Experimental Studies of Human-Computer Interaction:

Working memory and mental workload in complex cognition

Erik Wästlund

Department of Psychology, Gothenburg University, Sweden, 2007

© Erik Wästlund
Printed in Sweden
Department of Psychology
Gothenburg 2007
ISSN 1101-718X ISRN GU/PSYK/AVH--183—SE
ISBN 978-91-628-7066-9

For when one actually remembers,
this impression is what he contemplates,
and this is what he perceives.

~ *Aristotle*[1]

[1] Translated by J. I. Beare. Originally published in Ross, W. D. (Ed.) (1930). *The works of Aristotle* (vol. 3). Oxford: Clarendon Press.

Abstract: **Experimental Studies of Human-Computer Interaction:** Working memory and mental workload in complex cognition

Complex cognition is readily described as cognitive tasks requiring the coordination of multiple steps of processing or tasks exceeding short term memory capacity. Similarly, mental workload may be described as the use and temporary expenditure of a finite amount of information processing capacity. In the current study, consisting of eight experiments, the mental workload of complex cognition was manipulated through variations in the mode of presentation (**Study I**) with the information being presented either printed on paper or displayed on a computer screen as well as through variations in page layout (**Study II**) with the information being presented, either using a page layout designed to fit the computer screen or on a long page of scroll type. In **Study III,** the short-term memory demands of the complex cognitive tasks themselves were explored. The aim of the experiments was to investigate the effect of onscreen vs. paper presented materials on complex problem solving (**Study I**), the effect of page layout of onscreen presented materials on mental workload (**Study II**), and the configuration of short-term memory demands of complex problem solving (**Study III**).

The principal findings of the three studies may be summarized by the following points:

- Both Consumption- and Production of information is more effective when information is presented on paper rather than displayed on a computer screen (**Study I).**
- Consumption of information generates less mental workload when the page layout is adapted to fit the computer screen (**Study II:** Experiments 1 & 2).
- Problem solving processes, including both Consumption and Production of information, may be described both in terms of their reliance on either ST-WM or LT-WM (**Study III:** Experiments 1, 2 & 3) and in terms of their reliance on specific 'slave-systems' of the tripartite model (**Study III:** Experiments 1 & 3).

Taken together, **Studies I** and **II** show that the presentation of information on screen, versus in printed form, exerts detrimental effects on human information processing and that some of those effects may be attributed to differences in the navigational properties of the two media. In addition, **Study II** demonstrated that an adaptation of the page layout of the presented material so that it fitted its intended media, mental workload may be alleviated. Finally, the results of **Study III** showed that, in order to understand the memory demands of complex cognition, it is necessary to include elements of both the ST- and LT-WM paradigm of Ericsson & Kintsch and the tripartite model of Baddeley & Hitch.

Keywords: Information processing, Problem-solving, Working memory, ST-WM, LT-WM, Dual-task, Mental workload, Reading, Page layout, VDT, Convergent; Divergent.

Erik Wästlund, Dept. of Psychology, Karlstad University, SE-651 88, Karlstad, Sweden.
Tel: 0046 54 700 2528, Fax: 0046 54 83 91 65. Email: erik.wastlund@kau.se

Preface

This thesis consists of this summary and the following three papers, which will be referred to using roman numerals:

I: Wastlund, E., Reinikka, H., Norlander, T., & Archer, T. (2005). Effects of VDT and paper presentation on consumption and production of information: Psychological and physiological factors. Computers in Human Behavior, 21, 377-394.

II: Wastlund, E., Norlander, T., & Archer, T. (Submitted). Effect of Page Layout on Mental Workload: A *Dual*-Task Experiment.

III: Wastlund, E. & Archer, T. (Submitted). Working Memory Loads derived from Computer-based Primary- and Secondary-Tasks

Acknowledgements

This paper has only one author, but there would have been no paper without the help of my main advisor Trevor Archer and many others. To all those who have helped me (you know who you are); thank you. In particular, I wish to extend my sincere gratitude to colleagues and friends for their encouragement and support, to students and anonymous participants for lending me their time and lastly, and most importantly, to Carola and Hedda for filling my life with the stuff that really matters…

December 2006
Erik Wästlund

Contents

1 Introduction

The science of psychology and psychological measurement may be applied to practically all aspects of human behavior. Given the influx of computers in so many aspects of our day-to-day lives, it is no surprise that psychological tools, most notably those of cognitive psychology, have been used to understand the human aspect of human-computer interaction [HCI]. The aim of HCI studies is most often one of usability, i.e. to evaluate a user interface in order to facilitate human information processing (**Study 1**). This aim, however, may be expanded into encompassing general psychological mechanisms and thus generate knowledge, not only relating to a user interface in particular, but also about cognition in general **(Study II)**. Additionally, the tools and methods of HCI may be used without any direct interest in the user interface, but somewhat exclusively to investigate the psychology of complex cognition (**Study III**).

Although the scope of HCI today remains both wide and multi-faceted, in order to fully comprehend the effects of computers on cognition, one needs to start with the most basic aspect of computing, i.e. reading a text on a computer screen and comparing this with its predecessor - the printed page.

1.1 Problem statements

As computer usage increases, so too does the need for research into the effects of this development. Since computers are used mainly for information processing, in the professional setting at least, one critical important question is; how does computer usage influence human information processing? One obvious starting point for such a question is the comparison of information processing on a computer with its predecessor, i.e. the written paper (**Study I**).

As early as 1985, Belmore warned that text presentation on video display terminals [VDTs] was of a lower quality than text presentation on paper, leading to a poorer understanding of the material (Belmore, 1985). Furthermore, in a comparison between information searches on either paper or VDTs, the latter was shown to lead to a slower reading speed (Gray, 1991). Nine years after Belmores' study, Rice (1994) found no difference between reading from VDTs and paper with regard to text remembrance, implying that the mode of text presentation was not critical.Additionally, he found no difference regarding the level of understanding between text presented on paper and on VDTs. Nevertheless, reviews concerned with comparisons between text reading on VDTs and on paper maintained that paper text is superior (e.g. Ziefle, 1998). Explanations for the observed detrimental effects range from the quality of the equipment used in the studies, most notably the technical characteristics of cathode-ray-tube [CRT] monitors (Noyes & Garland, 2003) and line length limitations (Kolers, Duchnicky, & Ferguson, 1981), to the suggestion that reading from computer screens reduces working memory capacity (Mayes, Sims, & Koonce, 2001).

Although a few studies investigating this question are in existence, their results appeared mixed in their message. Also, points regarding the importance, and lack, of the matching of materials between the two modes of presentation have been raised. To avoid the confounding variables of page layout, e.g. line length, fonts, and kerning i.e. letter spacing, the materials in **Study I** were presented by means of a “pdf file”, either printed out on paper or displayed on a computer screen. As information processing may be divided into consumption and production of information, the investigation consists of two separate experiments, each investigating one type of information processing.

Among those that advocate that presentation on paper is superior, two explanations can be found in the literature, focusing on either (i) the quality of the equipment or (ii) the individual’s handling of the equipment. The former refers to the inferior quality of material presented on-screen, due to which the acquisition of information proceeds more slowly compared to paper presentation (e.g. Belmore, 1985; Gray, 1991). Under circumstances of a strictly limited schedule to complete the assignment, it is not surprising that conditions leading to greater time consumption in task completion will induce performance deterioration in a test of maximal performance. Nevertheless, the degree of consensus regarding the requirement for longer on-screen reading time, as opposed to printed material, is not quite convincing (e.g. Rice, 1994). The other school of thought focuses on the additional activity of handling the computer. The original or basic assignment of consumption of information is thus combined with the task of coping with computer operating requirements (Waern, 1989), resulting in a dual-task situation that may explain the observed performance deterioration (Jolicoeur & Dell, 1999; Koch & Prinz, 2002). Thus, the detrimental effects of on-screen presentation may be explained either in terms of reading speed or cognitive load. The latter may originate from additional cognitive resources being allocated to the reading process in order to compensate for the inferior presentation quality, or may be due to the dictates of the dual-task of reading and handling the computer.

In **Study II**, the focus was shifted from comparing modes of presentation to measuring mental workload associated with different types of page layout. Whereas one simply turns the page to move forward or backward through a printed document, navigation through an on-screen document entails using either a pointing device, for scrolling, or the arrow keys on the keyboard for paging (Piolat, Roussey, & Thunin, 1997). It is this difference in the navigational method that forms the basis of the proposition that the handling of the computer in conjecture with the original assignment, e.g. reading results in a dual-task situation, in turn resulting in higher mental workload. However, in addition to the activity of navigating itself, there is also the question of evaluating the effect of the procedure. Using printed material, this is simple: turn a page – end of story. On-screen documents, on the other hand, present no such clear cues regarding action completion. Viewing a document on-screen gives the impression that the document is something tangible, however a document like this is really just a text superimposed on a page resembling background graphic, and thus the dimensions of the page are, in theory, arbitrary. In practice, the page size of on-screen documents complies with printing standards such as A4 or US letter[2]. As most computer screens are unable to satisfactorily display a readable version of such a format, navigation has to be performed in steps, displaying different parts of the page. Thus, navigating through an on-screen document, in addition to the activity in question, also entails assessing the

[2] A4 (210*297 mm.) is the international standard size for printing paper. US Letter (215.9*278.4 mm.), by comparison, is slightly wider and slightly shorter.

effects of page movement. The problem is accentuated on the World Wide Web where long texts are often presented without any page-resembling background graphics at all, displaying the text in a hi-tech version of ancient text scrolls.

The purpose of **Study III** was to examine the mnemonic demands of complex problem solving. This shift from 'procedure-to-process' is motivated by an observation in **Study I** and a lack of certain predicted effects in **Study II**. In comparing the results of the two experiments in **Study I**, participants seem to have rated the mental workload of the two experiments differently. Although cross-experimental comparisons are problematic at best, the result is at least noteworthy. However, as the tasks in the two experiments of **Study I** differ in many respects, in **Study III** a set of tasks not quite so dissimilar were chosen for comparisons. Additionally, as the results of **Study II** showed no effect on the primary tasks of the secondary tasks, an additional aim of **Study III** was to evaluate the soundness of the novel secondary tasks used in **Study II**.

The eight experiments in **Studies I, II** and **III** may also be summarized in terms of the variations in information input, process demands their measurable outputs. Input was varied in terms of the mode of presentation (paper vs. screen) in **Study I**, in terms of page layout (screen vs. scroll) in **Study II**, and in terms of task configuration in **Study III**. In addition, the two experiments in **Paper 1** differed in terms of primary assignment (consumption of information in Experiment 1 and production of information in experiment 2) while the two experiments in **Study II** differed in terms of secondary task working memory load (phonological loop in Experiment I and visuo-spatial sketchpad in experiment 2). The effect of these variations in input was then measured in terms of the output of the cognitive processes. Thus, consumption of information was measured in terms of correct responses while production of information was measured in terms of the number of responses (fluency). Mental workload was measured in two ways, by means of subjective rating scales (**Studies I** and **II**), and by means of reaction times (**Study II** only). The input and output variables are summarized in Table 1.

Input			Process	Output		
Study: Exp	Comparisons	Memory Load	Assignment	Primary Task	Secondary Task	Subjective Ratings
I:1	P – S	Na	Cons.	Cr	Na	MWL
I:2	P – S	Na	Prod.	Fl	Na	MWL
II:1	Sn – Sl	Phono	Cons.	Cr	RT	MWL
II:2	Sn – Sl	Visuo	Cons.	Cr	RT	MWL
III:1	RC – Syl	Phono	Cons.	Cr	RT	---
III:2	RC – Syl	Visuo	Cons.	Cr	RT	---
III:3	He – Sni	Phono	Prod.	Fl	RT	---
III:4	He – Sni	Visuo	Prod.	Fl	RT	---

Table 1: Summary of input and output variables. P-S = Paper – Screen, Sn – Sl = Screen – Scroll, He – Sni = Headlines – Snippets, Na = Not applicable, Phono = Phonological loop, Visuo = Visuo-spatial sketchpad, Prod = Production of information, Cons = Consumption of information, Cr = Correct responses, Fl = Fluency, RT = Reaction Time, MWL = Mental workload.

1.2 Human Information Processing

In order to understand the process of working on a computer, one has to begin with basic Human Information Processing [HIP]. The most basic model of HIP consists of some type of informational input received by an individual who processes this information by means of cognitive operations and then delivers some type of output.

Figure 1: A simple general model of Human Information Processing.

In terms of the described model, **Study I** and **Study II** both examine the effects of variations in input whereas **Study III** is focused on the cognitive processes taking place between in- and output. More explicitly, **Study I** compares input presented on either a computer screen or in printed form. In **Study II**, all input is presented onscreen but instead varied in terms of graphical layout which lead to different interactive demands on the user. In **Study III**, focus is shifted from input to the memory demands of the cognitive processes which underlie all human information processing.

1.2.1 Consumption and Production of Information

From an HCI perspective, information may be seen as arising from, and returning to, the computer. From a psychological perspective, input can come from any sensory modality or from previously memorized information. In the same manner, output can be any type of action, both psychological and physiological. The processes in-between input and output, may, to a large extent, be divided into two categories: Consumption of information and Production of information. Both types of work (performance) are based upon cognition, briefly described as the process through which information is encoded, organized, stored, remembered, and applied/recalled (Martinsen & Kaufmann, 1999). Both the consumption and production of information entail the formation of new cognitive structures and novel combinations of ideas or thoughts. Ghiselin (1963) identifies both these consequences of cognitive processes as psychological aspects of creativity. The concept of creativity is complex and multi-faceted yet a fundamental distinction has been made between input-creativity, the ability or process of interpreting information and forming abstract connections between concepts, and output-creativity, the ability or process of generating new or cultivated material (Partridge & Rowe, 1994).

The input versus output-creativity distinction may be traced to Guilford (1967) who separates two forms of creative production based on the possible variations in the number of correct solutions. *Divergent* production is the production of one or more solutions, or attempts at explanations, which can be seen as samples from an infinite number of possible solutions. Since the only limitation arises from the individual's imagination, or evaluation and subsequent discarding of ideas, the possibilities are, in fact, limitless.

The absence of restrictions leads to difficulties in the qualitative evaluation of divergent production, a problem apparent in all discussions regarding art and music.

Amabile (1996) argues that a product is creative to the extent that suitable observers, independently of each other, judge the product to be creative. Suitable observers must possess expertise within the enterprise toward which the product is aimed. Thus, creativity may be seen as the quality of a product that is judged creative by competent judges. Although this may sound like a straight forward process, it is important to bear in mind that sets of judges may not agree.

Convergent production results in the synthesis of a correct solution to a problem, the right and unique answer (Guilford, 1967). Convergent production is based upon logic and deduction, two processes that are also studied within the context of reasoning (Markman, 2001). Although searching within divergent problem-solving is much broader than for convergent, the latter offers clearer criteria for goal-achievement (Guilford, 1967). Consumption of information presumes that individuals draw a specific conclusion regarding certain materials whereas production of information provides scope for any number of solutions. Consequently, the consumption and production of information may be described and measured in terms of convergent and divergent production.

The most prominent example of consumption of information, regardless of media, is of course reading, and thus consumption of information may be measured in terms of reading comprehension. It is, however, important that questions be constructed in such a manner that they do not solely tap memory but that they also require both an understanding of and deduction from the material. Production of information, on the other hand, is best described by the writing of a text. That said, it should be remembered that all writing, except possibly experimental Dadaistic poetry, also contains elements of consumption of information since the written text has to be evaluated if the plot is to evolve in a coherent manner. In the same way, assigning a title to a text document or assigning a subject line to an e-mail message entails the consumption of information, i.e. the contents of the text, in order to produce a legible a subject or title. Measuring production of information can thus be done in the form of writing assignments. In order to construct meaningful tests, of both consumption and production of information, it is, as always when constructing tests, important that the assignments bear resemblance to tasks that are familiar to the testee in question. Thus, it is pointless to ask your average participant to write a novel or read and understand a postgraduate text on quantum physics. This short description of the measurement of consumption and production of information points to two important methodological issues; one of assessment and one of generalizability. In order to successfully assess the ability of an individual, with some degree of precision, the task should include some items that the participant will pass and some that he or she will fail. When testing a group of participants, this principle must apply to all participants so optimally that all participants should pass some items and no participants should pass all of them. The question of generalizability pertains to ecological validity, i.e. if the behaviors observed in the studies reflect behaviors outside the controlled environment of the data collection situation. Ecological validity can be enhanced by using tasks that mimic tasks which the participants are accustomed to, and by presenting the tasks in a familiar setting.

Regarding the simple HIP model, what can be said about the production and consumption of information while working on a computer? First, the input arrow can be seen as symbolizing information presented on the computer screen. This seems obvious in relation to reading but, as shown, it is also important in relation to writing. The output arrow, on the other hand, may be interpreted as symbolizing information entered into the computer, either in the form of a novel text or by answering questions in a multiple-choice test. However, as consumption of information only rarely leads to some type of test, but instead is assumed to result in information being stored in memory and thus the output arrow could equally be seen as symbolizing the formation of memories. In the same manner, while production of information may be based on information presented on-screen, it can equally draw on information stored in memory. Last, but not least, both processes entail manipulation of information, either in order to organize meaningful memories or in order to produce novel ideas.

Although the classification of production processes makes it possible to discriminate between types of tasks, we need to look into the study of memory in order to understand how cognitive tasks are actually performed.

1.2.2 Working Memory

The history of memory begins with Simonides of Ceos (556 – 468 B.C.) who is attributed with the invention of visual memory aids (Yates, 1966). The aim of the technique advocated by Simonides was to memorize poems for later recitation. Although there are no records of Simonides' thoughts on the matter, the notion that memory is a function of separate subsystems was introduced by William James (1890), who distinguished primary from secondary memory, terms equivalent to the contemporary concepts of short-term memory [STM] and long-term memory [LTM]. The main difference between STM and LTM lies in their temporal qualities. Speaking with James, whose main focus was attention, the difference was one of past versus present. Today, the difference is still viewed as temporal but the STM is also seen as exerting much responsibility for the processing of information.

In its turn, the LTM is seen as consisting of several subsystems based on domains of memory. The lack of consensus regarding the classification of memory domains has resulted in the production of several LTM models (e.g. Squire, 1992; Tulving, 1999), which, despite their differences, display more commonalities. One basic distinction is made between the explicit or declarative and the implicit or non-declarative memory classes. Implicit / explicit refers to the degree of consciousness of access whereas declarative/non-declarative refers to the possibilities of communicating memory content to others. The explicit / declarative memories are divided into two classes based on personal involvement. Memories that are based on personal experience are labeled episodic memories whereas memories of facts are, in the term of Tulving (1999), labeled semantic memories. An important difference between the two concepts lies in the experiencing of the two. Remembering an auto-biographical experience, involves to some extent a 'time-lapse' back to the original incident, whereas remembering a fact does not involve any experience of the situation when the encoding took place. The implicit / non-declarative memories may, according to the view of Tulving, also be divided into two subclasses, procedural memories, which are memories for perceptual motor skills, and priming which is an association between stimuli and actions or thought.

As previously stated, the temporal qualities of STM have been known since James (1890) described the introspective studies of Ebbinghause. However, it was the findings of Miller (1956), i.e. that STM could contain 7 ± 2 pieces of information, and of the Petersons (1959), i.e. that STM-stored information was rapidly forgotten if people were distracted or otherwise preoccupied, that led to the popularization of the topic during the 1960s as well as to the inclusion of information processing as a part of STM.

Today, information processing, both in terms of on-line processing and temporary storage, is generally described and explored in terms of working memory (Duff & Logie, 2001). Most often, working memory is viewed as a multiple resource system containing several specialized subsystems. This notion is based on the work of Baddeley & Hitch (1974) who, over thirty years ago, proposed a model containing separate temporary storage systems for phonological information (the phonological loop) and visuo-spatial information (the visuo-spatial sketch pad). Activity in these storage systems is directed by the central executive which is not only seen as responsible for directing attention but also as responsible for on-line processing (Baddeley, 1996). This model has been applied successfully to a variety of cognitive activities such as language acquisition, syllogistic reasoning, mental arithmetic, and complex perceptual-motor skills (Duff & Logie, 2001).

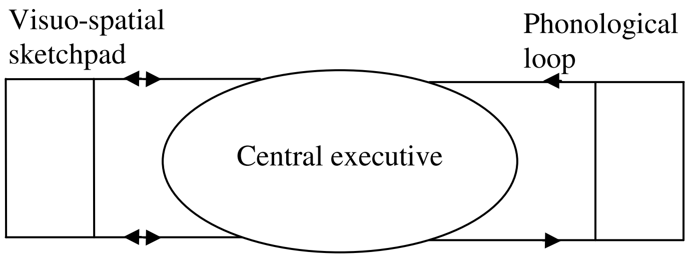

Figure 2: The working memory model proposed by Baddeley & Hitch (1974).

Despite, or perhaps because of, the success of the model, several alternative working memory models have emerged (see Miyake & Shah (1999) for a review). A specific point of divergence pertains to how the very narrow span of the working memory can account for the large processing demands made by complex problem solving. One plausible solution was the notion proposed by Ericsson & Kintsch (1995) who argued that in order to explain the working memory capacity during skilled performance, it is necessary to divide the working memory entity into a short-term working memory [ST-WM] and a long-term working memory [LT-WM]. During skilled cognitive activities, the ST-WM is supported by information processing strategies in the LT-WM and thus the ST-WM's capacity for the activity may surpass the usual constraints attributed to the working memory (Ericsson & Lehmann, 1996).

Although the models appear quite different at first glance, the ST-WM of Ericsson and Kintsch may be seen as containing the same representations of information as the 'slave-systems' of Baddeley and Hitch and whereas the latter postulates the manipulation of information as stemming from the central executive, Ericsson and Kintsch attribute this function to information processing schemata activated in the LT-WM. Irrespective of model, at the heart of the matter lies a short term memory system activated for both short time retention and as the central processing structure in the service of complex cognition. Despite the similarities of the two models and their proven merits, research into working memory seldom includes both of the models. In, **Paper III**, a more or less eclectic view of working memory will be adopted which combines the two aforementioned models. Thus, working memory will be seen as being comprised of the long and short term systems of Ericsson and Kintsch in which the 'slave-systems' of Baddeley and Hitch reside. The reason for combining and including both models, in **Paper III**, was to try to investigate complex cognition from a broader perspective than strict adherence to a specific model would allow.

1.2.3 Complex Cognition

Although there appears to be a paucity of agreement on what constitutes complex cognition (Kintsch, Healy, Hegarty, Pennington, & Salthouse, 1999), tasks exceeding typical working memory limitations (e.g. reading comprehension) and those requiring the coordination of multiple steps of processing are readily described as complex (e.g. syllogistic reasoning). The end-point of reading comprehension is to generate long-term memory structures containing integrated representations of the text. In short, the process leading up to this end-point may be described as the phonological loop of working memory being fed with single words, which are combined into phrases, sentences and so on. These phrases and sentences are then integrated with previously stored information, stemming from both the text and the long-term memory (Van Merriënboer & Sweller, 2005). Individual differences in reading comprehension have been explained as differences in working memory capacities (e.g. Daneman & Carpenter, 1980; Daneman & Merikle, 1996) but also in terms of more sophisticated and complex comprehension strategies loaded into the LT-WM (Ericsson & Kintsch, 1995; Kintsch, Patel, & Ericsson, 1999). In contrast, syllogistic reasoning is based on dual statements (e.g. Peter is taller than Paul and Paul is shorter than Mary) and the object is to determine which conclusions may be drawn from these statements. This objective is usually achieved using a third statement (e.g. Peter is taller than Mary) and participants are asked to indicate weather or not the final statement is true, false or impossible to deduce. Solving syllogistic tasks hardly leads to any long term memories of the statements or deductions. The key role of working memory in the solution of syllogisms has been demonstrated, by for instance, Gilhooly, Logie, Wetherick, & Wynn (1993) who concluded that there is a slight involvement of the phonological loop and a major involvement of the central executive.

Reading comprehension and syllogistic reasoning are thus similar insofar as they may both be shown to rely on working memory for processing, but differ in terms of LT-WM and long-term memory involvement. Additionally, they are both tasks wherein the object is to process information in order to form a valid conclusion, to which there is no alternative, which in Guilford's (1967) terms appears to constitute convergent production. Divergent production, on the other hand, involves tasks which are characterised by the lack of a single solution (Guilford, 1967).

In a similar fashion, just as convergent production relies on various cognitive processes, so too does divergent production. The writing of a long text involves a large number of cognitive processes. Information has to be retrieved from the long-term memory, reorganised via collaboration between LT-WM and ST-WM and finally be vocalised in the phonological loop. In terms of working memory demands, it has been shown that narrative texts imply more imagery and thus rely more heavily on the visuo-spatial sketchpad than do expository texts which place greater demands on the phonological loop (for a review see Olive, 2003). Additionally, as the composition of a long text entails the elaboration of information from a given starting point, it can be assumed that LT-WM processes are needed in order to form coherent texts longer than a single sentence. However, text production does not necessary entail elaboration. During 'note-taking', for instance, the object is condensation rather than elaboration. Note-taking refers to the production of short paragraphs of text containing, for instance, a shopping list or the minutes of a work meeting (Piolat, Olive, & Kellogg, 2005). It has been stipulated that the production of notes places high demands on the central executive functions as it concurrently involves comprehension, selection, and production processes. Additionally, it has been shown that 'note-taking' requires less executive mobilization than writing and that when note-taking is based on a text, the length of that text is positively correlated with cognitive effort (Piolat et al., 2005). One possible explanation for these findings could be that the amount of information concurrently being processed in the LT-WM is what defines the cognitive demands.

Independent of memory demands, information processing in a human-computer interaction setting, not only entails task completion but also the activity of controlling and inputting data into the computer, an activity which, in itself, draws on information processing capabilities. Basically, this handling consists of using either the keyboard or a pointing device such as a computer mouse or a touchpad. The type of memory process in use during a particular activity is determined by the user's experience of the activity in question and the configuration of the activity itself. Given that an activity entails few or no options and that the user is well-experienced in the activity, as is the case with a professional typist at work, motor-control (i.e. implicit memory) will be based more or less solely on procedural memory (i.e. 'how?') and will entail only small or no amounts of conscious processing. On the other hand, if an activity is new to the user and includes a series of options, as is often the case with new releases of operating systems, both motor-control and the choices will demand conscious processing (i.e. the involvement of explicit memory) and, thus, working memory resources. Additionally, there are activities during which motor-control might be automated but still reliant on conscious processing for the determination of task completion, e.g. scrolling through a long page of text.

1.2.4 Mental Workload

The concept of workload is more readily understood when used in a physiological context. Within the psychological tradition, however, the concept is rather ill-defined and applied in a variety of ways. In the context of Kahnemans' (1973) limited-capacity-model of attention, the concept of "mental workload" has been used while the term "cognitive workload" is associated with Wickens' (1984) multiple-resource-model of attention. Both models postulate a limited amount of attention that can be utilized for information processing and that, when demands exceed the available resources, processing will be impaired. The particular difference of interest between the two models lies in their views of attention, as based on single or on multiple resources. Support for the latter is derived from dual-task research showing that tasks similar in terms of input or output modality, or in terms of choice or judgment, lead to higher degrees of dual-task interference (Pashler, 1998). Today, the ‘multiple resources’ notion predominates (Xie & Salvendy, 2000). In addition, within the framework of the Cognitive Load Theory, the phrase cognitive load is used with the focus not being on attention per se but on working memory (Paas, Renkl, & Sweller, 2003).

To complicate the situation further, the depletion of information processing capacity may stem from both under and overload of stimuli (see Figure 3, for a graphical representation of the relationship between stimuli and workload)

Optimally, the amount of stimuli should fall within the range of ideal stimulation, i.e. the amount of stimuli should be adapted to an individual’s processing capacity. Such a situation leads to a low mental workload. As the amount of stimuli increases, unsurprisingly, so too does the mental workload. The same effect occurs, however, when the amount of stimuli decreases.

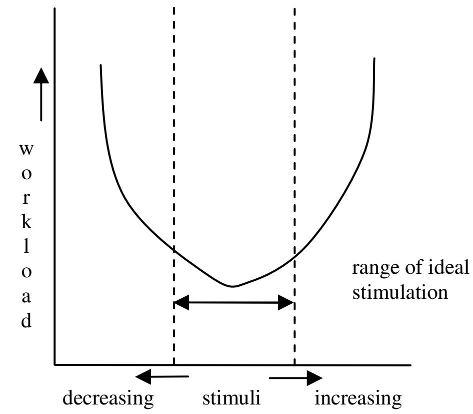

Figure 3: The relationship between amount of work and mental workload. Adapted from (Kumashiro, 1995).

Although decreases and increases in stimuli both lead to the same result, i.e. an increase in mental workload, they vary in the degree to which they draw upon the available resources. Whereas underload usually stems from stimuli that place high demands on attention but low demands on processing, e.g. monitoring for low occurrences of stimuli, overload can be the result of high demands on processing capacity in addition to high demands on attention. Examples of both under and overload can be found in the world of aviation (Wickens, Mavor, & McGee, 1997). Monitoring a radar screen for enemy aircrafts that may or may not appear, provides an example of a situation that places high demands on attention but low demands on processing capacity, thereby resulting in a high workload. Monitoring an air traffic control radar screen at an international airport, providing pilots with in-flight instructions, places an equal amount of strain on attention as well as high demands on processing capacity (Wickens et al., 1997).

In the present study, the term mental workload will be used in its most generic sense, i.e. referring to the use and temporary depletion of a finite amount of information processing capacity. The term mental workload is chosen in preference to cognitive workload and cognitive load as the latter two are to a greater extent associated with specific theories. Although Cognitive Load Theory places the source of attention within the working memory, as opposed to outside the working memory as is the case with both the attention-based models, the focus of Cognitive Load Theory is not in line with the current investigation. As the aim of the study is to investigate the effects of continuous information processing, as opposed to vigilance tasks, the depletion of processing capacity is assumed to stem from overload rather than underload.

The measurement of mental workload can be divided into three major categories (Wierwille & Eggemeier, 1993): (i) subjective measures, e.g. the NASA task load index [TLX] (Hart & Staveland, 1988) and The Subjective Workload Assessment Technique [SWAT] (Reid & Nygren, 1988) where respondents rate their subjective experience of a task on a series of scales; (ii) measures of physiological correlates, e.g. heart rate variability (e.g. S. Miyake, 2001) or eye activity (e.g. Van Orden, Limbert, Makeig, & Jung, 2001); (iii) and a wide variety of performance measures like error rates (e.g. van der Linden, Frese, & Meijman, 2003) or task completion time (e.g. Waters & Caplan, 2004). A special case of performance measures is found within the dual-task-paradigm where participants are subjected to two concurrent tasks. Previous research has shown that reaction time [RT] data offers a suitable secondary task measurement of mental workload (Verwey & Veltman, 1996; Wickens, 1984).

1.3 Notes on method

1.3.1 Dual-task experiments

A special case of performance measures is found within the dual-task-paradigm where participants are subjected to two concurrent tasks. Depending on the set-up of the experiment, different types of inferences may be made from the results. Within working memory research an often used technique is to investigate the memory demands of a certain primary task by testing the effects of various continuous secondary tasks previously shown to draw on a certain working memory resource. The result of two tasks taping the same resources is a performance deterioration on the performance of the primary- and/or secondary task. Frequently used secondary tasks include the *random generation task* whereby the participant generates random sequences of numbers or letters (the central executive), the *spatial tapping task* whereby the participant is instructed to tap sequentially the four corners of a square using a finger (the visuo-spatial sketchpad), and *articulatory suppression* whereby the participants are instructed to repeat syllables, words, or phrases aloud (the phonological loop) (A. Miyake & Shah, 1999).

An alternative to continuous secondary task is the application of probing tasks, whereby the participant is instructed to engage in the primary task, with intermittent interruptions by a secondary task probe. Previous research has shown that reaction time [RT] data to probing secondary tasks is a suitable measurement of mental workload (Verwey & Veltman, 1996; Wickens, 1984). The rationale behind using RTs as a measure of mental workload is that the amount of capacity demanded by the primary task will determine the reaction time to the occurrence of the secondary task (Brunken, Plass, & Leutner, 2003). Thus, by increasing the mental workload of the primary task, the reaction times to the secondary task ought to increase accordingly. The dual-task experiment may also be designed *in reverse*, i.e. if the mental workload of the secondary task is increased, performance on the primary task will decrease. The influence of the secondary task on the primary task has raised some concern regarding the techniques' usefulness (Paas, Tuovinen, Tabbers, & Van Gerven, 2003). However, if both primary and secondary task workloads are varied, the cognitive nature of the tasks themselves may be analyzed. In accordance with the Additive-Factor Method (Sternberg, 2001), interactions between the two tasks indicate that they are based on the same cognitive resource, whereas if the reaction times are additive, the tasks will utilize different mental resources.

Although the secondary tasks described above have been shown to be effective, they are rather impractical since participants need to be constantly monitored by the researcher as well as in solitude in order not to disturb each other. Thus, in **Studies II & III**, alternative secondary tasks are used.

1.3.2 Computerized experiments

As noted above, computers and computer applications are being used increasingly, this also being the case within the scientific community. Within the various disciplines of the behavioural sciences, computers are not only applied as tools for text processing and computing data, but also as a means of conducting experiments. The benefits of using computers in experimentation include controlling information, i.e. treating all subjects equally and the possibility of precise timing, both in terms of presenting materials as well as for the timing of responses. In addition, by utilizing web technology, computerized experiments allow the simultaneous testing of a theoretically infinite group of participants. It is important to note that utilizing web technology does not imply that an experiment will be made public via the Internet and thus open to unrestricted participation. Through *distributed web experimenting,* researchers conduct web-based experiments using participants recruited and assisted by local collaborators (Reips, 2002). At the outset of the experiment, the participants or collaborators log on to the same password-protected Internet-based experiment, thus access to the material is strictly regulated. Such an experiment can easily be transformed into an 'open-to-all' web experiment by removing the password protection and advertising the existence of the project on relevant websites. Taking experiments into the public domain via the Internet poses special problems, e.g. self-selection and diminished commitment to fulfilling the experiment. However, there are also benefits, e.g. the possibility to reach large numbers of participants of varying demographic backgrounds.

Web-based data collection utilities can be divided into server-side and client-side utilities (Reips, 2002). Server-side utilities include a web server for the storage and distribution of the experimental material as well as database applications for data collection and the distribution of dynamic content. Conducting an experiment solely on the server side, e.g. by means of a webpage, may lead to experimental errors in time measurement, due to lags in Internet traffic and visual discrepancies between various computer platforms. The solution to this is using client-side utilities, i.e. materials that are downloaded from the server and run on the participants' computers.

The ELMA CyberLab[3] web application, utilized in **Study II and Study III,** is a combination of server and client-side applications. When the participant clicks on the "start experiment" link, a small program is downloaded and run on his or her computer. The program then contacts the server-side database for group allocation and, based on this information, the participant is then shown a specific sequence of instructions and assignments. Throughout the experiment, ELMA CyberLab collects input from the user, in terms of both responses to questions/tasks and reaction times to various stimuli. After the experiment, the collected data is sent back to the server for storage in the database. The experimenter may at any time download the collected data from the database and import it into suitable statistical software for analysis.

2 Summary of studies

2.1 Introduction

In the following three papers, a total of 374 participants took part in eight experiments. **Study I** describes two experiments, each with 72 participants, concerning the effects of mode of presentation (paper vs. computer screen). **Study II** describes two experiments, with final samples consisting of 42 and 40 participants, respectively, concerning the effect of page layout. **Study III** describes four experiments regarding the memory demands of complex cognition with final samples consisting of 40, 32, 44, 32 participants, respectively.

2.2 Study I: Effects of VDT and paper presentation on consumption and production of information

General aim

The following two experiments were designed to measure the effects of mode of presentation (paper vs. computer screen) on consumption (Experiment I) and production of information (Experiment II). In addition, the participants rated their subjective experience of the assignments.

2.2.1 Experiment 1

Design

A 2*2 factorial design was used with Type of Presentation (Paper or Computer) and Gender as between-group variables. Dependent variables were consumption of information, measured in terms of correct responses on a reading comprehension test and ratings of subjective experience.

Instruments

a) Pre-experimental questionnaires to ensure group homogeneity:
   i. Demographics. A simple demographics survey containing questions regarding gender, age, university credits, and participants' computer usage.

[3]For the technically-minded: ELMA CyberLab consists of a series of Macromedia Flash files that communicate with a MySQL database via php scripts. ELMA CyberLab has been designed by the author.

ii. TRI – *Technology Readiness Index.* The TRI test (Parasuraman, 2000) is a multiple item scale design to measure the individual's propensity to embrace new technology.

iii. SE – *Stress and Energy*. The SE instrument allows individuals to rate their subjective experience of energy and stress (Kjellberg & Iwanowski, 1989).

iv. PANAS – *Positive- and negative affect scale.* The PANAS instrument (Watson, Clark, & Tellegen, 1988) measures the individual's degree of positive and negative affectivity through ratings using twenty adjectives.

v. PreSTH – *Stress –Tiredness – Hunger*. Three questions measuring how tired, stressed, and hungry subjects felt. Subjects were required to mark, with a cross, the point on a VAS scale that they considered most appropriate.

b) Dependent variables:

i. *The Higher Education Entrance Examination READ test.* The READ test is designed to measure Swedish language reading comprehension. The test contains five texts (averaging 1,000 words) each followed by four multiple choice questions. Participants where given thirty minutes to complete this test.

ii. PostSTH – *Stress –Tiredness – Hunger*. Three questions measuring how tired, stressed and hungry subjects felt. Subjects were required to mark, with a cross, the point on a VAS scale that they considered most appropriate.

Procedure

The experiment was conducted over a two-day period. On the first day, the text assignments were distributed via computer screen and on the second day, via paper. Since the participants were allowed to choose which day and time they would participate, they also, unwittingly, chose which assignment they would take part in. Upon arrival, the participants were instructed to turn off their cellphones and "take a seat". In front of them, they found an envelope containing the test materials and instructions relevant to all tests. The only difference between days one and two was that, during day one, the text assignments were presented on-screen by means of a pdf document, with the same document being presented in printed form during day two.

Results

The results of the data analysis showed that Type of Presentation had significant effects on both the READ test, where participants in the Paper assignment scored higher, and on the Post STH test, where participants in the Computer assignment reported higher ratings on the stress and tired scales and lower ratings on the hunger scale.

2.2.2 Experiment 2

Design

A 2*2 factorial design with Type of Presentation (Paper or Computer) and Gender as between group variables was used. Dependent variables were production of information, measured in terms of the number and quality of the responses during the "Headlined test" and ratings of subjective experience.

Instruments

a) Pre-experimental questionnaires to ensure group homogeneity:
   i. Demographics. (See Experiment 1).
   ii. TRI – *Technology Readiness Index.* (See Experiment 1).
   iii. SE – *Stress and Energy*. (See Experiment 1).
   iv. PANAS – *Positive- and negative affect scale.* (See Experiment 1).
   v. PreSTH – *Stress –Tiredness – Hunger*. (See Experiment 1).

b) Dependent variables:
   i. The verbal creativity test "Headlines" measures divergent production (Ekvall, 1969). The test assesses the ability to transform one unit of information into another, e.g. deriving from a complete newspaper article the essential text that might constitute a headline. The test requires the testee to write down as many headlines as possible for short newspaper articles; in the test's original design, this involved four newspaper articles where the time allocated to each article was three minutes. In the present study, this was modified to encompass 10 articles with a total time limitation of thirty minutes to equal the assignment time of Experiment 1. In addition to the quantitative judgements, the number of headlines, the test also included a qualitative assessment carried out by two journalists who subjectively rated every headline on a scale of 1 to 5.
   ii. PostSTH – *Stress –Tiredness – Hunger*. (See Experiment 1).

Procedure

(See Experiment 1).

Results

The results of the data analysis showed that Type of Presentation had a significant effect on the Headlines test whereby participants in the Paper assignment produced more headlines but that there were no differences in the quality of those headlines. Neither were there any significant effects on the Post STH test although there was a trend indicating that the participants in the Computer assignment reported higher ratings of stress.

2.3 Study II: The Effect of Page Layout on Mental Workload

General aim

The following two dual-task experiments were designed to measure the effects of the page layout of computer-presented text documents and the effects of working memory load (phonological loop in Experiment 1 and visuo-spatial sketchpad in Experiment 2) on reading comprehension and mental workload.

2.3.1 Experiment 1

Design

A 2*2 mixed factorial design was used with Layout (Screen / Scroll) as within group variable and PopType (PopUp / PopMem) as between group variable was used. To avoid possible order effects, Layout was balanced for order and text order, i.e. there were four subgroups in both PopType groups.

Instruments

a) Pre-experimental instruments to ensure group homogeneity:

i. Demo. A simple demographics survey containing questions about gender, age, university credits, number of times the participant had taken the Higher Education Entrance Examination, average hours of computer use per week, vision and vision correction.

ii. PreRightNow. The first instance of the RightNow test (see dependent variables for full description).

b) Independent variables:

i. Layout. To investigate the effects of page layout, the primary task text assignments were created in two versions (Screen and Scroll). In the Screen version, the text was divided up so that each page was the same size as the computer screen. In the Scroll version, the text was not divided up at all, i.e. it was presented as one long document.

ii. PopType. To investigate the effects of pre-loading the phonological loop, two types of pop ups (PopUp and PopMem) were used. In the PopUp assignment, participants had to respond to pop ups by simply clicking on a button labelled 'close'. In the PopMem assignment, the participants were first presented with a two digit target number. The following pop up showed a number, which may or may not have been the same number. The participant was then instructed to indicate whether or not this was the target number by clicking on a 'yes', 'no', or 'don't remember' button. Having done so, a new target number was then presented and the participant was allowed to continue reading the text.

c) Dependent variables:

i. Primary task. The primary tasks are text assignments forming part of the Higher Education Entrance Examination – READ. The test is designed to measure Swedish language reading comprehension. The original test consists of 10 pages containing five different texts, averaging 1,000 words, which are each followed by four multiple-choice questions. In this experiment, two of these texts were used. As the scoring is a simple summation of the correct answers, the respondents can receive a maximum of 4 points in each of the two assignments.

ii. Secondary task. During the course of the text assignments, the participants are interrupted by 10 pop ups covering the screen, with the measurement of interest being the mean reaction time to these.

iii. RightNow. This test is designed to measure the participants' current state of stress, general tiredness, and optical fatigue. The questions were phrased as thus: How *state in question* are you feeling right now? The participants were asked to respond to the questions using a Visual Analogue Scale [VAS] (range 0 – 100), with two extremities and a midpoint. The RightNow test was presented on three occasions during the experiment: i.e. before the first text assignment, and both before and after the second text assignment. It is thus possible to calculate the effect of Layout (Screen / Scroll) on the three states by subtracting the value reported post-assignment from the value reported pre-assignment.

d) Post-experimental evaluation:
   i. Performance. This test consisted of two questions, one positively and one negatively worded, which allowed the participants to rate, on a five point Likert scale, their own performance during the two assignments. The participants were asked to rate their performance during both the first and second text assignments.

Procedure

The experiment was conducted in a computer room containing 16 computers at Karlstad University. When the participants arrived, the only verbal instructions they were given were to "take a seat", to turn off their cellphones, and to look for all the information they might require on screen. Once all the participants had done this, they were then instructed to click the 'start experiment' button on their computer screens. When the participants started the experiment, the ELMA CyberLab randomly assigned them to one of the two PopType groups and to one of the four Layout subgroups.

Results

The results showed that there were no significant effects on reading comprehension but that the scroll page layout had led to a significantly higher mental workload.

2.3.2 Experiment 2

Design

(See Experiment 1).

Instruments

a) Pre-experimental instruments to ensure group homogeneity:
   i. Demo. (See Experiment 1).
   ii. PreRightNow. (See Experiment 1).

b) Independent variables:
   i. Layout. (See Experiment 1).
   ii. PopType. To investigate the effects of visuo-spatial load, two types of pop ups (PopSolE / PopSolD) were used. Both PopTypes are designed like a simple game of solitaire with a stack of four cards, one from each suit, face down located at the top of the screen. Below the cards, there are four foundations marked with the four suits. Upon clicking on the back of the top card, this turns over and the object is to drag this card to the corresponding foundation. The difference between the two PopTypes lies in the degree of difficulty in determining which foundation the card should be moved to. In the easy PopType (PopSolE), the cards in the stack were all aces and lay in the same order as the foundations, when viewed from left to right. In the difficult PopType (PopSolD), both value and suit were randomized.

c) Dependent variables:
   i. Primary task. The primary tasks were the same as in Experiment 1, except for the inclusion of two additional questions to each text, making the possible total 6 points for each text.
   ii. Secondary task. (See Experiment 1).
   iii. RightNow. (See Experiment 1).

d)Post-experimental evaluation:
   i. Performance. (See Experiment 1).

Procedure
(See Experiment 1).

Results

The results replicated those of Experiment 1, showing that there were no significant effects on reading comprehension, but that the scroll page layout led to a significantly higher mental workload.

2.4 Study III: Working Memory Loads derived from Computer-based Primary- and Secondary-Tasks

General aim

The following four dual-task experiments were designed to examine the memory demands of complex cognition in regards to both consumption and production of information. In Experiments 1 and 2, Reading comprehension is contrasted with Syllogistic reasoning. Additionally, the role of the phonological loop is investigated in Experiment 1 and the role of the visuo-spatial sketchpad is investigated in Experiment 2. In Experiments 3 and 4, text production is contrasted with text condensation. Also, the role of the phonological loop is investigated in Experiment 3 and the role of the visuo-spatial sketchpad is investigated in Experiment 4.

2.4.1 Experiment 1

Design

A 2*2 mixed factorial design was used with the primary tasks as within group variable and the secondary tasks as between group variable. To avoid possible order effects, the primary tasks were balanced for order, i.e. there were two subgroups in both secondary task groups.

Instruments

a) Pre-experimental instruments to ensure group homogeneity:
   i. Demo. A simple demographics survey to enable testing of group homogeneity and consisting of questions regarding gender, age, university credits, number of times the participant had sat the Higher Education Entrance Examination, average hours of computer use per week.

b) Primary tasks:

The objective of both the primary tasks (Reading Comprehension and Syllogistic reasoning) was to measure information processing using multiple choice questions. The difference between the two primary tasks lies in the amount of information that is necessary to process in order to answer the questions.

i. The primary task Reading Comprehension forms part of the Higher Education Entrance Examination – READ (a Swedish equivalent of the SAT test). This test is designed to assess Swedish-language reading comprehension. The original test consists of 10 pages containing five different texts, averaging 1000 words, with Reading Comprehension being measured via four multiple-choice questions per text. In the current experiment, one such text is used. As the scoring is a simple summation of the correct answers, the respondents could receive up to 4 points The questions are devised in order to be as varied as possible and are distributed in such a way that the whole text is used: it is thus necessary to form an overall understanding of the material in order to answer the questions. The text assignments are constructed in order to cover three cognitive levels; remembering, understanding, and deduction, where understanding and deduction predominate.
ii. The primary task Syllogistic Reasoning (Holmquist, 1974) used in this experiment contains forty syllogisms all of which include two premises (e.g. A is smaller than B and B is smaller than C) and a question (e.g. Is A smaller than C?). The respondent's task was to answer "Yes", "No" or "Impossible to know" using multiple-choice questions. Half of the syllogisms contain premises based on letters while the other half were based on names (e.g. Peter is taller than Paul. Paul is taller than Mary).

c) Secondary tasks:

To investigate the effects of increased memory load on the phonological loop, two types of pop-ups (PopUp and PopMem) were used. In both assignments, responses were performed by placing the mouse pointer over the relevant button and clicking the left mouse button. During the course of the each primary task, the participants were interrupted by 10 pop ups. PopUp incidence was based on a list of ten intermissions varying between 30 and 76 seconds. The order of the intermissions was randomized before each trial making it seem like the pop-ups appeared at random, from the participants' point of view. The dependent variable of the secondary task was the mean reaction time of the responses to PopUp occurrences. In order to minimize error variance, the trimmed mean was used, i.e. the highest and lowest RT value of each participant was excluded, thus, the individual mean RT was based on 8 of the 10 responses.

i. In the PopUp assignment, the participants were instructed to respond to pop-ups by simply clicking on a button labelled 'close'.
ii. In the PopMem assignment, the participants were first presented with a two-digit target number. The next pop-up showed a number which was either the same or different. The participant was then instructed to indicate whether or not this was the target number or not by clicking on a 'yes', 'no', or 'don't remember' button. Having done so, a new target

number was then displayed and the participant was allowed to continue reading the text.

d) Post-experimental evaluation:
   i. Performance. This test consisted of two questions, one positively and one negatively worded, which allowed the participants to rate, on a five point Likert scale, their own performance during the two assignments. The participants were asked to rate their performance during both the first and second text assignments.

Procedure

The experiment was conducted in a computer room containing 16 computers at Karlstad University. When the participants arrived, the only verbal instructions they were given were to "take a seat", to turn off their cellphones, and to look for all the information they might require on screen. Once all the participants had done this, they were then instructed to click the 'start experiment' button on their computer screens. Upon initializing the experiment, ELMA CyberLab randomly allocated each participant to one of the two secondary task groups and to one of the two primary task subgroups.

Results

The results showed that preloading the phonological loop had a detrimental effect on Syllogistic reasoning and that mean RTs were shorter during the Syllogistic reasoning task than during Reading comprehension.

2.4.2 Experiment 2

Design

(See Experiment 1).

Instruments

a) Pre-experimental instruments to ensure group homogeneity:
   i. Demo. (See Experiment 1).
b) Primary tasks:
   i. Reading comprehension (See Experiment 1).
   ii. Syllogistic reasoning (See Experiment 1)

c) Secondary tasks:

In order to investigate the effects of visuo-spatial load, two types of pop-ups (PopUp / PopUpV) were applied. In order to magnify the load on the sketchpad, the pop-ups occurred in sequences of four with half-second intervals between the click to close the pop-up and its next appearance, thus resulting in a total number of forty secondary task responses. However, in comparisons of secondary tasks the reaction time measurement used is from the appearance to closure of the first in the sequence.

   i. The secondary task PopUp was identical to the one used in Experiment 1.
   ii. PopUpV was also identical in all respects, except for the position on the screen where it was displayed. Instead of constantly popping up in the centre of the screen, the position of the PopUpV was randomized. Thus,

in order to click on the close button and continue, the participants were required to identify the exact position on the screen to were the pointer should be moved and then execute the movement.

d) Post-experimental evaluation:
  i. Performance. (See Experiment 1).

Procedure

(See Experiment 1).

Results

The results showed no effect of visuo-spatial load on any of the two primary tasks. The results of the secondary task analysis, however, replicated the results of Experiment 1 were mean RTs were lower during Syllogistic reasoning than during Reading comprehension.

2.4.3 Experiment 3

Design

(See Experiment 1).

Instruments

a) Pre-experimental instruments to ensure group homogeneity:
  i. Demo. (See Experiment 1).
  ii. LetterRain. The object of this test is to measure the participants' keyboard skills in order to ensure group homogeneity. During this one minute test, letters *rain* down from the top of the screen (with a maximum of four letters on the screen simultaneously) and the respondents task is to press the keys corresponding to the letters. When the correct key is pressed, the letter disappears and a new letter appears at the top of the screen. The dependent variable was the number of correct keystrokes.

b) Primary tasks:

The primary tasks Snippets and Headlines are both verbal creativity tests measuring divergent production. The main difference between the two tasks lies in the amount of information to be processed and produced for task completion. During both primary tasks, each encompassing a duration of ten minutes, the participants can switch between the three stimuli at any point. The dependant variable was character fluency.

  i. The primary task, Snippets, assesses the ability to transform and elaborate information. Participants are presented with three headlines and the task is to write a newspaper snippet that might accompany the given headline. The test is a variation of the ESL composition task (Sandlund, Linnarud, & Norlander, 2001), where the participants are given four random words as a starting point.
  ii. The primary task, Headlines (Ekvall, 1969), assesses the ability to transform one unit of information into another, in this case to condense,

from a newspaper snippet, the essential text to provide a headline. During the task, the participants are presented with three snippets and the object of the test is to construct as many headlines as possible for the snippets.

c) Secondary tasks:
   i. PopUp. (See Experiment 1).
   ii. PopMem. (See Experiment 1).

d) Post-experimental evaluation:
   i. Performance. (See Experiment 1).

Procedure

(See Experiment 1).

Results

The results showed no effect of preloading the phonological loop. The results of the secondary task analysis, however, showed that mean RTs were shorter during then Headlines test than during the Snippets test.

2.4.4 Experiment 4

Design

(See Experiment 1).

Instruments

a) Pre-experimental instruments to ensure group homogeneity:
   i. Demo. (See Experiment 1).
   ii. LetterRain.

b) Primary tasks:
   i. Snippets (See Experiment 3)
   ii. Headlines (See Experiment 3)

c) Secondary tasks:

In order to investigate the effects of visuo-spatial load, two types of pop-ups (PopUpCN / PopUpVN) were applied. The object of the secondary tasks was the same as in Experiment 2 i.e. to investigate the effects of visuo-spatial load. However, as the response execution of the primary tasks differs between Experiment 2 (mouse only) and Experiment 4 (keyboard only). The PopUps were adapted so that responses were executed by means of pressing keys instead of clicking on the muse. All pop-ups presented the participant with a number (0-9), upon which he or she had to press the corresponding key in order to continue with the primary task.

   i. During the low-load task (PopUpCN) the PopUp occurred at the same position on the screen every time and the number was always the same (0).
   ii. During the high-load task (PopUpVN), the pop-ups occurred at random positions and displayed a random one-digit number.

d) Post-experimental evaluation:
   i. Performance. (See Experiment 1).

Procedure

(See Experiment 1).

Results

The results showed that a visuo-spatial load had a grater detrimental effect on the Snippets task.

3 Discussion

3.1 Introduction to conclusions

The main findings of the three studies can be summarized in the following points:

- Consumption of information is more effective when the information is presented on paper than when displayed on a computer screen (**Study I:** Experiment 1).
- Equally, production of information is more effective when the information is presented on paper than when displayed on a computer screen (**Study I:** Experiment 2).
- Consumption of information generates less mental workload when page layout is adapted to fit the computer screen (**Study II:** Experiments 1 & 2).
- Problem solving processes, including both Consumption and Production of information, may be described both in terms of their reliance on either ST- or LT-WM (**Study III:** Experiments 1, 2 & 3) and in terms of their reliance on specific 'slave-systems' of the tripartite model (**Study III:** Experiments 1 & 3).

Since the basic aim of any experiment is to draw conclusions regarding the independent variable on the basis of measurements of the dependent variable, it is important to understand the configuration of the latter. The primary tasks, consumption and production of information, are both tests of maximal performance, i.e. they measure an individual's capacity to perform a given assignment. The object of the experiments, therefore, is to influence this capacity, to such extent that it is measurable, through variations in the mode of presentation (**Study I**) as well as the page layout (**Study II**) or through the use of secondary tasks (**Studies II & III**).

The main finding of **Study I** Experiment 1 was that consumption of information, measured via a reading comprehension test, is more difficult when the assignment text is presented on a computer screen than when it is presented in printed form. In other words, presentation of the material on a computer screen reduces an individual's capacity for consumption of information. The notion that the capacity for a test of maximal performance may be influenced rests on the assumption that there is a limit to the processing capacity of the working memory.

The main finding of **Study I,** Experiment 2, indicated that production of information, as assessed by testing creative production, resulted in a higher level of fluency when the material was presented on paper, albeit without affecting the quality of the headlines. Essentially, this finding confirms that of **Study I,** Experiment 1, and may be explained in the same manner**.** Nevertheless, methodological discrepancies between the two studies require some consideration. In addition to the assignments' different forms, i.e. consumption versus production of information, there was a discrepancy in the time dimension. The consumption of information assignment was arranged so that the participants were allowed 30 minutes in which to complete the assignment, whereas the production of information assignment was arranged so that the participants were required to spend 30 minutes on the task, the former necessitating completion and the latter requiring only compliance. In addition, the amount of text in the production of information assignment was less than 10% of the amount of text in the consumption of information assignment. It is important to bear in mind that the performance deterioration in the divergent problem-solving task only affected fluency and not quality. The lack of differences in assessed quality may be an effect of the assignment configuration, with regard to either the shorter period of time that participants were actually handling the computers, or to the assignment's divergent nature. Divergent production demands flexibility in the thought processes (Lubart, 2001). It is thus possible that the dual-task nature of the original assignment, combined with and interrupted by handling the computer, facilitates flexible thinking by ensuring that individuals move their focus of thought between different avenues.

In sum then, despite the methodological discrepancies observed, both **Study I,** Experiment 1, and **Study I,** Experiment 2, showed that on-screen presentation impaired performance. The finding that the participants in **Study I,** Experiment 1 (production of information), performed and responded in a similar fashion to the participants in **Study I,** Experiment 2 (consumption of information), in spite of the considerably smaller amount of text in the assignment, may be interpreted as an indication that cognitive load on a limited-capacity working memory, as opposed to reading speed, is a central component of the detrimental effects of text presented on computer screens. This notion corroborates that of Mayes et al. (2001), who showed that cognitive workload plays an important role for performance in computer-aided environments.

Since the documents used in **Study I** were based on the A4 format, the question is, then, whether or not the layout aspect of navigation can be identified as a source of the demonstrated detrimental effects of on-screen presentation. Based on this observation, the aim of **Study II** was to investigate the effects of page layout. It is again important to point out that the nature of the task of consumption of information, as measured through a test of reading comprehension, is one of measuring maximal performance and that the aim of the experiment is to manipulate and measure the ability to complete the task. Thus, in order for the manipulation to have a measurable effect, it needs to be of such magnitude that the mental workload is not just increased but increased to such an extent that performance is impaired. The results of both experiments in **Study II** indicate that performance was not affected by page layout to any measurable extent. However, it is important to remember that the reading comprehension task in **Study I,** Experiment 1, consisted of a total of 20 questions and that the participants worked on their assignment for 30 minutes. In the experiments in **Study II,** the timeframes and the number of questions were considerably smaller (10 minutes and 4 questions in Experiment I, and 7 minutes and 6 questions in Experiment II). Thus, in comparison, the manipulation was weaker and the measurements less precise in **Study II**.

In **Studies II** and **III,** a secondary task was added to the primary tasks of consumption and production of information. If the secondary task, which is also a task of maximal performance, utilizes the same memory resource as the primary task this consideration ought to lead to impairment of at least one of the two tasks. However, in order to generate measurable detrimental effects the two tasks have to, not only draw on the same resources, but also exceed available resources. Thus, as in **Study II**, when the primary task is manipulated, but not to the extent that it becomes impaired, the effects of manipulation on mental workload can be measured through the secondary task (Brunken, Steinbacher, Plass, & Leutner, 2002; Marcus, Cooper, & Sweller, 1996). In **Study III** on the other hand, the different primary tasks are contrasted with each other rather than slightly manipulated making the interpretation of RTs more complex as the differences might just as well stem from differences in the tasks memory demands as their inherent mental workload.

In terms of detrimental effects on the primary task by the secondary tasks the result of **Study III,** Experiment 3, show that the secondary task utilizing resources of the visuo-spatial sketchpad has detrimental effects on text elaboration and **Study III,** Experiment 1, show that syllogistic reasoning is impaired by pre-loading the phonological loop. No effects on the primary tasks of snippets (**Study III**) and reading comprehension (**Studies II and III**) were, however, observed. The latter results are in line with previous research showing that neither brief interruptions (Glanzer, Dorfman, & Kaplan, 1981) nor a three-digit working memory load (Baddeley & Hitch, 1974), while reading, has any significant effect on comprehending the material. The result of **Study III,** Experiment 1, however, appears to contradict a previous finding (Gilhooly et al., 1993) which asserted that the phonological loop only played a minor role in syllogistic reasoning. On closer inspection of the methodology and model used by Gilhooly et al. (1993), the findings do not seem so disparate, but rather warrant a reinterpretation of the their interpretations.

Taken together, these findings from **Studies II and III**, i.e. that the additional processing demands made by a secondary task on the ‘slave-systems’ of the Baddeley and Hitch model only affect problem solving tasks relying on the specific ‘slave-system’, confirms rather than elaborates the present notions of working memory processes. Nevertheless, the type of methodology introduced, i.e. web-distributed experimentation, offers a novel approach in this area of research. The advantages of the method not only provide the possibility of testing groups of individuals simultaneously, but also the opportunity of creating secondary tasks with a higher ecological validity. Although, concurrent articulatory suppression has been proven to be an effective secondary task, it hardly mimics any real-world phenomena, whereas different types of

PopUps occur quite frequently when using a computer and thus offer the clear potential for a promising avenue of development.

In terms of RTs, the main finding of **Study II,** Experiments 1 and 2, indicated that consumption of information generates less mental workload, as measured through reaction times to the secondary task, when page layout is adapted in order to fit the computer screen. The rationale used in this interpretation of the reaction time results is that the level of mental workload determines the cognitive resources available for reacting to the occurrence of a secondary task stimulus (Brunken et al., 2002; Marcus et al., 1996; Verwey & Veltman, 1996; Wickens, 1984), in this case the “PopUps”. There are two plausible explanations for the observed workload differences: Firstly, manipulating an on-screen text document via scrolling necessitates a shift of focus from the text to the action of controlling the page movement. The main difference between the screen and scroll layouts is the amount of visual cues facilitating control of the page movement. In the absence of visual cues, the reader is obliged to match the movement of the page with lines of text, which in turn has to be matched against the memory of the text, in order to determine whether or not the text has been moved a satisfactory distance or whether more scrolling is required. In terms of visual cues, the A4 format falls between the screen and scroll layouts as it is possible to use the top (but not the bottom) of the page as a meaningful navigational cue. Secondly, the absence of visual cues impedes the encoding of information in a two-dimensional space, thereby rendering the decoding of meaning more cumbersome (Piolat et al., 1997). It is noteworthy that, despite the fact that no differences were shown in terms of the primary task, a significant effect was found with regard to the secondary task. On the one hand, these findings support the notion of using dual-task experiments to investigate mental workload. However, on the other hand, they also raise further questions regarding the paper vs. computer issue that pertains to possible Type II errors in studies showing that no differences exist.

In terms of RTs the main finding from **Study III,** Experiments 1, 2 and 3, was that problem solving processes that draw more heavily on LT-WM memory elicit longer reaction times than do processes which are processed in the ST-WM. This interpretation is based on the finding of Cantor & Engle (1993), who not only showed that long term memory structure activation is an integral part of working memory performance, but also that the degree of activation may be measured through the use of reaction times. It is noteworthy that this finding was observed both when comparing reading comprehension with syllogistic reasoning (Experiments 1 and 2) and when comparing the Snippets and Headlines tasks (Experiment 3). It is equally noteworthy that the effect was shown independent of secondary task ‘slave-system’ load (phonological loop in Experiments 1 and 3 and visuo-spatial sketchpad in Experiment 2). Given the rather low power of the experiments, the fact that the effect was observed in three out of four experiments, despite there inherent differences, indicates that this is a reliably stable effect. As the primary tasks contrasted in **Study III,** have been shown to utilize different memory resources through the use of secondary tasks, it is more feasible that the observed differences in mean RTs are more a result memory demands than attributable to mental workload.

The use of a secondary task is not the only way of measuring mental workload. A more frequently used method consists of various forms of subjective rating scales. Ratings of subjective experiences usually take one of two forms: either the task is appraised or the respondent rates his or her current state. The two methods can be exemplified using the questions: "how cumbersome was the task" and "how tired did the task make you". Although the former can generally be said to be more precise, there is a greater risk of task appraisal than of rating the actual task. State rating, on the other hand, is less afflicted by task appraisal. However, on the other hand, it has less precision as it is not a rating of the task itself but of the affects of the task on the individual. In both **Study I** and **Study II**, the participants' subjective ratings of their current states were included.

The results of **Study I,** Experiment 1 (consumption of information), demonstrated that participants rated their subjective states of stress and tiredness higher in the group where the material was presented on-screen than when the material was presented in printed form. In **Study I,** Experiment 2 (production of information), no significant differences were observed but there was a trend indicating that the participants in the on-screen group reported higher levels of experienced stress. In the two experiments in **Study I,** rating was performed after task completion and the results were compared between the groups. As **Study II** involved both repeated measurement and counter balancing of the task order, ratings were carried out on three occasions, before and after the first task and then again after the second task. The difference between before and after a given task was then used as a measure of the task's effect on the individual. The results of the two experiments in **Study II** show no significant differences: neither between the type of layout nor between the type of secondary task. It is again important to remember the differences in respect of the assignment times in **Studies I** and **II**. The absence of any observed differences in **Study II** may be interpreted as being either non existent or too small, due to the short assignment time, and thus difficult for participants to evaluate and report (Mayes et al., 2001).

The observed increased stress and tiredness levels in **Study I,** Experiment 1, may be explained in the same way as the observed performance deterioration. Either the focus is on the quality of on-screen presented material, which necessitates a greater mobilization of both perceptual and executive cognitive resources in order to compensate for the deficiency of the information presentation, or one may focus on the dual-task of handling the equipment and the performance of the original assignment, which necessitates a greater mobilisation of cognitive resources to compensate for the added mental workload. Furthermore, in the light of the results of **Study II**, at least some of the added workload may be attributed to the differences in page display and their consequences for document manipulation. Irrespective of which particular mode of explanation is embraced, the heart of the matter pertains to an increase in mental workload.

3.2 Final conclusion

In brief, **Study I** indicated that the presentation of information on-screen vis-à-vis in printed form produced a detrimental effect on human information processing and **Study II** showed that some of that detrimental effect may be attributed to the differences in the navigational properties of the two media. In addition, the result from **Study II** demonstrates that, by adapting the page layout of the presented material to fit its intended medium, mental workload may be alleviated. Taken together, both **Study I** and **Study II** show that the presentation of information affects information processing and thus highlighting the importance of further human computer interaction research in this area. The results of **Study III** showed that, in order to understand the memory demands of complex cognition, it is important to include elements of both the ST- and LT-WM paradigm of Ericsson & Kintsch (1995) and the tripartite model of Baddeley & Hitch (1974).

From a theoretical point of view, the results show that, in order to understand the memory demands of complex cognition, it is important to include elements of both the ST- and LT-WM paradigm of Ericsson & Kintsch and the tripartite model of Baddeley & Hitch. This notion is based on the findings that reading comprehension and text production are similar in the sense that they both depend on LT-WM demands but that they differ in that, of the two, only text production is susceptible to an additional load on the visuo-spatial sketchpad. In the same fashion, syllogistic reasoning and the condensation of a short text are similar in that they both rely on ST-WM processes, whereas only syllogistic Reasoning is susceptible to an additional load on the phonological loop. Thus, future research investigating complex cognition and human information processing ought to expand current views on the structure of memory resources (see figure 4).

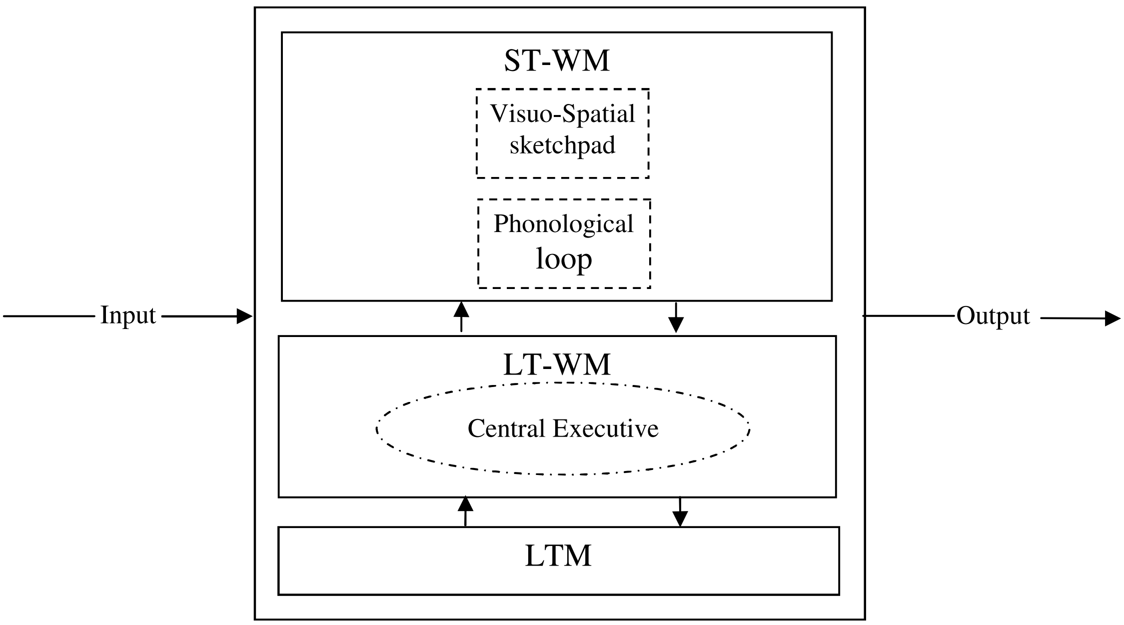

Figure 4: A proposed model for complex cognition that includes both the LT-/ST-WM and the tripartite working memory models [modified from Baddeley and Hitch (1974) and Ericsson and Kintsch (1995)]. The explicit processes of the visuo-spatial sketchpad and phonological loop are placed within the ST-WM whereas the implicit processes of the central executive are placed within the LT-WM.

From a methodological point of view, the most important aspect concerns the measurement of information processing power and mental workload using dual-task experiments. As the measurement of processing ability is, in essence, a measure of maximum capacity, the manipulation has to add sufficient mental workload to the original task so that task demands exceed the available processing resources. However, using secondary task reaction time data, it is possible to measure the available resources. Thus, when the primary task is manipulated, but not to the extent that it exceeds available resources, the effects of the manipulation on mental workload may be measured through the secondary task. A final methodological note ought to be made regarding the appropriateness of conducting dual-task studies of human information processing using distributed web experimenting. Not only is it possible to design experiments of high ecological validity, as tasks and graphics may be designed to resemble those usually encountered by participants, the possibility of group testing and easy data acquisition also makes it a very appealing alternative.

From a practical point of view, the results demonstrate two aspects of complex cognition. Firstly, comparing the effects of the mode of presentation can give us clues regarding what influences mental workload. It is, however, pointless to do this in order to advocate the abolishment of digitally presented material. Nevertheless, such findings are useful starting points in research concerning how to minimize mental workload during human-computer interaction. Secondly, and more importantly, in order to optimize the information processing ability during consumption of information of onscreen presented material, it is time to abandon the high-tech recycling of the ancient scroll and to break free of the traditions of printing, i.e. displaying written material in the standard A4 format and instead adapt the page layout to fit its intended medium.

4 References

Amabile, T. M. (1996). *Creativity in Context*. Oxford: Westview press.
Baddeley, A. D. (1996). Exploring the Central Executive. *Quarterly Journal of Experimental Psychology: Section A, 49*(1), 5-28.
Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), *Recent advances in learning and motivation* (Vol. 8, pp. 47–90). New York: Academic Press.
Belmore, S. M. (1985). Reading computer-presented text. *Bulletin of the Psychonomic Society, 23*(12-14).
Brunken, R., Plass, J. L., & Leutner, D. (2003). Direct Measurement of Cognitive Load in Multimedia Learning., *Educational Psychologist* (Vol. 38, pp. 53-61): Lawrence Erlbaum Associates.
Brunken, R., Steinbacher, S., Plass, J. L., & Leutner, D. (2002). Assessment of Cognitive Load in Multimedia Learning Using Dual-Task Methodology. *Experimental Psychology, 49*(2), 109-119.
Cantor, J., & Engle, R. W. (1993). Working-Memory Capacity as Long-Term Memory Activation: An Individual-Differences Approach. *Journal of Experimental Psychology: Learning, Memory and Cognition, 19*(5), 1101-1114.
Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450–466.
Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3*(4), 422-433.
Duff, S. C., & Logie, R. H. (2001). Processing and storage in working memory span., *Quarterly Journal of Experimental Psychology: Section A* (Vol. 54, pp. 31-48): Psychology Press (T&F).
Ekvall, G. (1969). *Rubriker: Användarmanual [Headlines: user manual]*. Malmö: Psykologiförlaget [Psychology Publishers].
Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*, 211-245.
Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence and maximal adaptation to task constraints, *Annual Review of Psychology* (Vol. 47, pp. 273): Annual Reviews Inc.
Ghiselin, B. (1963). Ultimate Criteria for Two Levels of Creativity. In F. Barron (Ed.), *Scientific Creativity*. New York: Whiley and Son.
Gilhooly, K. J., Logie, R. H., Wetherick, N. E., & Wynn, V. (1993). Working memory and strategies in syllogistic-reasoning tasks. *Memory & Cognition, 21*(1), 115-124.
Glanzer, D., Dorfman, D., & Kaplan, B. (1981). Short-term strage in the processing of text. *Journal of Verbal Learning and Verbal Behavior, 20*, 656-670.
Gray, S. H. (1991). Information search with dynamic text vs. paper text: an emperical comparison. *International Journal of Man-machine Studies, 35*(575-586).
Guilford, J. P. (1967). *The nature of human intelligence*. New York: McGraw-Hill.

Hart, S. G., & Staveland, L. (1988). Development of the NASA task load index (TLX): Results of empirical and theoretical research. In N. Meshkati (Ed.), *Human mental workload* (pp. 39–183). Amsterdam: North-Holland.
Holmquist, R. (1974). *Testbeskrivning Syllogismer I och II [Manual to the Syllogismer I and II Test]*. Malmö: Meddelande fron Metodoteket.
James, W. (1890). *The Principles of Psychology*. New York: Dover.
Jolicoeur, P., & Dell, A. R. (1999). Attentional and structural constraints on visual encoding. *Psychological research/ Psychologische forschung*(62), 154-164.
Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall.
Kintsch, W., Healy, A., Hegarty, M., Pennington, B., & Salthouse, T. (1999). Models of working memory: Eight questions and some general answers. In P. Shah (Ed.), *Models of working memory: Mechanisms of active maintenance and executive control*. New York: Cambridge University Press.
Kintsch, W., Patel, V. P., & Ericsson, K. A. (1999). The role of long-term working memory in text comprehension. *Psychologia, 42*, 186-198.
Kjellberg, A., & Iwanowski, S. (1989). *Stress/energiformuläret: Utveckling av en metod för skattning av sinnesstämning i arbetet. [The Stress/arousal questionnaire: Development of a method for assessment of mood at work]* (Undesrsökninsrapport [Research report] No. 1989:26). Solna: Sweden: Arbetsmiljöinstitutet [National Institute for Workplace Protection].
Koch, I., & Prinz, W. (2002). Process interference and code overlap in dual-task performance. *Journal-of-Experimental-Psychology: Human Perception and Performance*(28), 192-201.
Kolers, P. A., Duchnicky, R. L., & Ferguson, D. C. (1981). Eye movement measurement of CRT displays. *Human Factors*(23), 517-527.
Kumashiro, M. (1995). Practical measurement of psychophysiological functions for determining workloads. In E. N. Corlett (Ed.), *Evaluation of human work: A practical ergonomics methodology (2nd ed.)*. Philadelphia, PA: Taylor and Francis.
Lubart, T. I. (2001). Models of the creative process: Past, present and future. *Creativity Research Journal, 13*(3-4), 295-308.
Marcus, N., Cooper, M., & Sweller, J. (1996). Understanding Instructions. *Journal of Educational Psychology, 88*(1), 49-63´.
Markman, A. B. G., Dedre. (2001). THINKING. *Annual Review of Psychology, 52*(1), 223.
Martinsen, Ø., & Kaufmann, G. (1999). Cognitive Style and Creativity. In S. R. Pritzker (Ed.), *Encyclopedia of creativity*. San Diego: Academic Press.
Mayes, D. K., Sims, V. K., & Koonce, J. M. (2001). Comprehension and workload differences for VDT and paper-based reading. *International Journal of Industrial Ergonomics, 28*(6), 367-378.
Miller, G. A. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *The Psychological Review, 63*, 81-97.
Miyake, A., & Shah, P. (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. New York: Cambridge University Press.
Miyake, S. (2001). Multivariate workload evaluation combining physiological and subjective measures. *International Journal of Psychophysiology, 40*(3), 233-238.

Noyes, J. M., & Garland, K. J. (2003). VDT versus paper-based text: reply to Mayes, Sims and Koonce. *International Journal of Industrial Ergonomics, 31*(6), 411-423.
Olive, T. (2003). Working Memory in Writing:Empirical Evidence From the Dual-Task Technique. *European Psychologist*.
Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive Load Theory and Instructional Design: Recent Developments. *Educational Psychologist, 38*(1), 1-4.
Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive Load Measurement as a Means to Advance Cognitive Load Theory., *Educational Psychologist* (Vol. 38, pp. 63-71): Lawrence Erlbaum Associates.
Parasuraman, A. (2000). Technology Readiness Index [TRI] A Multiple-Item Scale to Measure Readiness To Embrace New Technologies. *Journal of Searvice Research, 4*(2), 307-320.
Partridge, D., & Rowe, J. (1994). *Computers and Creativity*. Wiltshire, England: Intellect Books.
Pashler, H. E. (1998). *The Psychology of Attention*. Cambridge, Mass.: MIT.
Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58*, 193-198.
Piolat, A., Olive, T., & Kellogg, R. T. (2005). Cognitive effort during note taking. *Applied Cognitive Psychology, 19*(3), 291-312.
Piolat, A., Roussey, J.-Y., & Thunin, O. (1997). Effects of screen presentation on text reading and revising. *International Journal of Human-Computer Studies, 47*(4), 565-589.
Reid, G. B., & Nygren, T. E. (1988). The Subjective Workload Assessment Technique: A Scaling Procedure for Measuring Mental Workload. In N. Meshkati (Ed.), *Human Mental Workload* (pp. 185-218). Amsterdam: North-Holland.
Reips, U.-D. (2002). Standards for Internet-based experimenting. *Experimental Psychology, 49*(4), 243-256.
Rice, G. (1994). Examining constructs in reading comprehension using two presentation modes: paper vs. computer. *Journal of Educational Computing Research, 11*, 153-178.
Sandlund, E., Linnarud, M., & Norlander, T. (2001). Effects of Stress versus Flotation-REST Relaxation on Creativity and Literacy Skills in Advanced English as a Second Language (ESL) Composition. *RASK:, 15*, 95-113.
Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans., *Psychological Review* (Vol. 99, pp. 195).
Sternberg, S. (2001). Separate modifiability, mental modules, and the use of pure and composite measures to reveal them. *Acta Psychologica, 106*(1-2), 147-246.
Tulving, E. (1999). Study of Memory: processes and systems. In M. Jelicic (Ed.), *Memory: systems, process, or function?* Oxford: Oxford University Press.
Waern, Y. (1989). *Cognitive aspects of computer supported tasks*. Chicester: Wiley.
van der Linden, D., Frese, M., & Meijman, T. F. (2003). Mental fatigue and the control of cognitive processes: effects on perseveration and planning. *Acta Psychologica, 113*(1), 45-65.
Van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive Load Theory and Complex Learning: Recent Developments and Future Directions. *Educational Psychology Review, 17*(2), 147-177.

Van Orden, K. F., Limbert, W., Makeig, S., & Jung, T. P. (2001). Eye activity correlates of workload during a visuospatial memory task. *Human Factors, 43*(1), 111-121.

Waters, G. S., & Caplan, D. (2004). Verbal working memory and on-line syntactic processing: Evidence from self-paced listening, *Quarterly Journal of Experimental Psychology: Section A* (Vol. 57, pp. 129-163): Psychology Press (T&F).

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*, 1063-1070.

Verwey, W. B., & Veltman, H. A. (1996). Detecting Short Periods of Elevated Workload: A Comparison of Nine Workload Assessment Techniques. *Journal of Experimental Psychology: Applied, 2*(3), 270-285.

Wickens, C. D. (1984). *Processing resources in attention*. New York: Academic.

Wickens, C. D., Mavor, A. S., & McGee, J. P. (Eds.). (1997). *FLIGHT TO THE FUTURE: HUMAN FACTORS IN AIR TRAFFIC CONTROL*. Washington, D.C.: NATIONAL ACADEMY PRESS.

Wierwille, W. W., & Eggemeier, F. T. (1993). Recommendations for mental workload measurement in a test and evaluation environment. *Human Factors, 35*, 263–281.

Xie, B., & Salvendy, G. (2000). Review and reappraisal of modelling and predicting mental workload in single- and multi-task environments., *Work & Stress* (Vol. 14, pp. 74-99): Taylor & Francis Ltd.

Yates, F. M. (1966). *The Art of Memory*. London: Pimlico.

Ziefle, M. (1998). Effects of display resolution on visual performance. *Human Factors, 40*(4), 554-568.

5 Appendix

Study I

Wastlund, E., Reinikka, H., Norlander, T., & Archer, T. (2005). Effects of VDT and paper presentation on consumption and production of information: Psychological and physiological factors. Computers in Human Behavior, 21, 377-394.

Study II

Wastlund, E., Norlander, T., & Archer, T. (Submitted). The Effect of Page Layout on Mental Workload: A Dual-Task Experiment.

Study III

Wastlund, E. & Archer, T. (Submitted). Working Memory Loads derived from Computer-based Primary- and Secondary-Tasks