilitation Approach is that it removes obstacles that prevent workers from working efficiently, which is something that contributes to an individual's well-being.
This chapter concludes with several suggestions for further research.

Samenvatting

Efficiëntie in arbeidsgedrag: een ontwerpbenadering voor modern gereedschap

De hier beschreven studie heeft primair tot doel een wetenschappelijke bijdrage te leveren aan het (her)ontwerpen van moderne gereedschappen, in casu personal computers. Frederic Winslow Taylor kan beschouwd worden als de eerste die het verbeteren van gereedschap onderwerp maakte van systematische studie. Het werk van Taylor kan gezien worden als een uitgangspunt voor deze studie. Hij observeerde mensen terwijl ze de gereedschappen gebruikten. Door middel van nauwgezette tijd- en bewegingsstudies trachtte hij de beste manier te vinden om een bepaalde taak uit te voeren.
In het *eerste hoofdstuk* wordt beargumenteerd dat Taylor's benadering naar moderne arbeidspsychologische inzichten niet meer voldoet, omdat de benadering zich richt op observeerbare gedragingen en derhalve vele relevante psychologische (niet observeerbare) processen, die met name bij moderne gereedschappen een belangrijke rol spelen, buiten beschouwing laat. Een andere reden voor het zoeken naar een andere benadering is het feit dat Taylor's benadering leidt tot het voorschrijven van werkmethoden en -strategieën (het idee van de 'one-best-way'). Volgens huidige A&O-psychologische inzichten is het beter dat mensen controle hebben over hun eigen werksituatie, waarbij met name de keuze van de werkmethoden en -strategieën erg belangrijk is.
Voorgesteld wordt om, in navolging van Taylor, efficiëntie als beoordelingscriterium voor gereedschap te nemen, maar daarbij de nadruk te leggen op de (psychologische) kosten die gepaard gaan met het uitvoeren van taken en het werken met de betreffende gereedschappen. Met andere woorden een gereedschap is beter dan een ander indien het mensen in staat stelt efficiënter te werken, d.w.z. minimaal dezelfde prestatie tegen lagere (psychologische) kosten.
Dit uitgangspunt is tevens de kern van de Action Facilitation Approach (Handelingsfacilitatie benadering), een nieuwe benadering voor het opstellen van ontwerprichtlijnen voor interfaces. De opbouw van de hier beschreven studie is zodanig gekozen dat de studie beschouwd kan worden als een initiële poging tot validering van de Action Facilitation Approach.

De studie is als volgt opgebouwd: in *hoofdstuk 2* wordt aandacht besteed aan algemene principes van 'behavior economics' en de ontwikkeling van de concepten 'psychologische kosten' en 'psychologische efficiëntie'. Psychologische efficiëntie heeft betrekking op de hoeveelheid inspanning die, volgens eigen perceptie, geleverd moet worden om een bepaalde prestatie te leveren.
Vervolgens wordt nagegaan in hoeverre dergelijke concepten aansluiten bij een bestaande theorie met betrekking tot arbeidsgedrag. Hierbij is gekozen voor Hacker's Handelingstheorie, omdat deze als de meest complete theorie over arbeidsgedrag wordt beschouwd. De theorie gaat met name over de organisatie

van het (arbeids)gedrag en de psychologische processen die dit gedrag controleren. Uitgangspunt is dat arbeidsgedrag is gebaseerd een actieve beslissing van de persoon zelf om te handelen. Dit betekent dat een opdracht eerst geinternaliseerd moet worden en geherdefinieerd tot een eigen individuele opgave. Deze individuele opgave vormt de basis voor het handelen. Dit impliceert tevens dat er geen 'one-best-way' is om een bepaalde taak uit te voeren, maar het principe van equifinaliteit vormt het leidende principe.
De eerder genoemde Action Facilitation Approach is gebaseerd op principes uit de Handelingstheorie. De AFA bestaat uit richtlijnen, gebaseerd op handelingstheoretische principes, voor het ontwerpen van interfaces. De veronderstelling is dat implementatie van deze ontwerpaanbevelingen zal leiden tot 'action facilitation', d.w.z. dat het hulpmiddel de werker ondersteunt bij de uitvoering van zijn taak. De operationalisatie hiervan is verbetering, of tenminste handhaving, van het prestatieniveau tegen lagere individuele kosten.

In *hoofdstuk 3* wordt nader ingegaan op het begrip 'mental effort' (mentale inspanning). Beargumenteerd wordt dat het begrip inspanning een betere indicatie is voor de psychologische kosten dan het vaak gebruikte begrip arbeidsbelasting. Inspanning doet meer recht aan de dynamische aspecten dat inherent is aan het begrip kosten. Terwijl het begrip arbeidsbelasting voornamelijk gerelateerd is aan de zwaarte van de taakeisen, is het begrip inspanning gerelateerd aan de taakeisen in verhouding tot het prestatie potentieel (beschikbare werkcapaciteit) van de taakuitvoerder. Met name het prestatie potentieel is aan veranderingen onderhevig. Een ander aspect is dat het leveren van inspanning verondersteld wordt het gevolg te zijn van een bewuste beslissing tot het leveren van inspanning om een bepaalde taak uit te voeren. Dit houdt in dat er ook sprake is van een motivationele component, hetgeen goed aansluit bij de handelingstheoretische uitgangspunten.
Vervolgens worden twee verschillende benaderingen met betrekking tot het begrip mentale inspanning besproken. Eén benadering is gebaseerd op het cognitief psychologisch onderzoek naar de architectuur van het menselijk informatieverwerkingsproces. Hierbij is de aanname dat er een centraal mechanisme is dat de verwerking van informatie stuurt, door middel van het richten van de aandacht. De implicatie is dat naarmate er meer aandacht nodig is voor het verwerken van informatie er ook meer mentale inspanning gevergd wordt. Mentale inspanning is hiermee dus rechtstreeks gekoppeld aan aandachtvragende informatieverwerking. De tweede benadering van mentale inspanning is vooral gericht op het in overeenstemming brengen van de actuele psycho-fysiologische toestand van de taakuitvoerder met de toestand die voor de betreffende taak vereist is. Deze psycho-fysiologische toestand heeft o.a. betrekking op de mate waarin personen geactiveerd zijn (vgl. verschil tussen slapen en wakker zijn) en heeft ook invloed op het functioneren van het informatieverwerkingsproces. Vermoeidheid kan ook worden aangeduid als een bepaalde psycho-fysiologische toestand, evenals de toestand na alcohol consumptie. Elk van deze toestanden heeft een invloed op het informatieverwerkingsproces, er kan derhalve ook worden gesproken van 'cogni-

tive states'. Indien er een discrepantie bestaat tussen actuele toestand en vereiste toestand dient er inspanning geleverd te worden om voor deze deficiëntie te compenseren.
Recentere modellen gaan uit van een integratie van beide benaderingen. Zo'n geïntegreerd model wordt in het navolgende onderzoek, dat gericht is op het meetbaar maken van mentale inspanning, dan ook gehanteerd.

In *hoofdstuk 4* wordt, nadat enkele methoden voor het meten van mental workload en mental effort zijn genoemd en besproken, geconcludeerd dat genoemde instrumenten meer op het (statische) begrip workload zijn gericht dan op effort. Derhalve wordt de ontwikkeling en constructie van een eigen beoordelingsschaal voor mentale inspanning besproken. Hierbij is gebruik gemaakt van de 'magnitude estimation' methode, een methode waarbij het schatten van de verhouding van een bepaalde stimulus met een vooraf gekozen standaard stimulus gebruikt wordt om de schaalwaarden van de betreffende stimuli (of items) te bepalen.
De resulterende schaal is vervolgens gebruikt in een onderzoek naar de werkbelasting van buschauffeurs. Dit onderzoek levert initiële informatie ten aanzien van de validiteit en betrouwbaarheid van de schaal. Het blijkt dat de schaal zeer gevoelig is voor veranderingen in taaklast en veranderingen in de psychofysiologische toestand (performance potential) van de persoon. Dit wijst er op dat de schaal inderdaad effort meet, zoals dat volgens het geïntegreerd effort-model wordt geconceptualiseerd. Tevens blijkt het instrument zowel in laboratoriumsituaties als in veld onderzoek betrouwbare beoordelingen te leveren van verschillende taaklast niveaus.

De validiteit van de Beoordelingsschaal Mentale Inspanning (BSMI) wordt middels een, in *hoofdstuk 5* beschreven, laboratorium onderzoek expliciet onderzocht. Hiertoe worden, met als uitgangspunt het geïntegreerde model voor mental effort (Mulder, 1986), zowel het prestatie potentieel als het taaklast niveau experimenteel gemanipuleerd. Daarnaast worden de wijze van informatieverwerking en de mate van controle over de taakuitvoeringssnelheid gemanipuleerd. De wijze van informatieverwerking refereert aan het onderscheid tussen bewust gestuurde informatieverwerking versus automatische verwerking van informatie. Dit onderscheid is gebruikt omdat de eerste manier van informatieverwerking inspannend geacht wordt te zijn, terwijl automatische verwerking verondersteld wordt geen inspanning te kosten. Een bijkomend argument is dat volgens diverse auteurs subjectieve meetinstrumenten, zoals de BSMI, niet gevoelig voor een dergelijke onderscheid zouden zijn.
De resultaten van het experiment laten duidelijk zien dat de BSMI-scores aan de (in hypothesen verwoorde) verwachtingen voldoet. De BSMI-scores zijn hoger indien personen een bepaalde taak uitvoeren, nadat ze reeds in een voorafgaande periode, inspannende taken hebben uitgevoerd. Hetgeen wijst op het feit dat personen moeten compenseren voor het feit dat hun actuele prestatie potentieel niet geheel in overeenstemming is met het niveau dat nodig is om de betreffende taak naar behoren uit te voeren. Tevens blijkt ook dat de hoeveelheid tijd die

personen dienen te besteden aan de taak (time-on task) van invloed is op de te leveren hoeveelheid inspanning.
Daarnaast blijkt dat de mate waarin personen invloed kunnen uitoefenen op de snelheid van het taakverloop van invloed is op de mate van inspanning die personen zeggen te leveren. Indien de mogelijkheid bestaat nemen personen bij moeilijkere taken meer tijd om de taak (beter) uit te voeren terwijl het inspanningsniveau gelijk blijft in vergelijking met een makkelijkere taak. Terwijl in situaties waarin personen geen invloed hebben op de snelheid van taakverloop er een toename is van het inspanningsniveau als de taak moeilijker wordt.
Deze resultaten geven aan dat personen, afhankelijk van de mate van controle die ze hebben op de taaksituatie, verschillende strategieën hanteren waarmee ze de hoeveelheid inspanning (oftewel de psychologische kosten) die ze moeten leveren reguleren. Een dergelijke strategie is bijvoorbeeld het uitruilen van inspanning voor tijd: met andere woorden langzamer werken, waardoor de hoeveelheid te leveren inspanning lager kan blijven. De hoeveelheid tijd kan ook als een 'kosten' factor worden beschouwd. Als gevolg hiervan kan men stellen dat het begrip psychologische kosten tenminste twee dimensies kent: tijd en inspanning.
Aanvullende evidentie voor de validiteit en betrouwbaarheid van de BSMI kan worden gevonden in ander onderzoek waarin de BSMI is gebruikt. Enkele van deze onderzoeken worden in dit verband kort besproken.
De conclusie van dit hoofdstuk is dat, in het licht van het gehanteerde model van mentale inspanning, de BSMI een betrouwbaar en valide instrument blijkt te zijn, dat zeer gevoelig is voor veranderingen in taaklast niveau en veranderingen in prestatie potentieel van de taakuitvoerder. Tevens blijkt ook dat veranderingen in wijze van informatieverwerking tot uitdrukkingen komen in de BSMI-scores. Opgemerkt moet worden dat een dergelijk ruim meetbereik ten koste gaat van de diagnostische eigenschappen van het instrument.

In de in *hoofdstuk 6* beschreven studie wordt de BSMI toegepast in de context van een interface-evaluatie onderzoek. Doel van dit onderzoek is empirische evidentie te verzamelen voor de validiteit van de eerder genoemde Action Facilitation Approach (AFA).
Hiertoe werden twee tekstverwerkers geselecteerd die op enkele relevante AFA-dimensies van elkaar verschillen. De juistheid van de keus werd middels het oordeel van acht experts gecontroleerd en bevestigd.
Vervolgens zijn beide tekstverwerkers in een experimentele vergelijking onderzocht. De hypothese was dat de tekstverwerker die het meest in overeenstemming is met de in AFA geformuleerde ontwerp richtlijnen personen het best in staat zou stellen hun taak efficiënt uit te voeren. Efficiëntie heeft betrekking op de verhouding tussen baten en kosten. De juiste uitvoering van de taak, resulterend in een output, wordt beschouwd als de opbrengst van arbeidsgedrag. Ten aanzien van de psychologische kosten van arbeidsgedrag wordt de in hoofdstuk 2, en in hoofdstuk 5 verder, ontwikkelde conceptualisering gehanteerd. Hierbij wordt het begrip psychologische kosten opgevat als bestaande uit twee (niet geheel onafhankelijke) dimensies: tijd en inspanning.

Ten aanzien van de uitvoering van het experiment is een combinatie van een 'binnen personen' en een 'tussen personen' design gebruikt. Hiervoor worden verschillende methodologische argumenten aangevoerd, variërend van beperking van het risico van 'transfer effects' tot de mogelijkheid om met kleine groepen proefpersonen toch nog significante verschillen te kunnen vinden.
Met betrekking tot de resultaten blijkt dat er inderdaad aanwijzingen zijn dat de tussen personen metingen beïnvloed zijn door 'transfer effects' en derhalve voor een vergelijkende evaluatie minder geschikt blijken te zijn. Wel blijkt uit dergelijke vergelijkingen dat personen een voorkeur uitspreken voor die tekstverwerker die hen in staat stelt efficiënt te werken, dat wil zeggen winst te boeken in tijd, of inspanning, of beide. Voorts zijn er aanwijzingen dat individuele preferenties een grote invloed hebben op de keuze van een werkstrategie, en dat sommige strategieën waarschijnlijk succesvoller zijn met het ene interface, terwijl andere strategieën wellicht succesvoller zijn met een ander interface.
De resultaten van het binnen persoon design laten zien dat de tekstverwerker die eerder werd aangeduid als zijnde het meest in overeenstemming met de AFA-richtlijnen, personen beter in staat stelde hun efficiëntie te verbeteren. Indien personen meerdere malen met de betreffende tekstverwerker een standaard taak uitvoerden, bleken zij (significant) beter in staat hun (psychologische) kosten te verminderen (d.w.z. sneller, of met minder inspanning, of beide hun taak volbrachten) dan personen die met de andere tekstverwerker die taak uitvoerden. Dit vormt een aanwijzing dat op z'n minst kan worden gesteld dat de met de AFA-richtlijnen in overeenstemming zijnde tekstverwerker meer potentie tot efficiëntie verbetering heeft.

Deze bevindingen kunnen worden opgevat als een initiële aanwijzing met betrekking tot de validiteit van de Action Facilitation Approach.
Voor de ontwerpers van systemen betekent dit dat, naast het feit dat aangetoond wordt dat efficiëntie als ontwerp-criterium kan dienen, het toepassen van de richtlijnen van de Action Facilitation Approach leidt tot systemen die mensen in staat stellen efficiënter te werken. Een andere, voor ontwerpers relevante, bevinding is het feit dat individuele verschillen met betrekking tot werkstijlen, -methoden, etc. van belang blijken te zijn. Dit wijst op de noodzaak om tot flexibelere systemen te komen, bijvoorbeeld door middel van 'differential dynamic design'. Gebruikers zouden meer in staat moeten worden gesteld om een bepaald hulpmiddel (gereedschap) aan te passen aan hun eigen voorkeuren en werkstijlen. Dit komt tegemoet aan de uitgangspunten die reeds in hoofdstuk 1 (en ook 2) werden verwoord, waarin werd gesteld dat taakuitvoerders zoveel mogelijk in staat gesteld moeten worden om hun eigen werkmethoden en -strategieën te kiezen, omdat dat hun beter in staat stelt hun psychologische kosten niveau te reguleren.

Hoofdstuk 7 vormt het slot hoofdstuk. Hierin worden de conclusies uit de eerder beschreven studies samengevat en op een rij gezet. De belangrijkste conclusies zijn dat de Beoordelingsschaal Mentale Inspanning (BSMI) een betrouwbaar en valide instrument blijkt te zijn om (een dimensie van) psychologische kosten te

meten, en dat er aanwijzingen zijn dat de Action Facilitation Approach een valide benadering lijkt te zijn.
Vervolgens wordt de AFA vergeleken met andere benaderingen voor het opstellen van ontwerp-richtlijnen, waaronder de Goals, Operators, Methods, and Selection rules (GOMS) benadering. Hierbij wordt onderscheid gemaakt tussen zg. 'taak georiënteerde benaderingen' en 'gedrag georiënteerde benaderingen'. De eerste benadering gaat ervan uit dat indien de uit te voeren taak bekend is, door middel van gedetailleerde analyse van de taak, ontwerp richtlijnen voor systemen zijn op te stellen. Beargumenteerd wordt dat een dergelijke benadering te veel lijkt op de in hoofdstuk 1 genoemde benadering van Taylor, en impliciet aanneemt dat er een 'one-best-way' bestaat om de taak uit te voeren. De tweede benadering richt zich op bestudering van arbeidsgedrag, oftewel het gebruik van het ontworpen systeem onder (zo mogelijk) reële omstandigheden. Deze benadering onderkent dat er meerdere manieren bestaan om een taak uit te voeren en dat het vooraf vastleggen van een bepaalde uitvoeringswijze van een taak de vrijheidsgraden van de taakuitvoerder inperkt. Dit zou op termijn wel eens minder efficiënt kunnen zijn omdat mensen hun kostenniveau dan niet meer kunnen reguleren. De waarde van de AFA ligt vooral daarin dat wordt getracht om belemmeringen om efficiënt te werken worden vermeden, hetgeen het welzijn van de taakuitvoerder kan vergroten.
Het hoofdstuk wordt besloten met enkele suggesties voor verder onderzoek.

References

Aalst, R. van, Kemper, H.C.G., Diederen, A.L.J. & Pijpers, V.T.J., (1986). Arbeidsbelasting van huisvuilbeladers na een reorganisatie van de huisvuildienst. *Tijdschrift Sociale Gezondheidszorg*, 64, 619-623.

Aasman, J., Wijers, A.A., Mulder, G., and Mulder, L.J.M., (1988), Measuring mental fatigue in normal daily working routines. In: P.A. Hancock and N., Meshkati (eds.), *Human Mental Workload*. Amsterdam: North-Holland publishers. p. 117-138.

Aasman, J., Mulder, G., and Mulder, L.J.M., (1987). Operator effort and the measurement of heart rate variability. In: *Human Factors*, vol. 29(2), 161-170.

Ackermann, D., (1986). Untersuchungen zum individualisierten Computer-dialog: Einfluss des operatieven Abbildsystems auf handlungs- und Gestaltungsspielraum und die Arbeitseffizienz. In: G. Dirlich, C. Freska, U. Schwatlo, und K. Wimmer, (Hrsgs.). *Kognitieve Aspekte der Mensch-Computer-Interaktion*. Berlin: Springer. p. 95-110.

Ackermann, D., (1987). *Handlungsspielraum, mentale Repräsentation und Handlungsregulation am Beispeil der Mensch-Computer-Interaktion*. Phil. Diss. Universität Bern.

Alluisi, E.A., (1967). Methodology in the use of synthetic task to assess complex performance. *Human Factors*, 9, 375-384.

Anderson, J.R., (ed.), (1981). *Cognitive Skills and their Acquisition*. Hillsdale, N.J.: Erlbaum.

Anderson, J.R., (1982). Acquisition of cogitive skill. *Psychological Review*, 89, 369-406.

Anderson, J.R., (1983). *The Architecture of Cognition, Cambridge*, Mass.: Harvard University Press.

Arensman, R.P.C., (1990). *Handelingsfacilitatie: een nieuw uitgangspunt bij mens-computer interface evaluatie*. Doctoraal scriptie, TUD, Vakgroep Techniek, Arbeid & Organisatie/ UvA, Vakgroep Arbeids- & Organisatiepsychologie.

Arnold, A.G., and Boogert, A.H., (1987). *The facilitation approach to the design of information systems. Experiences fom a training course with system designers*. Paper presented at the Third European Congres on Psychology of Work and Organization, Antwerp, April 13-15.

Aronson, G., (1989). Changed qualification demands in computer-mediated work. *Applied Psychology*, 38(1), p. 57-71.

Bainbridge, L., (1979). Verbal reports as evidence of the process operator's knowledge. *International Journal of Man-Machine studies*, 11, 411.

Battman, W., (1988). *Verhaltensökonomie: Grundannahmen und eine Anwendung am Fall des kooperativen Handelns*. Frankfurt am Main: Peter Lang.

Berghuis, M., Brink, H., Krayema, W., Verbeek, M., Witte, R., & Zijlstra, F., (1980). *Taakconflicten: een onderzoek naar normen en rijstrategie van buschauffeurs in stadsdiensten.* Rapport. Rijksuniversiteit Groningen: Vakgroep Arbeids- & Organisatiepsychologie, Themagroep Noord-Nederland.

Borg, G.A., (1962). *Physical Performance and Perceived Exertion.* Dissertation at University of Copenhagen (Denmark).

Borg, G.A., Bratfisch, O., and Dornic, S., (1971). On the problem of perceived difficulty. *Scandinavian Journal of Psychology,* vol. 12.

Borg, G.A. (1976). *Simple rating methods for estimation of perceived exertion.* Stockholm: Institute of Applied Psychology, Report.

Bovy, P., (1986). Routekeuzegedrag. *De Psycholoog,* Vol. 21 (12), 616-620.

Boxtel, A. van, and Jessurun, M., (in press). Amplitude and bilateral coherency of facial and jaw-elevator EMG activity as an index of effort during a serial choice reaction task. Accepted for publication in *Psychophysiology* (to appear in Sept. 1993).

Broadbent, D.E., (1958). *Perception and Communication.* London: Pergamon.

Broadbent, D.E., (1971). *Decision and Stress.* London: Academic Press.

Card, S.K., Moran, T.P., and Newell, A., (1983). *The Psychology of Human-Computer Interaction.* London: Lawrence Erlbaum.

Casali, J.G., and Wierwille, W.W., (1983). A Comparison of rating scale, secondary-task, physiological, and primary-task workload estimation techniques in a simulated flight task emphasizing communications load. In: *Human Factors,* vol. 25 (6), 623-641.

CBS, (1991). *Sociale Economische Maandstatistieken,* Vol. 4 en 7,. Den Haag.

Cooper, G.E., and Harper, R.P. (1969). *The use of pilot ratings in the evaluation of aircraft handling qualities.* NASA Ames Technical Report NASA TN-D-5153, Moffet Field, CA: NAA Ames Research Center.

Cowan, N., (1988). Evolving conceptions of memory storage, selective attention and their mutual constraints within the human information-processing system. *Psychological Bulletin,* 102, 163-191.

Damodaran, L., (1981). The role of user support. In: B. Shackel (ed.), *Man-Computer Interaction: Human Factors Aspects of Computers and People.* Rockville, Maryland: Sijthoff and Noordhoff.

Dellen, H.J. van, Aasman, J., Mulder, L.J.M., and Mulder, G., (1985). Time domain versus frequency domain measures of heart rate variability. In: J.F. Orlebeke, G. Mulder, and L.J.P. van Doornen (eds.), *The Psychophysiology of Cardiovascular Control.* New York: Plenum Press. p. 353-374.

Derrick, W.L., (1981). The relations between processing resources and subjective dimensions of operator workload. In: R. Sugarman (ed.), *Proceedings, 25th Annual Meeting of the Human Factor Society.* Santa Monica, CA.

Derrick, W.L., (1988). Dimensions of operator workload. *Human Factors,* vol. 30, 95-110.

DIN (1986), Deutsches Institut für Normung e.V., *Norm paper VDU Workstations, principles of dialogue design,* nr. 66 234 Part 8, January 1986.

Donchin, E., (1981). Surprise!..........Surprise? *Psychophysiology,* 18, 493-513.

Doorn, L. van, and Zijlstra, F.R.H., (1988). Variation in response functions complicates the evaluation of scales. In: W.E. Saris (ed.), *Variation in Response Functions: a Source of Measurement Error in Attitude Research*. Amsterdam: Sociometric Research Foundation, p. 87-97.

Drenth, P.J.D. en Sijtsma, K., (1990). *Testtheorie - Inleiding in de Theorie van de Psychologische Test en zijn Toepassingen*. Houten: Bohn Stafleu Van Loghum bv.

Drury, C.G., (1990). Designing ergonomics studies and experiments. In: Wilson, J.R., and Corlet, E.N., (eds.), *Evaluation of Human Work: a Practical Ergonomics Methodology*. London: Taylor & Francis. p. 101-129.

Dzida, W., (1985), Ergonomische Normen für die Dialoggestaltung. In: H.F. Bullinger (ed.), *Software-ergonomie 1985 Mensch-Computer Interaktion*. Stuttgart: Teuber Verlag, p. 430-444.

Dzida, W., Herda, S., & Itzfeldt, W.D. (1978). User-perceived quality of Interactive Systems, *IEEE Transactions on Software Engineering*, vol. SE-4(4), p. 270-276.

Eason, K.D., (1981). A task-tool analysis of manager-computer interaction. In: B. Shackel (ed.), *Man-Computer Interaction: Human Factors Aspects of Computers and People*. Rockville, Maryland: Sijthoff and Noordhoff.

Edwards, W., (1954). The theory of decision making. *Psychological Bulletin*, 51, 380-417.

Eekels, J., (1982). *Industriële Doelontwikkeling*. Delft; Delftse Universitaire Pers.

Egan, D.E., (1988). Individual differences in human-computer interaction. In: Helander, M., (ed.), *Handbook of Human-Computer Interaction*. Amsterdam: North-Holland. p. 543-568.

Eilers, K., Nachreiner, F., und Hänecke, K., (1986). Entwicklung und Überprüfung einer Skala zur Erfassung subjektiv erlebter Anstrengung. *Zeitschrift für Arbeitswissenschaften*, (4), 215-224.

Fibiger, W., et al., (1984). Cortisol and catecholamines changes as functions of time-of-day and selfreported mood. *Neuroscience and Behavioral Reviews*, vol. 8, 523-530.

Fleishman, E.A., and Quaintance, M.K., (1984). *Taxonomies of Human Performance*. New York: Academic Press.

Frankenhauser, M., (1975). Experimental approaches to the study of catecholamines and emotions. In: L. Levi (ed.), *Emotions - Their Parameters and Measurement*. New York: Raven Press. p. 209-234.

Frankenhauser, M., (1986). Stress at work: psychobiological and psychosocial aspects. *International Review of Applied Psychology*, vol. 35, 287-299.

French-Lazovik, G., and Gibson, G.L. (1984). Effects of verbally labelled anchor points on the distributional parameters of rating measures. In: *Applied Psychological Measurement*, vol. 8 (1), 49-57.

Frese, M., Sabini, J., (1985). *Goal-directed Behavior* : the Concept of Action in Psychology. Hillsdale, N.J : L. Erlbaum Associates.

Frese, M., & Brodbeck, F.C. (1989). *Computer in Büro und Verwaltung. Psychologisches Wissen für die Praxis*. Berlin: Springer Verlag.

Gardiner, M.M., and Christie, B., (Eds.), (1987). *Applying Cognitive Psychology to User-Interface Design*. Chichester: Wiley & Sons.

Gopher, D., and Navon, D., (1980). How is performance limited: testing the notion of central capacity. *Acta Psychologica*, 46, 161-180.

Gopher, D., and Braune, R., (1984). On the psychophysics of workload: why bother with subjective measures? *Human Factors*, vol. 26(5), 519-532.

Gopher, D., and Donchin, E., (1986). Workload - an examination of the concept. In: K. Boff, L. Kaufman, & J.P. Thomas (eds.), *Handbook of Perception and Human Performance*, Vol. II - Cognitive Processes and Performance, New York: Wiley and Sons, chapter 41.

Gould, J.D., (1988). How to design usable systems. In: Helander, M., (ed.), *Handbook of Human-Computer Interaction*. Amsterdam: North-Holland. p. 757-789.

Greif, S., (1991). The role of German work psychology in the design of artifacts, 1991. In: J.M. Carroll (ed.), *Designing Interaction - Psychology at the Human-computer Interface*. Cambridge: Cambridge University Press. p. 203-226.

Hacker, W., (1973). Arbeits- und Ingenieurpsychologische Beiträge zur sozialistischen Rationalisierung der Volkswirtschaft. In: W. Hacker, K.-P., Timpe, und M. Vorwerg (Hrsgs.), *Arbeits-, ingenieur- und sozialpsychologische Beiträge zur sozialistischen Rationalisierung*. VEB Verlag der Wissenschaften: Berlin.

Hacker, W., (1978). *Allgemeine Arbeits- und Ingenieurspsychologie, Psychische Struktur und Regulation von Arbeitstätigkeiten*. Bern: Huber Verlag.

Hacker, W., (1986). *Arbeitspsychologie. Psychische Regulation von Arbeitstätigkeiten*. Bern: Huber Verlag.

Hacker, W., Raum, H., Rentzsch, M., und Völker, K., (1987). *Bildschirmarbeit - Analyse und Gestaltung rechnergestützte Arbeit*. Berlin: Verlag die Wirtschaft Berlin.

Hacker, W., Volpert, W., and Cranach, V. von, (eds.), (1982). *Cognitive and Motivational Aspects of Action*. Amsterdam: North-Holland.

Hackman, J.R., & Oldham, G.R., (1975). Development of the Job Diagnostic Survey. *Journal of Applied Psychology*, 60, 159-170.

Hamblin, P.L., (1974). Social attitudes: magnitude, measurement and theory. In: H.M. Blalock (ed.), *Measurement in the Social Sciences*. London: MacMillan. p. 61-120.

Hart, S.G., and Staveland, L.E., (1988). Development of NASA-TLX (Task Load indeX): Results of empirical and theoretical research. In: P.A. Hancock, and N., Meshkati (eds.), *Human Mental Workload*. Amsterdam: North-Holland publishers. p. 139-184.

Hasher, L., & Zacks, R., (1979). Automatic and effortful processing in memory. *Journal of Experimental Psychology*, 108, 356-388.

Heemstra, M.L. (1988). *Efficiency of Human Information Processing: a Model of Cognitive Energetics*. Acad. diss. Vrije Universiteit van Amsterdam.

Helander, M., (ed.), (1988). *Handbook of Human-Computer Interaction*. Amsterdam: North-Holland.

Hellinga, P., (1985). *De constructie van de Groninger Adjectief Checklist, de GACL*. Groningen: Heymans Bulletin, HB-85-771-Ex.

Hockey, G.R.J., (1984). Varieties of attentional state: the effects of the environment. In: R. Parasuraman et al (eds.), *Varieties of Attention*. New York: Academic Press. p. 449-484.

Hockey, G.R.J., (1986a). A state control theory of adaptation and individual differences in stress management. In: G.R.J. Hockey, A.W.K. Gaillard, and M. Coles (eds.), *Energetics and Human Information Processing*, Dordrecht: M. Nijhoff Publishers. p. 285-298.

Hockey, G.R.J., (1986b). Changes in operator efficiency as a function of environmental stress, fatigue, and circadian rhythms. In: K.R. Boff, L. Kaufman, & J.P. Thomas (eds.). *Handbook of Perception and Human Performance, vol. II - Cognitive Processes and Performance*. New York: Wiley & Sons. chapter 44.

Hockey, G.R.J., and Hamilton, P., (1983). The cognitive patterning of stress states. In: G.R.J. Hockey (ed.), *Stress and Fatigue in Human Performance*. New York: Wiley and Sons. p. 331-362.

Hockey, G.R.J., Coles, M., & Gaillard, A.W.K., (1986). Energetical issues in research on human information processing. In: G.R.J. Hockey, A.W.K. Gaillard, and M. Coles (eds.), *Energetics and Human Information Processing*, Dordrecht: M. Nijhoff Publishers. p. 3-22.

Horn, L.A. ten, & Zijlstra, F.R.H. (1990). *Informatie Technologie in de Maatschappij: Toepassing, Beleid en Perspectief*. Leiden: Stenfert Kroese.

Howard, S., and Murray, D.M., (1987), A taxonomy of evaluation techniques for HCI. In: H.-J. Bullinger, and B. Shackel (Eds.), *Human-Computer Interaction - INTERACT '87*. Amsterdam: Elsevier Science Publishers BV. p. 453-460.

Iwanowa, A., und Hacker, W., (1984). Das Tätigkeitsbewertigungssystem (TBS); Ein Hilfsmittel beim Erfassen potentiell gesundheits- und entwicklungsfomder objektiver Tätigkeitsmerkmale. *Psychologie und Praxis, Zeitschrift für Arbeits- und Organizationspsychologie*, 28 jrg. Heft 2, 57-66.

Johanssen, G., Moray, N., Pew, R., and Rasmussen, J., Sanders, A.F., Wickens, C.D., (1979). Final report of experimental psychology group. In: N. Moray (ed.), *Mental Workload: Its Theory and Measurement*. New York: Plenum Press. p. 101-114.

Jorna, P.G.A.M, (1985). Heartrate parameters and the coping process under water. In: J.F. Orlebeke, G. Mulder, and L.J.P. van Doornen (Eds.), *Psychophysiology of Cardiovascular Control: Models, Methods, and Data*. New York: Plenum Press, p. 827-840.

Kahneman, D., (1973). *Attention and Effort*. Englewood Cliffs (N.J.): Prentice Hall.

Kantowitz, B.H., and Knight, J.L., (1976). Testing tapping time-sharing I: auditory secondary task. *Acta Psychologica*, 40, 343-362.

Kantowitz, B.H., Hart, S.G., & Bortolussi, M.R. (1983). Measuring pilot workload in a moving base simulator: I. Asynchronous secondary choice-reaction task. *Proceedings of the Human Factors Society*, 27, 319-322.

Kantowitz, B.H., Hart, S.G., & Bortolussi, M.R., (1983). Measuring pilot workload in a moving base simulator: II. Building levels of workload. *Proceedings of the Annual Conference on Manual Control,* 20, 359-371. Ames Research Center, Moffet Field, CA: NASA.

Karasek, R., and Theorell, T., (1990). *Healthy Work - Stress, Productivity, and the Reconstruction of Working Life.* New York: Basic Books.

Karat, J., (1988). Software evaluation methodologies. In: Helander, M., (ed.). *Handbook of Human-Computer Interaction.* Amsterdam: North-Holland. p. 891-904.

Karat, C.-M., (1992). Cost-justifying human factors support on software development projects. *Human Factors Society Bulletin*, vol. 35(11). p. 1-4.

Kerr, E.B., and Hiltz, S.R., (1982). *Computer-mediated Communication Systems.* New York: Academic Press.

Kieras, D.E., and Polson, P.G., (1985). An approach to the formal analysis of user complexity. *International Journal of Man-Machine Studies*, 22, 365-394.

Kieras, D.E., (1988). Towards a practical GOMS model methodology for user interface Design. In: Helander, M., (ed.). *Handbook of Human-Computer Interaction.* Amsterdam: North-Holland. p. 135-157.

Kompier, M.A.J., (1988). *Arbeid en Gezondheid van Stadsbuschaffeurs.* Acad. Diss. R.U. Groningen. Delft: Eburon.

Kompier, M.A.J., en Marcelissen, F.H.G., (1990). *Handboek Werkstress.* Amsterdam: NIA.

Kompier, M., Linden van der, T., Meijman, T., Barendsen, P., Bullinga, R., Bij van de, D., Cavalini, P., Dormolen, M., & Zijlstra, F., (1982). *Taakopvattingen van Buschauffeurs, een poging om de opgave-struktuur van buschauffeurs te meten.* Rijksuniversiteit Groningen: Vakgroep Arbeids- & Organisatie-psychologie, (rapport uitgebracht aan de Werkgroep Ziekteverzuim van de Gemeente Groningen).

Landauer, K., (1988). Research methods in Human-Computer Interaction. In: Helander, M., (ed.). *Handbook of Human-Computer Interaction.* Amsterdam: North-Holland. p. 905-928.

Lehmann, M., et al., (1982). Katecholamineausscheidung bei Körperlichen und Konzentrativen Belastungen. *International Archives of Occupational and Environmental Health*, vol. 50, 175-186.

Leontjev, A.N., (1964). *Probleme der Entwicklung des Psychischen.* Berlin: VEB Verlag der Wiss.

Leontjev, A.N., (1979). *Tätigkeit, Bewusstsein, Persönlichkeit.* Berlin: VEB Verlag der Wiss.

Leplat, J., (1989). Cognitive skills at work. In: Bainbridge, L., Ruiz Quintanilla, S.A., (eds.), *Developing Skills with New Information Technology.* Chichester: Wiley & Sons.

Lewin, K., (1926). Untersuchungen zur Handlungs- und Affektpsychologie. Vorbemerkungen über die psychischen Krafte und Energien und über die Struktur der Seele. *Psychologische Forschungen*, 7.

Lodge, M., (1981). *Magnitude Scaling. Quantitative Measurement of Opinions.* Beverly Hills: Sage Publications.

Logan, G.D., (1988a). Towards an instance theory of automatization. *Psychological Review*, 95, suppl.

Logan, G.D., (1988b). Automaticity, resources and memory: theoretical controverses and practical implications. *Human Factors*, 30, 583-598.

Lundberg, U., & Frankenhauser, M., (1980). Pituitary-adrenal and sympatic-adrenal correlates of distress and effort. *Journal of Psychosomatic Research*, 24, 125-130.

Mackay, C., & Whittington, M.C., (1983). Ergonomics and human error. In: Ridley, J. (ed.), *Safety at Work.* Butterworths.

Martin, J., (1973). *Design of Man-Computer Dialogues.* New Jersey: Prentice-Hall.

Massaro, D.W., (1975). *Experimental Psychology and Information Processing.* Chicago: Rand-McNally.

Mayer, R.E., (1988). From novice to expert. In: Helander, M., (ed.), *Handbook of Human-Computer Interaction.* Amsterdam: North-Holland. p. 569-580.

McClelland, J.L., (1979). On the time relations and mental processes. An examination of systems of processes in cascade. *Psychological Review*, 86, 287-330.

Meijer, T., (1989). *Mentale Informatie Werk.* Literatuurstudie. Arbeids- & Organisatie psychologie, Tilburg University.

Meijman, Th.F., en O'Hanlon J.F. (1984). Workload, an introduction to psychological theories and measurement methods. In: P.J.D. Drenth, Hk. Thierry, en Ch.J. de Wolff (eds.), *Handbook of Work and Organizational Psychology.* New York: Wiley and Sons. p. 257-288.

Meijman, T., Cavalini, P., Cremer, R., Dormolen, M. van, Gloerich F., Hellinga, P., en Noord, F. van, (1985). *Onderzoek Taakbelasting Rijexaminatoren CBR*, Rapport. Vakgroep Funktieleer & Arbeidspsychologie, RU Groningen.

Meijman, T, Zijlstra, F., Kompier, M., Mulders, H., and Broersen, J., (1986). The measurement of perceived effort. In: D.J. Oborne (ed.), *Contemporary Ergonomics 1986.* London: Francis & Taylor, p. 242-246.

Meijman, Th., Ouwerkerk, R. van, Mulder, L., en Vries-Griever, A. de, (1989). Hartslagvariabiliteit: het onderzoek van mentale efficiëntie. In: T.F. Meijman (red.), *Mentale Belasting en Werkstress*, Assen: Van Gorcum. p. 62-83.

Meijman, Th.F. (1989). Belasting en herstel: een begrippenkader voor het arbeidspsychologisch onderzoek van werkbelasting. In: Th.F. Meijman (red.), *Mentale Belasting en Werkstress - Een Arbeidspsychologische Benadering.* Assen: Van Gorcum. p. 5-20.

Meijman, Th.F., (1991). *Over Vermoeidheid.* Acad. Dis. R.U. Groningen. Amsterdam: Studie Centrum Arbeid & Gezondheid.

Meijman, Th.F., en Mulder, G., (1992). Arbeidspsychologische aspekten van werkbelasting. In: P.J.D. Drenth, Hk. Thierry, en Ch.J. de Wolff (red.), *Nieuw Handboek Arbeids- & Organisatie Psychologie*. Deventer: Van Loghum Slaterus (hoofdstuk 2.11).

Moray, N., (1982). Subjective mental workload. In: *Human Factors*, vol. 24, no. 1, 25-40.

Miller, R.B., (1962). Task description and analysis. In: R.M. Gagné (ed.), *Psychological Principles in System Development*. New York: Holt, Rinehart & Winston.

Miller, G.A., Galanter, E., and Pribram, K.H., (1960). *Plans and the Structure of Behavior*. New York: Holt.

Mulder, G., (1980). *The Heart of Mental Effort*. Acad. Thesis. University of Groningen (The Netherlands).

Mulder, G., (1986). The concept of mental effort, and its measurement. In: Hockey, G.R.J. et. all. (eds.) *Energetics and Human Information Processing*, Dordrecht: M. Nijhoff Publishers. p. 175-198.

Mulder, L.J.M. (1988). *Assessment of Cardiovascular Reactivity by Means of Spectral Analysis*. Acad. Thesis, Rijks Universiteit Groningen.

Mulders, H.P.G., Meijman, Th.F., and O'Hanlon, J.F., (1982). Differential psychophysiological reactivity of city busdrivers. In: *Ergonomics*, vol. 25(11), 1003-1011.

Mulders, H.P.G., Dormolen, M., en Broersen, S., (1988). Differentiële reactiviteit in relatie tot arbeidsbelasting en gezondheid. In: B.G. Deelman & G. Mulder (red.), *Nederlands Psychologisch Onderzoek 3; Bijdragen aan de Experimentele Gerontologie en het Onderzoek naar Mentale Belasting en Stress*. Lisse: Swets & Zeitlinger. p.219-238.

Navon, D., and Gopher, D., (1979). On the economy of the human processing system. *Psychological Review*, 86, 254-255.

Neuberger, O., (1985). *Arbeit*. Stuttgart: Enke Verlag.

Newell, A. and Simon, H.A. (1972). *Human Problem Solving*. Englewood Cliffs (N.J.): Prentice-Hall.

Nickerson, R.S., (1986). *Using computers: The Human Factors of Information Systems*. Massachusetts: The MIT Press.

Norman, D., and Bobrow, D.J., (1975). On data and resource limited processes. In: *Cognitive Psychology*, 7, 44-66.

O'Donnell, R.D., and Eggemeier, F.T. (1986). Workload assessment methodology. In: K.R. Boff, L. Kaufman, J.P. Thomas (Eds.), *Handbook of Perception and Human Performance, Vol. II: Cognitive Processes and Performance*. New York: Wiley & Sons. p. 42-1 - 42-49.

O'Hanlon, J.F., (1981), Boredom: practical consequences and a theory. In: *Acta Psychologica*, vol. 49, 53-82.

Olson, J.R., (1987). Cognitive analysis of people's use of software. In: J.M. Carroll (Ed.), *Interfacing Thought*: Cognitive Aspects of Human-Computer Interaction. Cambridge, MA: MIT Press. p. 260-293.

Olson, J., and Olson, G.M. (1990). The growth of cognitive modelling in human-computer interaction since GOMS. In: *Human-Computer Interaction*, Vol. 5, 221-265.

Otto, P., (1983). Voraussetzungen der Automatisierung von Hand- und Kopfarbeit. In: Sonntag, P. (Hrsg.), *Die Zukunft der Informationsgesellschaft*. Arnoldshainer Schriften zur interdisziplinären Ökonomie, Bd 5. Frankfurt a.M.: Haag und Herchen, 142-165.

Payne, S.J., and Green, T.R.G., (1986). Task action grammars: a model of the mental representation of task languages. *Human Computer Interaction*, 2, 93-133.

Perdok, J. (1984). *Methods to study mental workload in one-man-manning situations*. Paper presented at the 'SecondInternational Bridge Design and Operation Forum, Hamburg august 1984. MARIN Wageningen.

Polson, P.G., (1987). A quantitative theory of human-computer interaction. In: J.M. Carroll (ed.). *Interfacing Thought: Cognitive Aspects of Human-Computer Interaction*. Cambridge, MA: Bradford, MIT Press. p. 184-235.

Porat, M.U., (1977). *The Information Economy*. Washington DC: Department of Commerce, Office of Telecommunications.

Posner, M.I., (1964). Information reduction in the analysis of sequential tasks. *Psychological Review*, 71(6), 491-504.

Posner, M.I., and Snyder, C.R.R., (1975). Attention and cognitive control. In: R.L. Solso (ed.), *Information Processing and Cognition: the Loyola Symposium*. Hillsdale (NJ): Erlbaum. p. 55-86.

Pribram, K.H., and McGuinness, D., (1975). Arousal, activation and effort in the control of attention. In: *Psychological Review*, vol. 82 (2), 118-149.

Rabbitt, P. (1979). Current paradigms and models in human information processing. In: V. Hamilton, and D.M. Warburton (Eds.), *Human Stress and Cognition: An Information Processing Approach*. Chichester: Wiley & Sons. p. 115-140.

Rasmussen, J., (1981). Models of mental strategies in process plant diagnosis. In: J. Rasmussen and W. Rouse, (Eds.), *Human Detection and Diagnosis of System Failures*. New York: Plenum Press. p. 241-258.

Rasmussen, J., (1983). Skills, rules, and knowledge: signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man and Cybernetics*, vol. 13, p. 257.

Reid, G.B. and Nygren, T.E., (1988). The subjective workload assessment technique: a scaling procedure for measuring mental workload. In: P.A. Hancock and N., Meshkati (eds.), *Human Mental Workload*. Amsterdam: North-Holland publishers. p. 185-218.

Roe, R.A. (1984). Taakstructuur en software-ontwerp. In: R.F. Wijk van Brievingh & F.G. Peuscher (red.), *Methodisch Ontwerpen van Medisch Technologische Apparatuur*. Delft: Delft University Press, p. 142-160.

Roe, R.A. (1988). Acting systems design - an Action Theoretical Approach to the Design of Man-Computer Systems. In: V. de Keyser, T. Qvale, B. Wilpert, and S.A. Ruiz Quintanilla (Eds.), *The Meaning of Work and Technological Options.* London: Wiley & Sons. p. 179-195.

Roe, R.A., (1990). *Bepaald Anders: een Eco-Technologische Kijk op Informatie-Arbeid in de Dienstverlening.* Assen: Van Gorcum.

Roe, R.A., and T, Meijer, (1990). Action facilitation and mental information work - towards an approach to within-job design. In: P. Richter and W. Hacker (Eds.), *Mental Work and Automation. Proceedings of 6th International Symposium on Work Psychology.* Technische Universität Dresden. p. 55-60.

Roe, R.A. en Zijlstra, F.R.H., (1991). Arbeidsanalyse ten behoeve van (her)ontwerp van functies: een handelingstheoretische invalshoek. In: J.A. Algera (red.), *Analyse van Arbeid vanuit Verschillende Perspectieven.* Lisse: Swets & Zeitlinger, p. 179-243.

Roe, R.A., Zijlstra, F.R.H., Schalk, M.J.D., Meijer, T., Taggenbrock J., en Neervoort, P. (1993). *Mentaal Informatie Werk: verslag van een survey.* report. Work and Organizational Research Centre of Tilburg University.

Ronde, P. de, (1992). *Evaluatie 'Test Beeldschermwerk'.* stage verslag. Vakgroep Arbeids- & Organisatie psychology, Tilburg University.

Rubinstein, S.L., (1962). *Sein und Bewusstsein.* Berlin: VEB Verlag der Wiss.

Russell, B., (1921). *The Analysis of Mind.* New York: Macmillan.

Sabini, J., Frese, M., and Kossman, D.A., (1985). Some contributions of Action Theory to Social Psychology: Social Action and actors in the context of institutions and an objective world. In: M. Frese and J. Sabini (eds.), *Goal Directed Behavior: The Concept of Action in Psychology.* Hillsdale (NJ): Erlbaum. p. 249-260.

Saltzman, E., and Scott Kelso, J.A., (1987). Skilled actions - a task dynamic approach. *Psychological Review,* 94, 84-106.

Sanders, A.F., (1983). Towards a model of stress and human performance. *Acta Psychologica,* vol. 53, 61-97.

Saris, W.E., Bruinsma, C., Schoot, W., & Vermeulen, C., (1977). The use of magnitude estimation in large scale survey research. In: *Mens en Maatschappij,* vol. 52, 369-395.

Saris, W.E., Neijens, P., and Doorn, L. van, (1980). Scaling social science variables by multimodality matching. *Methoden en Data Nieuwsbrief,* 2, 3-21.

Saris, W.E. and Stronkhorst, L.H., (1984). *Causal Modeling in Nonexperimental Research: an Introduction to the LISREL Approach.* Amsterdam: Sociometric Research Foundation.

Schäfer, G., (1988). *Functional Analysis of Office Requirements: a Multiperspective Approach.* Chichester: Wiley & Sons.

Schneider, W., and R.M. Shiffrin, (1977). Controlled and automatic human information processing. I: detection, search, and attention. *Psychological Review,* vol. 84. p. 1-66.

Schönpflug, W., (1985): Goal Directed Behavior as a Source of Stress: Psychological Origins and Consequences of Inefficiency, In: M. Frese & J. Sabini: *Goal Directed Behavior: The Concept of Action in Psychology*, Hillsdale (NJ): Erlbaum. p. 172-188.

Schönpflug, W., (1986a). Behavior Economics as an approach to Stress Theory. In: M.H. Appley and R. Trumbull (eds.). *Dynamics of Stress - Physiological, Psychological, and Social Perspectives*. New York: Plenum Press. p. 81-98.

Schönpflug, W., (1986b). Effort regulation and individual differences in effort expenditure. In: G.R.J. Hockey, A.W.K. Gaillard, and M. Coles (eds.), *Energetics and Human Information Processing*, Dordrecht: M. Nijhoff Publishers. p. 271-284.

Semmer, N. (1984) *Stressbezogene Tätigkeitsanalyse; Psychologische Untersuchungen am Arbeitsplatz*. Weinheim: Beltz.

Sherridan, T.B., (1980). Mental workload: What is it? Why bother with it? *Human Factors Society Bulletin*, 23, 1-2.

Sherridan, T.B., and Simpson, R.W., (1979). *Toward the Definition and Measurement of the Mental Workload of Transport Pilots*. (FTL Report R 79-4). Cambridge, MA: Massachusetts Institute of Technology, Flight Transportation Laboratory.

Shiffrin, R.M., W. Schneider, (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attention and a general theory, *Psychological Review*, vol. 84, p. 127.

Shneiderman, B., (1979). Human factors experiments in designing interactive systems. *Computer Institute of Electrical And Electronic Engineers Publication*, 12, 9-19.

Singley, M.K., and Anderson, J.R., (1985). The transfer of text-editing skill. In: *International Journal of Man-Machine Studies*, 22, 403-423.

Sitter, U., de, (1981). *Kwaliteit van de Arbeid. Op Weg Naar Nieuwe Fabrieken en Kantoren*. Deventer: Kluwer.

Skipper, J.H., Rieger, C.A., and Wierwille, W.W., (1986). Evaluation of decision-tree rating scales for mental workload estimation. In: *Ergonomics*, vol. 29 (4), p. 585-599.

Sperandio, J.C., (1972). Charge de travail et regulation des procèssus operatoires. Le Travail Humain, 35, 85-98.

Sperandio, J.C., (1978). The regulation of working methods as a function of workload among air traffic controllers. *Ergonomics*, 21, 193-202.

Sperandio, J.C., (1980). *La Psychologie en Ergonomie*. Paris: Presses Universitaire de France.

Stassen, H.G., Johannsen, G., and Moray, N., (1988). Internal representation, internal model, human performance model and mental workload, In: *Proceedings IFAC Man-Machine Systems*, Oulu, Finland. p. 23-32.

Sternberg, S., (1969). Memory scanning: mental processes revealed by reaction time experiments. In: *American Scientist*, vol. 57, 421-457.

Stevens, S.S., (1966). A metric for the social consensus. In: *Science*, vol. 151, 530-541.

Stevens, G.C., (1983). User-friendly computer systems? A critical evaluation of the concept. *Behaviour and Information Technology*, 2, 3-16.

Stewart, T.F.M., (1981). The specialist user. In: B. Shackel (ed.), *Man-Computer Interaction: Human Factors Aspects of Computers and People*. Rockville, Maryland: Sijthoff and Noordhoff.

Swanbom, P.G., (1982). *Schaaltechnieken, Theorie en Praktijk van Acht Eenvoudige Procedures*. Meppel: Boom.

Taylor, F.W., (1907). *On the Art of Cutting Metals*. New York: A.S.M.E.

Taylor, F.W., (1911). *The Principles of Scientific Management*. New York: Harper.

Taylor, F.W., (1972, reprint) *Scientific Management*, comprising 'Shop Management', 'The Principles of Scientific Management'; Testimony before the Special House Committee. Greenwood Press, Westport, Connecticut.

Teiger, C., (1978). Regulation of activity: an analytical tool for studying work load in perceptual motor tasks. *Ergonomics*, 21, 203-213.

Thayer, R.E., (1967). Measurement of activation through self report. *Psychological Reports*, 20, 663-678.

Thayer, R.E., (1978). Factor analytic and reliability studies on the activation-deactivation adjective checklist. *Psychological Reports*, 42, 747-756.

Treisman, A., (1969). Strategies and models of selective attention. *Psychological Review*, 76, 282-299.

Treisman, A., & Davies, A., (1973). Divided attention to eye and ear. In: S. Kornblum (ed.), *Attention and Performance IV*. New York: Academic Press.

Trist, E.L., (1981). The evolution of sociotechnical systems as a conceptual framework and as an action research program. In: A.H. van de Ven & W.F. Joyce (eds.), *Perspectives on Organization and Behavior*. New York: Wiley & Sons.

Trist, E.L., and Bamforth, K.W., (1951). Some social and psychological consequences of the longwall method of coal-getting. *Human Relations*, 4, 1-38.

Ulich, E., (1978). Über das Prinzip der differentiellen Arbeitsgestaltung. In: *Industrielle Organisation*, 47, 566-568.

Ulich, E., (1984). Psychologie der Arbeit. In: *Management-Enzyklopädie* (Band 7, pp. 914-929). Landsberg: Moderne Industrie.

Ulich, E., (1991). *Arbeitspsychologie*. Stuttgart: Verlag Poeschel.

Ulich, E., Udris, I., Frei, F., und Alioth, A., (1980). Zur Frage der Individualiserung von Arbeitstätigkeiten. In: W. Hacker, und H. Raum (Hrsg.), *Optimierung von kognitiven Anforderungen*. Berlin: VEB Verlag der Wiss. p. 65-69.

Veer, G.C., van der, (1990), *Human-Computer Interaction: Learning, Individual Differences, and Design recommendations*. (Acad. Thesis), Amsterdam: Vrije Universiteit.

Velden, J.M, van der, Arnold, A.G., en Roe, R.A., (1989). *Zoeken en Vinden? Evaluatie van het user interface van een online bibliotheek zoeksysteem met betrekking tot gebruiker, taak en gebruik*. Delft: Delftse Universitaire Pers.

Veldman, H., (1992). *Hidden Effects of Noise as Revealed by Cardiovascular Analysis*. Acad. Thesis. Groningen: Rijks Universiteit Groningen.

Vidulich, M.A., and Tsang, P.S., (1986). Techniques of subjective workload assessment: a comparison of SWAT and the NASA-Bipolar methods. *Ergonomics*, 29(11), 1385-1398.

Volpert, W., (1975). Die Lohnarbeitswissenschaft und die Psychologie der Arbeitstätigkeit. In: W. Volpert und P. Groskurth (Hrsgs.). *Lohnarbeitspsychologie*. Frankfurt, Main: Fisher. (pp. 11-169).

Volpert, W., (1982). The model of the hierarchical-sequential organization of action. In: W. Hacker, W. Volpert, and M. von Cranach (eds.), *Cognitive and Motivational Aspects of Action*. Amsterdam: North Holland.

Volpert, W., (1990). Work design for human development. In: R. Budde, C. Floyd, R. Keil-Slawik, and H. Züllighoven (eds.). *Software Development and Reality Construction*. New York: Springer.

Vroom, V.H., (1964). *Work and Motivation*. New York: Wiley & Sons.

Winer, B.J., Brown, D.R., and Michels, K.M., (1991). *Statistical Principles in Experimental Design*. New York: McGraw-Hill. (third edition).

Watson, J.B., (1919). *Psychology from the Standpoint of a Behaviorist*. Philadelphia: J.B. Lippincott.

Westerink, B.H.C., Bosker, F.J., & O'Hanlon, J.F., (1982). Use of alumina, Sephadex G 10 and ion exchange columns to purify samples for determination of epinephrine, norepinephrine, dopamine, homovanillac acid and 5-hydroxyindoleacetic in urine. *Clinical Chemistry*, 28, 1745-1748.

Wickens, C.D., (1984). *Engineering Psychology and Human Performance*. Columbus (O): Merill.

Wickens, C.D., (1986). The effects of control dynamics on performance. In: K.R. Boff, L. Kaufman, J.P. Thomas (Eds.), *Handbook of Perception and Human Performance, Vol. II: Cognitive Processes and Performance*. New York: Wiley & Sons. chapter 39.

Wickens, C.D., (1987). Attention. In: P.A. Hancock (ed.), *Human Factors Psychology*. Amsterdam: Elsevier Science Publications. p. 29-80.

Wiethoff, M., Gunther, Th., Mulder, G. en Mulder, L., (1988). Mentale inspanning bij slaapdeprivatie. In: B.G. Deelman en G. Mulder (red.), *Nederlands Psychologisch Onderzoek 3: Bijdragen aan de Experimentele Gerontologie en het Onderzoek naar Mentale Belasting en Stress*. Lisse: Swets & Zeitlinger. p. 159-172.

Williges, R.C., and Wierwille, W.W., (1979). Behavioral measures of aircrew mental workload. *Human Factors*, 21, 549-574.

Wijers, A.A., (1989). *Selective Visual Attention: a Psychophysiological Approach*. Acad. diss. Groningen University.

Winsemius, W., (1969). *Taakstructuren, Storingen en Ongevallen*. Groningen: Wolters Noordhoff. Leiden: TNO/NIPG.

Zijlstra, F.R.H., and L. van Doorn, (1985). *The Construction of a Subjective Effort Scale*, Report, Dept. Social Sciences & Philosophy, Delft University of Technology.

Zijlstra, F.R.H., en Meijman, Th.F., (1988). Subjectieve Mentale Inspanning: gecontroleerde informatieverwerking en/of 'executive resource control'?. In: B.G. Deelman en G. Mulder (red.), *Nederlands Psychologisch Onderzoek 3: Bijdragen aan de Experimentele Gerontologie en het Onderzoek naar Mentale Belasting en Stress*. Lisse: Swets & Zeitlinger. p. 239-250.

Zijlstra, F.R.H. und Roe, R.A., (1988). Ansätze zur Beanspruchungs- und Belastungsmessung in Verbindung mit neuen Technologien. In: E. Frieling und H. Klein (Hrsgs.), *Rechnerunterstützte Konstruktion - Bedingungen und Auswirkungen von CAD*. Bern: Huber Verlag. p. 258- 267.

Zijlstra, F.R.H., en Meijman, Th.F. (red.), (1989). *Mentale Inspanning: Nader Bekeken*. Groningen: Instituut voor Experimentele en Arbeidspsychologie.

Zijlstra F.R.H., P.C. Cavalini, M. Wiethoff and Th.F. Meijman, (1990). Mental Effort: strategies and efficiency. In: P.J. Drenth, J. Sergeant en R. Takens (Eds.), *European Perspectives in Psychology. vol. 1*. Chichester: Wiley and Sons. p. 435-447.

Zijlstra, F.R.H., Buurman, R. den, Horn, L.A. ten, Regensburg, R.E., (1989). *Test Beeldschermwerk*. Delft: Delft University Press.

Appendix A:

The AFA-checklist was made in the Dutch language. Only the dimensions have been translated in English.

I. Try to support the process of action preparation, i.e. orientation to the task and forming an action plan, by offering adequate information about the system.

Ondersteun handelingsvoorbereiding, i.h.b. de oriëntatie op de taak en het opstellen van handelingsplannen, door informatie en hulpmiddelen aan te reiken.

1). Bij 'gesloten' vragen worden alle alternatieven getoond.

2). De relaties tussen vragen zijn duidelijk.

3). De gebruiker kan zelf 'macro's' definiëren.

4). De gebruiker kan zelf toetsen definiëren.

5). De gebruiker kan zelf toetsen herdefiniëren.

6). De gebruiker kan zelf bepalen welke informatie, zoals bijv. de gekozen instellingen, op het beeldscherm worden getoond.

7). op ieder moment is het mogelijk hulpinformatie op te vragen zonder daarbij lopende activiteiten te onderbreken.

8). Foutmeldingen zijn 'zelfverklarend'.

II. Try to achieve an un-interrupted execution of action plans by presenting adequate feedback about progress and results of activities.

Draag bij aan een ononderbroken, vlotte uitvoering van handelingsplannen, door het geven van signalen en feedback over het verloop en resultaat van activiteiten.

1). Informatie over de gekozen instellingen wordt, indien dit is gewenst, getoond.

2). Foutmeldingen zijn duidelijk geformuleerd.

3). Extra informatie over een foutmelding kan worden opgevraagd.

4). Foutmeldingen worden tijdig gegeven zonder de taak onnodig te onder-breken.

5). Bij een foutmelding wordt aangegeven hoe het beste verder te gaan.

6). Het is mogelijk een printvoorbeeld op het beeldscherm op te vragen.

7). Waarschuwingen en adviezen worden gegeven indien fouten dreigen op te treden.

8). Het programma bevat hinderlijke fouten, zgn 'bugs'.

9). Toetsaanslagen worden op het beeldscherm geëchoed.

10). Er wordt aangegeven dat een commando is geaccepteerd.

11). Er wordt aangegeven dat een commando wordt uitgevoerd.

12). Indien gewenst wordt het resultaat van een uitgevoerd commando getoond.

III. Make changes in action plan and action execution possible.

Laat ruimte voor verandering van handelingsplannen en hun uitvoeringswijze.

1). Ieder commando is omkeerbaar of teniet te doen d.m.v. een herstel functie.

2). 'Kortsluit' mogelijkheden zijn aanwezig voor ervaren gebruikers.

3). Bij iedere conditie is een 'escape' mogelijkheid aanwezig.

IV. Support, if necessary, parallel execution of various activities.

Ondersteun gelijktijdig uitvoeren van verschillende activiteiten

1). De gebruiker kan 'vooruit' typen zonder op de computer te hoeven wachten.

2). Een document kan worden afgedrukt zonder dat daarbij het uitvoeren van commando's, of het verwerken van tekst gestaakt dient te worden.

3). Het programma verwerkt commando's en tekst voldoende snel zodat er zo min mogelijk gewacht dient te worden.

4). Het programma geeft, indien gewenst, informatie over de lopende activiteiten.

5). Het programma geeft een waarschuwing bij het verlaten van niet afgesloten activiteiten.

V. Offer means to support the supervisory process, especially with regard to anticipation of coming actions.

Biedt middelen voor supervisie van handelingsuitvoering i.v.m. anticipatie op toekomstige (deel)handelingen.

1). Macro's kunnen worden gemaakt door handelingen op een normale wijze uit te voeren en op te slaan, zonder dat hiervoor allerlei extra handelingen hoeven te worden verricht.

2). De laatst uitgevoerde commando's worden bewaard en kunnen weer worden opgevraagd.

3). Een overzicht van toekomstige interactiestappen wordt getoond.

VI. Take into account that people have limited capacities with respect to cognitive, sensorial, and motorial mechanisms.

Houdt rekening met de beperkte capaciteit van de cognitieve, sensorische en motorische mechanismen.

1). De cursor staat op de plaats waar ingave wordt verwacht.

2). De beeldscherm layout is consistent en uniform.

3). Foutmeldingen verschijnen op een vaste plaats op het beeldscherm.

4). Menu's bevatten minder dan 8 items.

5). Beeldscherm (kleur)instelling is wijzigbaar.

6). De cursorgrootte is instelbaar.

7). Capslock en Numlock conditie wordt op het beeldscherm aangegeven.

VII. Take into account that people strive towards lowering the level of regulating their actions.

Houdt rekening met de tendens bij de gebruiker de handelingsefficientie te optialiseren en probeer dat te ondersteunen; sta verandering in regulatieniveau toe.

1). Het programma vraagt (soms) om redundante informatie.

2). Het is mogelijk zelf gedefinieerde macro's te gebruiken.

3). Commando's kunnen via menu's of direct gegeven worden.

4). Voor eenzelfde bewerking zijn verschillende commando's mogelijk.

5). Experimenteren wordt aangemoedigd.

VIII. Try to accommodate to the user's working-style, and working-methods, and other relevant differences between users (skills, knowledge, etc.).

Zoek zoveel mogelijk aansluiting bij de (taal)kennis, vaardigheden en de manier van werken van de gebruiker.

1). De dialoog tussen gebruiker en computer verloopt in de Nederlandse taal.

2). Commando's, uitleg en informatie zijn helder en duidelijk geformuleerd.

3). Zoals de tekst op het beeldscherm staat wordt het ook geprint (what you see is what you get).

4). Alle functies en commando's zijn voldoende duidelijk zodat ze zonder nadere raadpleging van de handleiding zijn te gebruiken.

5). De gebruiker kan kiezen of hij een overzicht van de basiscommado's permanent op het beeldscherm getoond wil hebben.

Over de auteur, Fred Zijlstra

Geboren in 1956 te Bolsward. Behaalde in 1983 het doctoraalexamen in de psychologie, specialisatie Arbeids- & Organisatie psychologie, aan de Rijks Universiteit Groningen.
Was nadien enige tijd werkzaam als onderzoeksmedewerker bij de Stichting Werk en Welzijn te Leeuwarden. Hierna volgde een dienstverband bij de Faculteit Wijsbegeerte & Technische Maatschappijwetenschappen aan de Technische Universiteit Delft in combinatie met een aanstelling bij de sectie Experimentele - & Arbeids psychologie aan de Rijks Universiteit Groningen. In deze periode is het onderzoek dat in dit boek is beschreven uitgevoerd. Is sedert 1990 werkzaam als universitair docent bij de sectie Arbeids- & Organisatie psychologie van de Katholieke Universiteit Brabant. In deze periode is het Laboratorium voor Informatie Arbeid opgericht.
Publiceerde in de loop der tijd onder meer over de onderwerpen mentale werkbelasting en inspanning, taakanalyse & taakontwerp, gebruiksgemak van informatietechnologie. Huidige onderzoeksaktiviteiten hebben betrekking op de ontwikkeling van een onderzoeksprogramma waarin de invloed van onderbrekingen (storingen) op het uitvoeren van mentale arbeid wordt onderzocht.