ELABORATION OF TEXT FOR SKILL ACQUISITION

by

JONATHAN JAMES STUART, B.S.c, M.Sc.

A Doctoral Thesis
Submitted in partial fulfilment of the requirements
for the award of Doctor of Philosophy of the
Loughborough University of Technology

September, 1991

Supervisor: Dr A. Shepherd
Department of Human Sciences

ACKNOWLEDGEMENTS

I wish to express my gratitude and appreciation to my supervisor, Dr. A. Shepherd, for his encouragement, constructive criticism and friendship. I would also like to thank my friends and colleagues in the Department of Human Sciences for their support; particularly the research students in CERG.

The donation of many hours of time by the Loughborough Students who took part in experiments has been priceless. I am also very grateful for the financial assistance provided through the efforts of the Welfare Office.

I dedicate this thesis to my family, without whose guidance, sense of perspective and love I would not have been able to complete this work.

"He was turning the method of rationality against itself, turning it against its own kind, in defense of an irrational concept, an undefined entity called Quality.

He wrote: "(1) Every instructor of English composition knows what quality is. (Any instructor who does not should keep this fact carefully concealed, for this would certainly constitute proof of incompetence.) (2) Any instructor who thinks quality of writing can and should be defined before teaching it can and should go ahead and define it. (3) All those who feel that quality of writing should be taught anyway, can benefit by the following method of teaching pure quality in writing without defining it."

He then went ahead and described some of the methods of comparison that had evolved in the classroom."

Robert M. Pirsig 1974
Zen and the Art of Motorcycle Maintenance
p 208

There is a Quality to the prose of instructional text, but there are also means of measuring its effectiveness. This is a science.

"The concept of information design , though, challenges us to concentrate on the development of planning procedures and to relegate visual expression to a subsiduary role. ... Planning for communication in this extended sense is no less creative in nature than graphic art; the process of planning still relies on intuitive problem solving skills. But the most skilled problem-solver is limited by his knowledge-base.
... If there is no genuine foundation for the process of systematic enquiry we are engaging not in science, but in what Liam Hudson (1972) called 'the cult of the fact'.

Robert Waller (The Open University)
Functional information design : research and practice
Information design journal vol 1, 43-50, 1979

THESIS SYNOPSIS

The thesis investigates the elaboration of text to improve the acquisition of computing skill by casual users. Manuals are currently the main method by which these users are helped to acquire such skills. However, since there is no control over the way that the manual will be made use of, and no control over the sequencing of task learning, manuals are rarely entirely satisfactory in supporting learning processes. The issue investigated in the present thesis is whether manuals can be modified by the elaboration of text in order to support learning. The thesis also investigates skill measurement techniques, methods of specifying training devices and the applicability of controlled laboratory experiments to this applied setting.

A comparison of computing skill measurement techniques showed that the most valid and practical method consisted of short tasks to be carried out using the computing system. These tasks were used as the basis for the study of the effectiveness of manuals.

Examination of the pertinent literature in the light of the problems of producing texts for casual users and a model of skill acquisition based on hierarchical task analysis produced several theoretical devices for improving manuals. A formal method of applying the devices to a standard text enabled the controlled examination of the theories.

Pilot studies of the usefulness of the texts were carried out. These uncovered unexpected features of both the experimental setting and the embodiment of the elaboration techniques in electronic manuals.

The effect of the appearance of the texts was examined in a study that compared the effect of graphic design techniques verses one elaboration method. Although there was a significant initial affect of the elaboration on time to complete the tasks, the effect of the graphical design and later affect of the elaboration on unaided performance was insignificant.

A further, more controlled study of several elaborations showed that the task space allowed the subjects to follow several strategies. This applied problem area cannot usefully be studied using a factorial type design without removing the opportunities for idiosyncratic approaches to the task that characterise the area. An examination of individual subject records proved to be a more useful paradigm. Distinct types of behaviour arising from the nature of the different materials were observed. Implications for further research in this area were discussed.

The thesis has established that the category of users described in the literature as 'casual users' exists and that they exhibit the behaviours identified in anecdotal reports.

It has brought a rigour in the design of texts, tests and experimental method to this area of study and developed techniques that can be used by other workers for further research.

The studies have provided original descriptions of the behaviour of subject's use of manuals for this type of task.

The results and discussion of both the main study and study of the effects of graphic design of texts on learning provide a strong argument for the need of a designer of texts and tasks rather than the slavish application of human factors research.

CONTENTS

CHAPTER 1

REQUIREMENT FOR AN IMPROVED METHOD OF ELABORATING TEXT FOR SKILL ACQUISITION

The inadequacies and cost of current training methods for the casual user have been well documented. This thesis examines the problems of training the casual user and evaluates the use of elaborated text as a solution.

Since the introduction of computer systems that are inexpensive when compared to the cost of providing individual tuition, manufacturers have moved towards providing training in the form of manuals (Uhler, 1981; Uhler, 1984). The current forms of manual provided with systems are disliked by users and described as inappropriate for their requirements (Gilb & Weinberg, 1977; Scharer 1983; Carroll & Carrithers, 1984)

Although manuals are the cheapest part of a computer system they can strongly determine its cost effectiveness (Eames & Starr, 1965; Dobson, 1991). Sophisticated computer systems are only of value if the user is properly trained and/ or the instruction manual is well written (Sullivan & Chapanis, 1983). Changes in the instructional material can have a large effect on the efficiency of the man-machine system (Chapanis, 1965; Broadbent, 1978).

Improvements have been made to the production of manuals (Farkas & Williams, 1990) but the main concern of technical authors is still to ensure that the documentation is up to date and complete. This is an essential component of a manual but it does not address the learning problems of the computer user, just the minimum information requirements of the manual.

(Eason, 1979) stated that system designers traditionally assumed that users have:

1. The ability to communicate in computer terms.
2. Training to teach them to communicate in computer terms.
3. Willingness to communicate
4. The opportunity to take part in the design of the system.

Traditional machine training also makes the assumption of regular, competent users whereas many computer users may carry out tasks intermittently and will not always be competent at carrying out those tasks. In addition computer systems change and evolve while most machines are relatively static (Mayer, 1967) - the information worker is required to develop new transformations of information as part of their job (Schuck, 1985)

In the design of highly specific computer tools that are the main part of a person's job it is usual for all of these assumptions to be true. These types of jobs are now a decreasing proportion of the total tasks undertaken with computers; it is increasingly common for computers to be used by 'the man in the street' as part of his job or daily activities.

Training solutions to resolve the conflict between design assumptions, the cost of training and the needs of the casual user are required.

Concerning general education in response to the increasing demand for information, increasing population, rapidly increasing body of knowledge and lack of teaching resources (Bruner, 1966) proposes that education should aim to help people to :

1.Adapt readily to change.
2.Be able to abstract the underlying principles of subjects.

The training requirements of the casual user are a subset of these general educational needs, determined by the nature of the computing task and the relationship between the person's job tasks and the task of learning to use the computer system. The literature review of this area is described in Chapter 2. In it the characteristics of the casual user in terms of their use of computers and their approach to learning computer application packages are described.

The use of computers does not dominate the work of casual users; computers are used as a means to an end. Casual users are not prepared to spend a large amount of time using the system or studying instruction manuals for the information that they require. Instead, after a honeymoon period of learning, they make do with the set of commands they they have learnt, make assumptions about the rest of the system and may forget how to perform tasks because of poor instruction or lack of practice.

These users require longer retention of the skills that they originally learnt, increased performance at carrying out similar work tasks and an increased ability to solve new problems.

Previous studies of means of improving the training of casual users were criticised on a number of points. The studies frequently studied one shot learning over a short period of time; this is quite different from the actual learning situation of casual users. Many of the studies did not carefully control the construction of the training devices or base their measurement of skill on tasks that were tested for their validity or derived from a model of the skill.

A means of measuring the skill of the users at carrying out computing tasks is required in order to test the efficacy of different training approaches. In Chapter 3 several tests of skill and means of administering the tests are examined in relation to a test of skill based on the users actual work tasks. The aim of the experiment was to find the method that most closely correlated with the results of the work task while requiring the least use of time, marking skill and resources.

While the study revealed that the use of short test questions based on editing tasks was the most useful technique for skill measurement it also revealed that the substance of the tests, the actual work tasks to be performed, needed further examination in order to produce a test battery that could reflect the different aspects of skill. A model of skill is developed, in Chapter 4, that provided a basis for the development of training solutions to the problems of casual users and a means of selecting tasks for the measurement of skill. Skill is operationally defined in order to overcome the theoretical problem of distinguishing between transfer and problem solving and to describe the generic form of tests of skill.

The factors to be manipulated in order to improve skill are external (or directly observable) variables that lead to learning - the instructional materials and procedures and the student's learning behaviour. The external variables that result from learning are the basis for the measures of performance. The internal or cognitive variables are discussed in the thesis as a means of selecting practical methods of improving learning. The literature relevant to the aspects of the model; studies of memory, reading, in relation to text processing for skill acquisition is used pragmatically in order to solve this applied problem. Chapter 5 discusses this literature in relation to the training requirements of casual users with a particular emphasis on the use of text for learning a skill. These studies match the use of text by computer users; many previous studies are not appropriate for this applied problem because they have examined reading a text for recall. It was concluded that the most appropriate solution was the controlled elaboration of text using devices such as the provision of goals, examples of procedures, explanations of procedures, a means of increasing the effort the subjects expend on the learning of the task. For comparison both a device for producing a standard text designed to approximate current manuals and a core text that excluded all elaborations were also selected for further study.

The means of producing the texts is described in Chapter 6. Previous studies of elaboration have produced texts according to very loose rules, often incorporating a number of devices at once and describing the production of the texts in general terms only, if at all. A method of producing texts based on a description of the tasks that incorporate the methods selected in the previous chapter is described. This consists of the hierarchical task analysis of the task and the addition of material corresponding to the training hypotheses.

The texts are incorporated into an electronic manual in order to study their use and the performance of subjects. Chapter 7 describes this process, the selection of tasks for tests of skill and the design of studies of the effect of the texts on skill acquisition. Several pilot studies were carried out in order to exclude problems in the design and materials. It was found that the means of increasing subject's effort could not be practically implemented; that the description of the task, errors and help given to subjects needed be formalised and that the presentation of materials for the tests of the retention of skill and transfer needed to be balanced across all conditions.

Several subjects complained about the graphical layout of the texts; graphic designers confirmed that there were means of improving the layout. Chapter 8 describes a test of the effect of graphic design on skill acquisition. The performance of subjects using both graphically enhanced and plain forms of elaborated and unelaborated texts was measured with the texts available for use and afterwards with the texts withheld. It was found that although good graphical design improves initial speed at answering questions this effect is soon lost. The elaborations led to slower initial performance but enhanced performance on the recall tasks. It was concluded that the plain graphic design of the electronic manuals would not affect skill acquisition and so the materials could be used as initially designed.

A study of five forms of elaboration carried out according to the improved design, described in Chapter 9, revealed that the task space under study allowed the subjects to adopt many more strategies than was originally anticipated. Critical incidents during learning affected the performance of the subjects more that any overall effect of the elaboration. This led to a wide range of scores within each group - preventing a comparison with other groups. It appears that a factorial paradigm for the study of this applied area is not appropriate; instead an examination of the pattern of critical incidents that lead to skill acquisition is required.

The effect of one particular critical incident, that of accidentally finding the information required in later tasks did not improve performance as expected. Instead the incident was a result of the subject's strategy for using the manual which led to slower performance with heavier use of the manual. It is suggested that a more sophisticated model of manual use for skill acquisition is required along with a means of studying the patterns in the behaviour of subjects in order to continue the study of this applied problem.
The strengths and weakness of the techniques developed in this thesis are described in the discussion. The carefully defined method of manual production, though leading to an unusual appearance, is a necessary tool for the study of text elaboration - without such control any findings are open to debate. The incorporation of many of the features of this complicated applied area has led to studies that revealed a large number of useful qualitative results but has caused problems in the examination of the quantitative results because of the range of behaviours allowed. Forms of experimentation that can chart the effects of training methods in this area without resort to the exclusion of the factors that characterise the problem need to be developed.

CHAPTER 2

LITERATURE REVIEW AND DESCRIPTION OF THE TRAINING NEEDS OF CASUAL USERS.

Introduction

The aim of this thesis is to reveal the issues and problems surrounding the task of training computer users, to articulate a framework for producing a practical solution and to test the adopted approach using both laboratory and field experiments.

There are four main topics to be covered in the description of the problem of providing training for casual computer users: a description of the characteristics of this particular user population; the nature of the task that they are to perform; their current learning behaviours and the training that they currently receive.

From these descriptions the main training issues will be selected and a method for arriving at solutions to the problems will be set out. The rest of the chapter will describe the means of arriving at that end result. First of all a definition of skill and an outline of the theory of skill used in this thesis will be established along with a section on the measurement of skill. Once this foundation has been laid the training strategy will be presented. The training strategy will provide the user with a text containing a basic description of the constituent sub-tasks of the task (a core text lifted from the HTA) which will be embellished with extra text designed to improve their performance.

Characteristics of the casual user

Accounts in the literature have emphasised different characteristics of the casual user (Cuff, 1980; Allwood, 1986). The main points that are consistent throughout accounts and which are pertinent to this study have been selected. Casual users comprise of people with a wide variety of computing experience and job needs; reviews of particular sections of the group have emphasised different qualities. This review will concentrate on the aspects of the people that are pertinent to training provision.

Casual users employ computers infrequently (Martin, 1973), spending the rest of their day on other aspects of their work. The use of the computer is not necessarily forced by their job or social role (Codd, 1974). They are more dedicated to the performance of their work rather than the use of the computer. Although the computer may be necessary for their job they try to minimize any time spent using the machine that is not leading to completion of their work. This includes time spent learning computing skills (Eason, 1982); partly as a consequence they undergo little or no initial training ((Barrett, Thornton, Cabe, 1968; Damodaran, 1974). They are usually highly skilled in their work tasks but have little technical knowledge of computers; they have mental models of computing tasks that are closer to their work task than to the system (Kennedy, 1975). Their computing skill tends to asymptote at relative mediocrity (Rosson, 1983; Neilsen, Mack, Bergendorff & Grischkowsky, 1986), though they may have 'pockets' of computing expertise (Shapiro and Kwasny 1975). They have a poor perception of the potential of their computing systems to help them with their work.

Damodaran (1974) describes the relationship between job type and training need. She found that specialist users required more knowledge of the facilities available and how to control them; clerks wished to know more about the context of their computer-task within the computer system and company organization while managers wanted quick specific training (due to the short amount of time they were prepared to spend on learning to use the system). This illustrates that the training solution must either be job specific or capable of being altered by the user to meet their own needs.

Casual users prefer to expend little effort on system use rather than on their work - Eason & Damodaran (1974), describe this as the tendency for users to "satisfice". This is similar to the psychological phenomenon of conceptual set, where subjects continue to use known problem solving strategies, even when simpler strategies would have worked. It is this, according to Eason, which is an illustration of Zipf's (1965), "Principle Of Least Effort". However, they may not have the opportunity to avoid using the computer (Eason 1982). The users prefer to make a reasonable guess rather than approach tutorial medium directly (Cuff 1980), again in order to minimise the effort expended on using the system.

Their limited knowledge means that naive users are "at risk" from error which may colour their future attitudes towards the system. (Eason 1982).

These are the basic characteristics that describe the work situation of the population known as casual users, as described in the literature. The label 'casual user' fits a particular profile rather than any particular user. In order to be more specific about the people discussed in this study some more details of their tasks needs to be provided. The users in this study are those that use application programs rather than programming languages. With the advent of 4th Generation programming languages the distinction between the two types of tool is blurred; the shared features and issues in these and other areas of skill will be discussed later.

The term 'intermittent user' would perhaps be more suitable for some of the members of the population discussed in this study in that they only use the computer for a relatively small proportion of their time at work. However some of the users may utilise the computer for most of their day and still be classed as casual users because they use parts of the system only occasionally. It is therefore not the degree of computer usage that is important, it is the range and frequency of computing tasks that they carry out. Someone who carries out the same task over and over again with relatively short gaps between the sessions is clearly not a casual user. One that spends the same amount of time at the terminal but carries out a range of computing tasks, each of which is not carried out frequently and perhaps never in the same manner, bears a closer resemblance to the profile. The essential point is that the user has not reached the degree of expertise in carrying out his work on the computer that it is effortless and without error; tasks need to be remembered, new methods need to be explored in order to carry out his job. The terms 'casual user' and 'user' will be used interchangeably to describe the users in this study.

Problems of the task

The problems of the task for the casual user are not just aspects of usability resulting from a poorly defined interface. Using a computer application package is essentially a hard task (Mack, Clayton & Carroll, 1983) that is made more difficult by the lack of formal trainng and practice available to the users.

Eason (1979) proposes that the type of help required depends upon the frequency that the task is carried out and on the complexity of the task. Again, a selection of levels of help should be available

The user's primary goal is to use the computer to carry out his work task; the profile of these determines the pattern of computer use and the system operations carried out. This consists of infrequent use of the tool's functions, a likelihood that system tasks will need to be combined in a novel fashion in order to perform new work tasks and a strong possibility that new computer applications will need to be learnt from time to time.

The users are experts at their work tasks. Their problem is to determine how to combine system operations in order to carry out their familiar work. To use Norman's (1986) phrasing they need to form task to system bridges of understanding in order to be able to perform.

It is the range of system operations available to the user and the possibility for combining these within the constraints of syntax that make computer application programs particularly difficult to master. It is not just a matter of learning a short sequence of operations, whose combinations and effects can be learnt by rote; there is always a wide selection of actions to be coordinated according to the user's needs. Once the system's operations have been memorised the user's learning task has not been completed, he must learn to orchestrate them to use the tool to its full potential.

The problem is sometimes further complicated by the user's need to use more than just one system. If the systems have similar functions, for example both of them are word processors, they may both operate according to the same basic principles and only details vary. Transfer of learning will occur at the conceptual level but not at the syntax level E.g. Singley & Anderson, 1985. This is not always the case - compare screen and line editors. Even if transfer occurs at the conceptual level learning will not necessarily be faster than if the similarities were at the syntax level - some minor differences in syntax can be harder to remember and lead to more errors than conceptual differences can. Once the user has mastered one package learning a new package is not necessarily straightforward, although some transfer will occur

Good software design cannot eliminate all of the training problems. With ever increasing computing power combined with progressive advances in the ability of designers to use this power, applications packages will contain more functions covering more work tasks. Advances in software design ought to eliminate some of the other problems that users encounter when learning to use a new system. These are primarily the sparsity of the interface - frequently packages provide few cues to help the user understand the current state of the system (leading to a feeling of disorientation) and poor feedback about the effect of the user's actions (leading to a lack of reinforcement and lack of confidence), and a lack of consistency in the interface (giving the impression of elusiveness and slipperiness). Poor design can compound these flaws with incomprehensible error messages couched in system terms that provide little information about the cause of the error (Carroll, 1982; Mack, Lewis and Carroll, 1983).

Errors in software package usage can be of two types. The slip may be contrary to the correct operation of the machine, in which case the system should not comply and should inform the user. In 'fool proof' systems such errors are harmless; in more fallible systems such errors may cause the computer to 'crash'. The second type of error is to correctly execute the operations in the package in a manner that does not match the intentions of the user. If noticed the action may either be easily reversible or may throw the system into a different state requiring action by the user in order to get back to his original task. Unnoticed errors lead to even more problematic states where recovery will not consist of a single action. Knock on effects of errors make learning a difficult and frustrating process for users; one slip can take the user from a well known action sequence into an alien area of the system (Mack, Lewis and Carroll, 1983).

Current behavioural profile of the user

Casual users tend to forget major and minor details of system use (Cuff, 1980). It is not only through mistakes that users enter unknown features and states of the system. It is their normal tendency to explore tool features without adequate knowledge or after reference to instructions. According to Carroll (1982) there are several causes of this behaviour.

Firstly, users can be frustrated by the sequence of tuition or information provision in manuals. As stated above, a major characteristic of these users is that they have a production bias - they want to use the tool to produce work, not just for the sake of using the tool, and they want to spend minimal amount of time using the tool in a non productive manner. This includes the time spent following the instruction manual in order to learn the system according to the sequence of reading prescribed by the author. They would rather explore the system themselves in order to perform their tasks, taking the risk of making mistakes because they feel that this is a more time efficient way of learning. As a learning strategy self exploration can be effective - learning through trial and error the intelligent user can acquire a rich and vivid memory for the use of operations, providing the sequence of learning experiences are fortuitous. This method of learning will be discussed later.

People do not engage in trial and error learning in a random manner unless the task to be learnt provides few cues about how to perform it. As natural pattern matchers people are quick to make generalisations from observation and apply analogies to current situations. While acquiring computing skills this often means making incorrect generalisations from the syntax of previously learnt commands (Mack, Lewis and Carroll, 1983; Carroll & Rosson, 1986) or applying analogies from more familiar office tools, such as the typewriter or filing cabinet (Allwood & Wikstrom, 1986). The terms generalisation and analogy are both subsumed by the term mental model, of which Carroll & Olson (1988) say '... the models that learners spontaneously form are incomplete, inconsistent, unstable in time, over simple, and often rife with superstition'. Another behaviour that arises from user's desire to minimise time spent learning system tasks is what Carroll & Carrithers (1986) calls an assimilation bias.

Not only would they prefer to spend time using the system 'productively', they would also prefer not to learn anything that they aren't forced to. Other authors describe this as satisficing. It is characterised by users making the best of the commands that they already know rather than learning other commands that may be more efficient to use but require time to be spent learning to use them.

Users who show a comprehensive knowledge of these commands, which may be high level macros or infrequently required shortcuts, are called 'power users'. For instance it is possible to multiply numbers on a calculator by adding the same number over and over again instead of learning to use the multiplication button. An example from computing can be seen in the Unix environment where power users often construct sophisticated shells tailored to their own work needs and use a wide range of Unix and 'C' commands to bend the system to their requirements. Satisficers in this domain do not make full use of the system functions and frequently use repetitive sequences of key strokes that power users would either construct to form a simple command or use one existing command to save time and effort. The commonest explanation for satisficing is that these users are unwilling to expend the time and effort to learn the more powerful functions because they believe that they at least break even by continuing with their existing knowledge while making a few extra key strokes (See Eason, 1982). In some situations this is a correct and useful judgment. In others such perseverance is wasteful yet users continue, possibly because they do not know that other methods exist or because they dislike the process of learning with its inherent ups and downs.

The fear of making mistakes can be very limiting. Carroll (1982) draws a comparison with the attitude of computer game players, especially adventure game players, for whom mistakes and errors are challenges to be overcome and sources of new information. The difference between the two groups can be largely attributed to the different motivations of the users - workers lose time and perhaps work as a consequence of an error, game players may at worst take a knock to their pride. But another underlying cause is a fear of failure associated with trying to master what many believe to be a simple tool and a fear of the computer itself. Although this thesis will not deal specifically with the fears of users it is hoped that the improved performance and skill acquisition discussed here will reduce the possibility of 'failure'. Schuck (1985) expresses a need for an improved pedagogy within information technology work places in order to favour skill development through both the acceptance of mistakes as part of the learning process and the encouragement of questioning.

Questions such as 'How can I do this more efficiently ?' and 'Can I use the computer to perform this task ?' raised by the user are at the root of learning activities. A study by the author of question asking by users and a similar study by Briggs (1990) has shown that although all users are skilled at asking how they can perform particular work tasks it is only the most experienced of users who are able to ask questions about system operations. This finding ties in with the assertion that casual users have mental models that are closer to that of the task than that of the system (Kennedy, 1975). With a person at hand to answer their questions about carrying out work on a computer this deficiency would not be a hindrance. The problem lies in the way that training materials are currently constructed. They aim to describe the system and the effects of the commands rather than how to use the commands to good effect at work; inexperienced users are unable to look up their question in a manual and read the answer. This is the task to system bridge of Norman (1986) mentioned earlier; users need to learn how to use commands to perform tasks rather than just learn how to operate commands.

The way that users use manuals varies from person to person and according to their expertise and individual approach to learning. The variety of manual usage has not been documented, only the amount of use, for that reason the profiles given below are based on the authors experience and is designed to illustrate the range of learning behaviours shown by users rather than to draw a specific point about their needs and is not intended to restrict this discussion to the use of manuals for training. Some users prefer never to use a manual but to always rely on others for information and advice or, in contrast, never use a manual because they wish to explore the system for themselves. A few steadily work their way through the manual from cover to cover; most just dip in here and there wherever they feel is appropriate. More experienced users tend to use the manual only to find the answers to specific questions about the package while exploring the system or use them first of all in order to make comparisons with other systems that they have mastered. This is not to say that these users explore and learn in an efficient manner, some assume that they can solve any problems they may produce by ill considered exploration but find that they are unable to resolve the mess they have created. Manuals are not the only form of existing training but they are the most common device supplied with the system by the manufacturers to aid the use of their product (Uhler, 1981). The training available to users is discussed below.

However, Eason (1984) also points out that for the intermittent user, evidence exists to show that there is a "honeymoon" period in the first stages of use, in which the user is prepared to devote time to trial and error in order to become acquainted with the system. This is then replaced by a phase in which task performance dominates the user's thinking. Maximum utilisation of the latter period should therefore be made when helping users to master computer applications. Eason and Damodoran (1974) found that the extent to which a person likes/dislikes a system is likely to influence their motivation to learn about that system, and so positive reinforcement (intrinsic or extrinsic) in the initial stages of use is important. Also, learning to use computers would seem to be influenced by user's initial perception of them (Kennedy, 1975). The learners may be left "scarred and scared" after a bad first experience, and so preparation for computer interaction is important.

Current available training

The main problem that users encounter is how to get at the information that they require. More often than not they don't know what information, in computing terms, they want. i.e. they don't know exactly which question to ask. This may be why local experts are so popular, they transform the users ambiguous query into a narrow question then point the user to the appropriate location of the information (O'Malley, 1986).

Another form of assistance is the provision of 'sign-posts', such as headings and subheadings. These appear to provide a context which assists in the understanding process. Also, several experimental studies have suggested the benefits of outlining the main ideas in a section at the beginning, or summarising them at the end, as a useful aid in remembering, (Ausubel, 1960; Ausubel and Robinson, 1969).

According to Allwood and Wikstrom (1986) most users are not provided with any initial training. Damodaran (1974) states that 40% of users receive no formal training, 27% are provided with manuals and 58% receive on the job training; 67% wanted more training. Barrett et al (1968) when examining users satisfaction with a bibliographic information retrieval system is surprised to find that only 54% of users received formal instruction on using the tool and all would like more training. Eason (1982) comments on user's dissatisfaction with the level of training and user's continuing need for more training. The finding that manuals seem to be the most prevalent form of training is not surprising considering the market forces at work in the computing industry. As the cost of systems has come down users have not been able to afford or justify attending courses. Most of the new users, the casual, users could not justify taking time off work in order to learn a particular computing system because computing did not comprise a large proportion of their work. As a solution manufacturers provided manuals for the systems, at first aimed at systems managers or company trainers who could then train the end users, but subsequently aiming them at the end users themselves who more and more are the sole users of machines and responsible for their own training (Uhler 1981).

In environments where more than one person uses a particular package some will be more expert than others. Of those some will be sought out by the less experienced for aid. The term 'local expert' is applied to these informal trainers. This form of training, including on the job training by colleagues is one of the major forms of training for casual users (Eason 1976). Most users would prefer to ask a person for help than use a manual (Wright, 1979; O'Mally, 1986). Unfortunately such individuals are frequently unrewarded for this role and so the facility is a transient one of varying quality.

Advice centres, such as local computing services reception or experts who are 'on call', provide the same function as local experts but are more formally organised, the staff are usually trained to carry out this occupation and receive credit or payment for their support (Lang, Lang and Auld, 1981).

Both local experts and advisory staff have the choice of providing evolutionary or compensatory help (Damodaran, 1977). Compensatory help is the provision of sufficient information for the user to overcome his current difficulties. Evolutionary help is more forward looking and aims to prevent the user from having to ask the staff about this type of problem again. It may involve pointing out the manuals and information sources available, telling the user how to diagnose his problem, point out areas of misunderstanding and describing why a particular solution works in this case.

Schuck (1985) proposes that instead of the orthodox hierarchical management structure in which the worker is not encouraged to question or experiment with established practices a more open, supportive atmosphere should be promoted in which the worker is free to play with information so that its meaning can be understood. This perspective does not describe the area in sufficient detail to allow the construction of a training program, nor does it describe all of the differences between established assumptions behind training and the assumptions that hold in Information Technology training.

Damodaran (1977) similarly draws a distinction between minimum training that ensures that users can provide suitable input and obtain appropriate output for their task (compensatory training) and training that meets the users needs after they have mastered the essentials of a system and keeps them up-to-date with changes in the system (evolutionary training). She suggests that the criteria for effective support are that it is :

1. Evolutionary.
2. Formally recognised as a job role with appropriate career opportunities.
3. On demand, point of need.
4. The goals of support staff are congruent with the needs of the users.
5. Includes an adequate range of experience.

Eason (1979) stated that system designers traditionally assumed that users have:

1. The ability to communicate in computer terms.
2. Training to teach them to communicate in computer terms.

3. Willingness to communicate on the computer's terms.
4. The opportunity to take part in the design of the system.

In the design of highly specific computer tools that are the main part of a person's job it is usual for all of these assumptions to be true. These types of jobs are now a decreasing proportion of the total tasks undertaken with computers, it is increasingly common for computers to be used by 'the man in the street' as part of his job or daily activities. Training solutions to resolve the conflict between design assumptions and the needs of the user are required.

A major problem with documentation used to be that it was of a very poor quality (Chapanis 1965). It was frequently missing, incomplete, inaccurate or obsolete (Wright, 1984). In contrast good documentation is regarded by the users as complete, conducive to easy learning and has a logical arrangement of the material (Guillemette, 1987). Compared to the cost of hardware very little money was spent on its development (Eames and Stark, 1965). This is unfortunate because good manuals can save time, improve efficiency and save money (Alvari and Weiss, 1985). The same is true for any effective training system.

An examination of current documentation produced by major manufacturers shows that most have learnt to get the information content of documentation right first time, but this is not always the case. Documentation also used to be written by the system designers (Weber, 1985) and consequently was highly technical and system centred (Guillemette, 1987) but now manufacturers currently provide less technical manuals aimed at novices. As yet there do not appear to be many documents aimed at helping the user with their work rather than on how to use the system. Users require task to system knowledge bridges in order to make the tool useful in their work (Norman 1986).

Despite improvements in the quality of documents users have problems following instructions, though the instructions are frequently not read at all (Szhchinski, 1979; Wright, 1981). Poor documents are noticeable for their absence from the user's side; according to Wright (1984) good documents are rarely noticed.

Previous Studies of the area

One of the earliest discussions of the application of human factors to man-computer interaction provides a useful perspective:

> "One view is that ... research in this area is futile. Computer technology is moving so fast that knowledge obtained through experimentation is bound to be obsolete before it can be applied. Design improvements are discovered faster by trial and error than they could be by lab. research. ... if research in this area is to be worthwhile it must be directed towards the discovery of general principles which can then be translated into specifications of particular system design."
> (Nickerson, 1969 p503)

Human factors research in the design of systems and documentation has produced many independent results leading to lists of guide-lines and rules of thumb. This may be due to a degree of variation between non-manipulated aspects of the experiment e.g. two experiments examining the efficacy of two different manuals , one based on a windows package, the other on a command line package and both using different control conditions. or a concentration on aspects of the system rather than on the psychology of the user.

Many of the results have not been applied to less constrained, non-laboratory training situations and so no real idea is given of the usefulness of the findings alone or relative to other training methods.

A criticism that can be made of many studies in this area is that they have treated text as a static stimuli divorced from its function within the user's work context; they have required subjects to read text then recall items because they have volunteered for the experiment rather than make a text available for use in by subjects who wish to learn a particular task. A widely used index of a text's usefulness is the readability index (Taylor, 1953). Kern (1985) concludes that the subjects used for the creation of readability indexes are of a different population from the target population and hence are of dubious validity and that measures of readability at the work place do not correlate with the readability indexes, nor do they correlate with indexes of actual usage of the manuals.

Diehl and Mikaleky (1981) suggest that readability indexes are of little use for technical writing: adjusting the index does not make the text easier to use. They propose that the degree of usefulness depends on the purpose of the reading, the familiarity with the text and the importance of the information for the job. They classify the purpose of the reading into 'reading to do', 'reading to learn' and 'reading to assess'.

Kalt and Barrett (1973) attempted to combine access structures (highlighting in this case) with mathemagenic devices (advanced organisers, post questions, concrete illustrations and delayed review) in a text that was designed to be used by engineers for both learning the job and for reference. The engineers were given unlimited time in a classroom to study the text with the intent of learning the job from it. After a delay of one hour they were given six questions designed to test their knowledge of key ideas (provided by experts in the area), asked to complete some look up tests and give a subjective rating of the text.

They found that the experimental group retained more of the key ideas, scored better on the look up tests and preferred the text as well as rating it as easier to use for reference than a similar group of engineers using a plain text. These results were less pronounced with the more experienced engineers.

It is likely that asking the engineers to read the document as if they were going to learn from it has an effect very different from the behaviour of an engineer faced with a situation new to him that leads him to resort to the manual. The measure of retention (questions about main topics) seems very coarse with respect to the sort of knowledge that computer users will need to retain.

Wright and Reid (1973) carried out an experiment to assess both the ease of use of a manual and the amount of material retained after using the manual. The formats of the manuals were tables, algorithms and short sentences. They found that although tables were the most effective way of representing the information when the subject had access to the text, they led to the least retention of the material - performance recall was highest for short sentences. They proposed that these results were due to the difficulty of encoding tables as compared to the sentences.

Sicht et al (1972) and Caylor et al (1973) developed Job Reading Task Tests (JRTT) based on the concept of a job reading task by Kern

(1970) in order to better assess reading skill requirements of jobs. The job reading tasks were compiled from workers examples of the situations when they required information and the sources that they used to find the answers.

Kern (1985) proposes that instead of basing the evaluation of a manual on its fit with a set of guide-lines and the opinions of managers the process should include both 'dry runs' (target users carrying out job reading task tests) and job-site evaluations that compare the behaviour using a new manual with existing base-line data i.e. comprehensive feedback from the users.

Pugh (1977) found that skilled readers adopted different reading strategies in response to different types of questions about a text. These included scanning and search reading (answering 'How?' and 'What?' questions), skimming (deciding how to or whether to use a text), receptive reading ('What is the author saying ?') and responsive reading ('What do you think about what the author is saying ?'). It may be the case that different results for the same mathemagenic behaviours have been caused by the adoption of different reading strategies by the readers in response to different types of questions being presented by the experimenters. Walker (1979) also notes the need for text designers to start looking at the actual use of texts by users rather than at the results of standard reading tests.

An extension to the study of user's use of texts is to examine their information needs while they use a package. Allwood and Wikstrom (1986) asked four subjects to 'think aloud' while learning to use a word processing and a database system. They found that the commonest source of learning difficulties was due to poor knowledge of basic computer concepts (screen window, file) and that few errors were caused by the inappropriate use of analogies. However:

1. The authors point out that it is very difficult to attribute specific errors to particular analogies.

2. Only four subjects were used in the experiment, all of whom were very intelligent and very experienced computer users.

3. Even with training 'thinking aloud' is a very difficult skill which even when mastered may change the way that the user carries out the learning task.

A more ecologically valid method of studying user's information needs is the 'question asking' technique, as used by Kato (1986). Users are requested to use a system without documentation and are encouraged to use a tutor as a source of information. The tutor is trained to reveal only the necessary information and is overseen by the experimenter. Although the question asking activity is a far more natural behaviour for the user, the nature of the questions is almost certainly determined by the nature of the entity providing the answers (manual, computer or person). The method does indicate what problems the user encounters in which contexts, what information they require, which features are harder to learn and understand, how users arrive at understanding (to some extent) and in what ways users misunderstand the system. This is similar to the usability testing of a manual - although it reveals practical problems with the current version it does not provide a means of developing a more appropriate model of the user or approach to the design of documentation.

This criticism can also be applied to studies that survey user's needs. Sackman (1970) looked at whether interactive systems increased learning speed because of the immediate presentation of results by studying his own process of learning to use a system. Surveys of user opinions of advisory services have been carried out (Lang et al 1981, Dawson et al (1982) as the basis for improving the services. Such work gives an idea of what the users feel is lacking, but the users are not always the best judge. Rosson (1984) quantified computer experience among users then looked at the effect of their level of experience on learning, using and evaluating a text editor. The more experienced users tended to learn from manuals, less experienced (here secretaries) preferred to learn from friends and tended not to experiment with the system though they knew that such activity could be a valuable way of learning. The study does not investigate whether the secretaries could be helped to learn more effectively from manuals or why it is that experts are able to use manuals and the secretaries are not.

Myatt et al (1986) attempted to manipulate the form of training given to users to promote computer literacy. They attempted to find out if the skills learnt when mastering a computer application would aid people to learn other applications or computer languages and vice versa. Programming skills are transferable to other languages but application program skills are a relatively narrow set that lead to little transfer even to other applications.

In a similar manner Kennedy (1975) attempted to manipulate the medium that people learned about computing from (by giving them a manual or allowing them to practice on the system) and examined the effect of a person's attitudes towards computers on their learning about computers. Unfortunately none of the people in the manual condition bothered to read the manual. Attitude did not seem to make a difference, this is not surprising because the test used (a word association test) had not been tested for validity. He did find that those people who were more tense in the learning situation had trouble learning, but this does not distinguish between cause and effect.

Both of these studies have carried out very coarse manipulations of the training, with little control over the specific details or reference to an underlying model.

Cohill and Williges (1985) examined the retrieval of help information for novice users, Czaja et al (1986) looked at different forms of training: manual based, instructor or on-line help. The main problem with both of these studies is that they do not look at the micro-level of the training methods, just at the top level that they are manipulating. As a consequence one is unable to tell if one condition is better that the other because of the general training method or because of the approach/style of the method. For instance Caja et al (1986) found no difference between a manual based course and an instructor based course. After analysis of the results they realised that the instructor merely lectured to the learners, an approach similar to the manual. Their on-line condition was worse than the other two because, they feel, the computer did not allow the user to make mistakes and thus learn how to recover from errors. Cohill and Williges (1985) conclude that user initiated, hard copy help that allows for browsing is better that on-line help. They concede that had the program that initiated and selected the help been better, or a windowing scheme for the help been provided, the results may have been different.

A study that examined the problems of casual users and produced a method of overcoming some of their problems was carried out by Carroll and Carrithers (1984 a and b). They studied the way that users learn about a system by exploring rather than following the instructions of the manual. They found that this activity leads to a large amount of time being spent on recovering from unnecessary errors (errors caused by using parts of the system that they do not initially need to know about). Their solution was to block off these features from the user and force them to use only a sub-set of features of the system (Carroll and Carrithers 1984a, 1984b). In their own terms "system functions that new users typically do not need, but which can be springboards for errors and confusion, were disabled." (Carroll and Carrithers 1984a).

The authors found that more training wheels subjects completed the word processing task, spent 20% less time on the task, reached the typing area of the word processor almost twice as fast as those subjects learning in the unrestricted system, spent less time recovering from errors and performed better on a system comprehension test.

Their approach to training seems to be well founded and effective. Users don't follow manuals because they don't tell them what they want to know at that particular instant, which appears to be how to get to the typing area and how to print. Manuals do not allow users to follow their natural learning tenancies, following instructions does not promote learning in the same way that exploration does (Lewis and Mack 1981). The experimental results indicate that users perform faster and with less errors under the training wheels condition.

What is not clear is just how useful training wheels would be in normal working situations. The authors tested training wheels under two experimental conditions. In the first they were asked to produce a given letter. In the second the subjects were asked to learn how to use the system then produce a letter. The difference appears to be minor but it did make a difference to the subjects behaviour. While in both conditions the subjects split their time up into a learning phase then a working phase in the first condition their behaviour was guided by the goal of learning what they needed to know in order to produce the letter, in the second they had a more global goal of learning to use the system. The authors are aware of the differences between the two conditions but they do not attempt to draw any parallels between the subjects' goals in the experiment and their goals when at work. If field tested training wheels may produce very different results simply because of the different goals of the users.

The experiment shows how long each group took to reach the goal once. It does not illustrate learning in any direct sense, though the comprehension test does back up the claims. It only describes the benefits of the system during the very first stages of learning. The experiment does not show how well the training wheels subjects would perform once the restrictions have been removed. The system needs to be studied under field conditions.

Another aspect of the situation of the casual user is the requirement to move from one application package to another by transferring their knowledge to the new situation. Annett and Sparrow (1985) stated that one of the central problems in the practical application of transfer theory is in identifying they key elements of the task and predicting how they will interact. At the moment the only tool is intuition and experience. This problem is reflected in the studies of transfer between computing tasks.

Studies that have isolated parts of computer application tasks and examined the transfer between application programs conclude that positive transfer is due to higher level system concepts while reduced performance is attributed to syntactic differences in the user interfaces of the systems (Karat et al 1985, Ziegler et al 1987, Singley and Anderson 1985, Polson et al 1987). More work is needed on the theory of transfer before any progression can be made beyond recommendation to design for consistency.

Waern (1985) discussed transfer in a slightly different manner. Rather than looking at similarities between command names and actions she discusses the interaction between different goal-conditions-methods on different systems. She proposes that designers should look at the number of new rules that users have to learn, the transfer between rules (based on the similarity between conditions and consequent actions) and the status of the rules where transfer is likely to occur. Unfortunately it is difficult to predict what prior knowledge the user will apply and how automated the rules will be.

Hierarchical task analysis (HTA) produces break-downs of tasks in a formal, rigourous manner resulting in plans very similar to Waern's (1985) goal-conditions-methods. Shannon (1985) and Sinclair (1987) applied HTA to computer applications in order to investigate the feasibility of using this method for developing computer training programs.

Shannon (1985) examined the use of a computer based help facility that was based on a HTA of editing skills. The results of this pilot study illustrated a few difficulties with help systems in general (difficulties involved with learning the use of menus, using menus and displaying the help information) which tended to obscure the benefits of the substance of the help.

Sinclair (1987) used HTA to improve both the implementation of a software package and the overall training system; the resulting package needs to be field tested against existing methods, a difficult problem because she was looking at commercial software used within a company.

HTA is a useful tool for decomposing tasks but it does not solve training problems unless it is used as part of an overall training methodology (Shepherd, 1987).

Summary of the problem of casual users

Casual users tend to forget how to use previously learnt tasks. They perform a wide variety of work tasks and so need to be able to bend their system knowledge to the new situations. They may even need to learn to use several systems. Thus any savings in subsequent learning that can be encouraged through their initial training would be desirable. They make incorrect assumptions, generalisations and use of inappropriate analogies leading to errors. If their attempts to solve new tasks through these processes could be enhanced performance would be significantly improved. Satisficing is common due to exploration which is often haphazard, undirected and unsuccessful, error ridden and fear provoking and thus seen as an uneconomic and unattractive effort. They prefer to spend as little time as possible learning to use the system so any training materials provided should aid self exploration and educated guessing as much as possible. Training cannot prescribe how to carry out every work task or how to cope with every eventuality so users should be helped to solve problems for themselves. Evolutionary rather than compensatory help would therefore be the most appropriate form of training.

Training requirements

The characteristics of the user, their tasks, their present learning and performance characteristics and the current training provided have been described. These have given a picture of the problem to be investigated but have not described the nature of a fitting solution.

If the efficacy of the solution is to be tested the outcome must be observable. Stating that the user must 'know how to' use the system is not a useful approach because 'knowing' by itself cannot be tested. The criteria for the success of the technique must be in terms of the observable behaviour of the users. Thus from the above account the users must be able to show:

1. Longer retention of the skills that they originally learnt.
2. Increased performance at carrying out similar work tasks.
3. Increased performance at solving new problems that differ significantly from the original tasks.

These criteria will be elaborated in the discussions of the nature and measurement of skill. The study of the method of achieving these training goals should treat text as a tool to be used for learning. It should examine long term learning and should focus on material that is produced in a controlled and carefully defined manner based on an underlying model of the user.

What needs to be considered when developing a manual are the questions 'What does the user need to know in order to be skilled at using the tool?' and 'How can it be ensured that the reader gains from the manual the information that he needs to know , so that he can not only solve his immediate problem but become more skilled so that the manual becomes obsolete ?'

CHAPTER 3

PRACTICAL PROBLEMS OF MEASURING SKILL

Introduction

The previous chapter described the nature of casual users, the tasks that they perform and their approaches to learning. Their training requirements were outlined and the limitations on the sort of training that could be provided was described. Critical examination of previous studies of means of improving the training for casual users revealed two main items relevant to this chapter. Many of the studies used measures of skill that had questionable validity with respect to the work tasks of casual users both in the tests of skill applied and the time scale over which skill changes were measured. In order to develop and test strategies for training, tests that are valid, reliable and practical must be developed. Some studies examined user's performance at the task of editing entire documents; this chapter examines the practical problems of this approach and the validity of alternative, easier methods of skill measurement.

The aim of this experiment is to examine the problems of measuring computing skill. Schumacher & Waller (1985) describe some of the problems of measuring performance, list the following useful criterea : time available, goodness-of-fit, reliability, validity, ease of use and cost. This chapter describes the examination of a number of tests of computing skill according to these criterea, with particular emphasis on the validity and ease of use of the tests. Schumacher & Waller (1985) notes that 'Validity is frequently glossed over with cursory comment or simply not considered.' (p379).

In previous studies in the context of word processing, the method of assessing the changes in skill has been to measure user's performance at editing large documents (Briggs, 1990). In this and other areas skill has also been assessed by measuring performance on particular aspects of the computing skill that are held to be particularly representative or central to the particular package use, for instance Coombs, Gibson & Alty (1981). Sometimes the tests have been conducted on the computer (Reder, Charney & Morgan, 1986) others using just pencil and paper (Nowaczyk, 1984).

Although the assessment of skill using a complete document ought to provide a valid measure of the skill it is felt that there are practical problems of assessing performance in this manner. The experimenter is unable to control the exact means the user adopts to solve problems and it requires access to the computer on which the package is run. Pencil and paper tests that are focussed on particular aspects of the skill may provide a quick, accurate and portable measure of the skill. But the tests may not be a valid measure of the skill needed in the work place.

This piece of work consists of two parts. Firstly, skill will be measured using the 'edit a document' type task henceforth referred to as the 'document test'. The practical problems of using this type of test will be discussed, as will the theoretical problems. The results of the test will be used to form the subjects onto three balanced groups.

Each group will then receive the same set of paper and pencil questions tests of skill. The set of questions will be of different types, to match the range of tests that have been applied in the past (Karat, Boyes, Weisberger & Schafer, 1985; Reder et al., 1986). The groups will be allowed different levels of access to the computer in order to vary the medium that the tests are to be carried out in - if the tests can be carried out using pencil and paper alone (i.e. without requiring a computer) then this form will be the easiest and cheapest to implement. The correlation between the scores on these tests and those from the document tests will be assessed and the practical problems of instigating and marking the tests will be described. The chapter will end with recommendations of the easiest and most valid method of testing computing skill.

Performance test

This test was designed to simulate tasks in the work environment. A range of tasks (20) were selected and used as the basis for 'scrambling' a document, similar to the selection of tasks assembled by Briggs (1990).

The tasks were selected according to the level of the task operations in the command menus. The use of any sort of analysis of the task or skill as the basis for the selection of tasks is unusual; frequently 'a document' or 'a typical task' is selected by the author for their main test of skill. A test of skill such as this, whatever the medium, would normally be based upon a task analysis and should cover all of the items revealed by the analysis; this experiment is not a bench test but an examination of the effect of the test type and medium of the test on the performance of subjects. Details of the levels are given below.

Levels of cues in Word

Listed below are the 'levels' of the actions that can be performed with Microsoft Word 3.0. 'Levels' refers to the depth of the task operations within the menu structure. The lists were composed in order to justify the selection of different tasks in the second experiment which examines the effect of the medium of presentation of the tests. In the 'computer presentation' the user has access to all of the menus in the package and is able to pick up from these cues about the performance of the task. In the 'computer screen' presentation the user is able to see only the top level of the menu hierarchy and in the 'pencil and paper presentation' even these cues are not available as aids. The selection of levels according to depth in the menus does not necessarily reflect the relationship between the cues and ease of recall of the commands - some cues will be more effective than others. However, this method should ensure that the tests are fairly evenly balanced in terms of clues towards the answers.

Description of Levels

Level 1

{Mouse only
Main menu only}

- All highlighting.
- All deleting and inserting from keyboard.
- Printing.
- Quiting.
- Undo.
- All moving within the text.
- Change overtype mode.
- All typing of text.
- Use of one glossary.
- Open and size windows.

Level 2

{The above,plus
First level of menus
Bottom level of menus, using right mouse key}

- Use of all glossaries.
- Format character.
- Select menu for formatting.
- Move through help.
- Jump to a page.
- Select method of jumping.
- Library Autosort
- Select library option.
- Change Options.
- Select from the print menu.
- Replace text.
- Search for text.
- Load a file.
- Select from the Transfer menu.
- Split a window at a given line.

Level 3

{All of the above, plus
Second level features (Eg Transfer, Delete)}

- All other commands except those in level 4
- Print Direct
- Print File
- Print Glossary
- Print Merge
- Print Queue
- Transfer Glossary - Merge, Save, Clear

Level 4

{Features requiring use of a sub program of the package}

- Library Spell.
- Library Run.
- Gallery.
- All Format commands
- All Print Options
- Repaginate

The tasks that were selected are listed below:

1. Format Left 'Chapter 4'
2. Format Centre 'Some useful commands'
3. Underline (Chapter 4, ...)
4. Indent left margin
5. Page width .
6. Indent start of each paragraph (7).
7 Justify paragraphs (4)
8. Swap part of first sentence
9. Swap second part of second paragraph.
10. Page Break
11. Tab out table - NOTE spaces v. tabs !
12. Swap paragraphs.
13. Change lower case to capitals
14. Inset paragraph (Left and Right)
15. Turn off indent for that paragraph.
16. Merge document, at correct place.
17. Split part of second document .
18. Search and Replace disc for disk.
19. Spelling check.
20. Corrections to phrases by altering words.

Design

The order of the tasks were not rotated within the document because their particular combination in any instance produces a unique document that poses particular editing problems for the subjects. Tasks were not rotated in previous studies.

Subjects

The subjects were 30 male and female first year undergraduate psychology, ergonomics and 'introduction to human factors' students mostly aged 19 though some were older 'mature students'. All had received 10 weeks computing training from the same instructors, emphasising training on the Microsoft Word word processing package on the RM Nimbus networked microcomputer. Previous to the course the users had varying amounts of computing background.

Instructions and aid

A copy of the instructions given to subjects can be found in Appendix 1.

In response to questions subjects were informed that they were allowed to copy the document out. Although the concept of merging files was not explained, subjects who had difficulty in loading the first file were given assistance. It was felt that if this had not been done many of them would have given up before starting. Due to problems with the network subjects were later informed that they did not need to save the file or print it out, just inform the experimenter that they had finished. Subjects in distress or blocked on a particular problem were asked to do as much as they could or go on to a different part of the document; it was emphasised that they were free to leave whenever they wished. Subjects were requested to make every effort possible not to delete their work accidentally.

Subjects were allowed one hour in which to complete the editing at which point the experimenter saved and printed their work. If the subjects took less than one hour the time taken was noted.

Results

Practical problems of testing

Marking

Each task was examined separately. A note was made of whether the task was correctly completed and whether the correct method of carrying out the task was used. The latter was judged according to the formatting characters revealed using the 'Options, Visible: complete' command. For instance, indents formed by tabs or spaces appear different to those formed by formatting the paragraph with an indent.

Unfortunately the marking was not as easy as was hoped.

For example:

1. Titles in the text could have been placed on the left most edge of the text not through the user's actions but because they had not re-formatted the margins - paragraphs not indented (correctly) but also not indented where they should be.
2. The second document may have been typed in by hand instead of merged. If the copy was visibly perfect there was no way to tell which method the subject had opted for.

Optimal performance

For a skilled subject the easiest - least key strokes - method of carrying out the tasks was to use the special functions in the word processor rather than using simple editing commands. However, for those users who were unsure of their ability to use the functions the best path was to use the simple commands. This led to occurrences of subjects re-typing large portions of the text and/or using spaces for formatting instead of editing the original documents.

This meant that the experiment did not constrain the subject to use particular methods of achieving goals. Particular attention will need to be paid to the strategy that the users adopted in solving problems otherwise there will be no means of separating the scores of subjects who took the same amount of time but used different methods.

It is not possible to completely categorise the methods used or in some cases whether or not a given task has been conducted. See the examples above.

Help System

Subjects were able to use the built in help system in the word processor - it is impossible to turn this facility off in this package. In tests where the use of the help system is taken as an integral part of the computing skill this will not pose a problem.

Starting Problems

Some subjects were unable to load the document that was to be edited. This may have been because there had been a gap of several weeks between the end of the training course and testing. In order to gain information about the subject's other skills people in this predicament were assisted by the experimenter by loading the document for them.

Deleting the document

One subject accidentally deleted all of the work that he had carried out just before the end of the hour. In a work situation he would have been given a score of zero. However he was clearly fairly adept at using the system; it would have been inapropriate to have classified him for the next part of the experiment according to a zero score. It was therefore decided to drop the subject from the study.

Stress

Many people find the tasks of learning to use and using a word processor highly stressful. Many of the subjects in this study had only been using the package for ten weeks and knew that they were not very skilled. This, coupled with the test situation (which prevailed despite assurances that the experimenter was interested in their results whatever they were and that the results would be confidential and would not have any bearing on there marks for their academic work) proved to be highly stressful for some of the subjects, especially when they found they were unable to carry out a task or undo an error. One subject decided to withdraw from the study for this reason.

Theoretical problems of this type of test.

As stated above, the test does not constrain the subject to use a particular method of attaining a goal. Differences in strategy are not always revealed from an examination of the finished text.

Marking is not as clear cut as it was originally envisaged - it is sometimes difficult to tell whether the subject performed a particular task then obscured the fact by carrying out a second task incorrectly.

There is always the problem of balancing up the time it took the subject to complete the editing against his score for the task and also the strategy that he used. This issue will be discussed later when the numerical results have been described.

Numerical Results

Tables 1 and 2, shown below, display the totals per subject and the totals per task.

Table 1
Subject Totals - Performance Test

Subject	Total Correct	Total using spaces	Total also attempted	Grand TOTAL	Time Taken
1	9	9	2	18	dnf
2	8	3	9	11	57
3	10	3	7	13	52
4	3	1	16	4	dnf
5	7	2	11	9	dnf
6	16	3	1	19	47
7	5	0	15	5	dnf
8	2	0	18	2	dnf
9	10	3	7	13	55
10	9	5	6	14	dnf
11	13	7	0	20	48
12	7	3	10	10	dnf
13	10	4	6	14	dnf
14	5	3	12	8	dnf
15	13	1	6	14	53
16	15	2	3	17	60
17	14	4	2	18	50
18	11	7	2	18	dnf
19	6	2	12	8	dnf
20	5	2	13	7	dnf
21	15	4	1	19	58
22	11	4	5	15	60
23	11	7	2	18	dnf
24	17	0	3	17	59
25	7	6	7	13	dnf
26	10	9	1	19	dnf
27	8	5	7	13	dnf
28	15	3	2	18	34
Means	9.66	3.64	6.64	13.30	

Table 2
Scores per Task - Performance Test

Task	Level of question	Total Correct	Total using spaces	Grand TOTAL	% using spaces
1	L4	24	0	24	0
2	L4	8	10	18	56
3	L4	19	1	20	5
4	L4	11	4	15	27
5	L4	13	3	16	19
6	L4	4	19	23	83
7	L4	6	3	9	33
8	L1	25	1	26	4
9	L1	24	1	25	4
10	L1	7	1	8	13
11	L1	5	17	22	77
12	L1	20	2	22	9
13	L4	2	17	19	89
14	L4	7.5	7	14.5	48
15	L4	9	3	12	25
16	L4	12	7	19	37
17	L1	18	0	18	0
18	L3	15	1	16	6
19	L4	16	5	21	24
20	L1	25	0	25	0

The aim of this part of the experiment is to provide a basis for forming the subjects into balanced groups and to examine the problems of using this type of test for measuring skill.

Categorisation of subjects

The first method of splitting the subjects is according to their total score on the test.

This is made up of the tasks that the marker judged that they performed correctly on and those that they attained the goal of the task by using simple editing skills. The relationship between the ranking of the subject's number of tasks 'correctly' performed and the ranking of their total score, and the ranking of the number of tasks that they performed 'simply' and their total score is shown in graphs 1 and 2. The correlations are .855 (n=28, p<0.05) and .647 (n=28, p<0.05) respectively. This shows that the number of questions answered using the 'correct' method is a better indicator of their total score than the number of 'simple' answers. The graphs of the scores for these measures indicate that subjects, whatever their ability take the same strategy in carrying out the tasks - they use the same proportions of simple and correct methods for attaining the task goals.

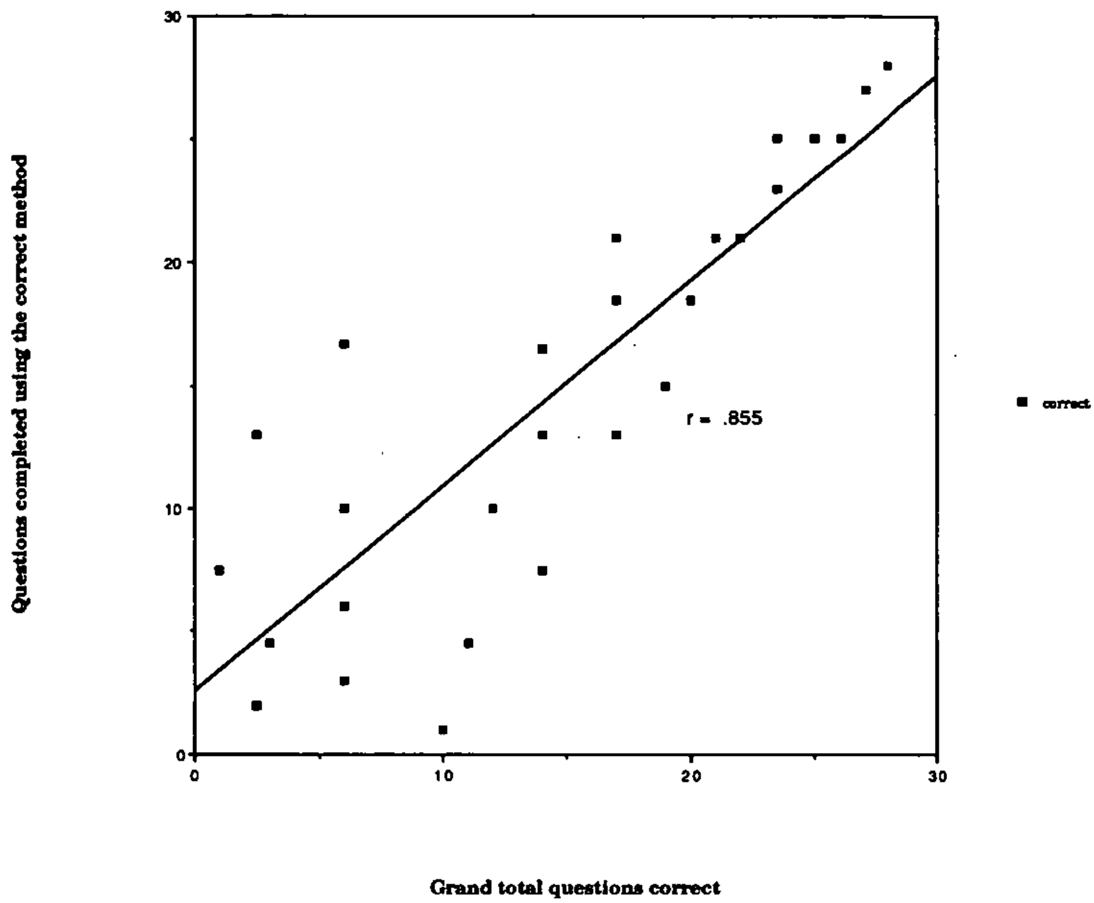
Subjects could not be split into groups of low level users/ high level users so the allocation of subjects into three matched groups depended more on total score than on strategy, though care was taken to place the odd individual in an appropriate group so that not only scores but also strategy and time taken was taken into account. It did not prove possible to measure the number of tasks attempted because of the global effect of some of the tasks on the document. It was felt that a measure of tasks attempted taken from the final document rather than a record of the session was very unrepresentative of the activity during the session.

Analysis of effect of question type

No overall pattern of the effect of the level of the task functions within the application package menus emerged, though the Level 4 tasks were attempted by using spaces slightly more frequently than the level one tasks - 37% versus 20% respectively, though not a significant difference. By ignoring the results for the Level 4 tasks that can not be imitated by using other functions this figure rises to 47%.

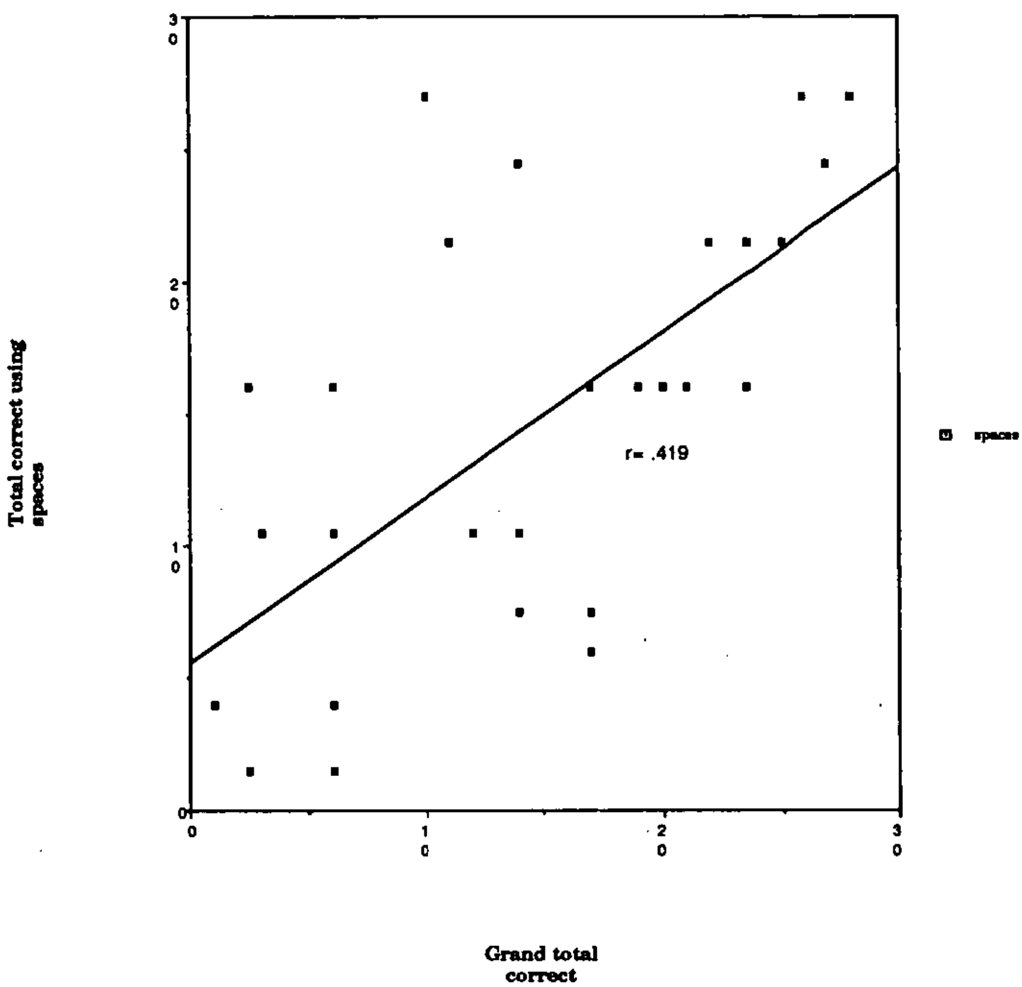
Graph 1

Total questions completed using the correct method against grand total questions correct



Graph 2

Total correct using spaces compared with grand total correct



Summary

This test has provided a test of subjects skill at using this word processor on a task that closely approaches the sort of task they perform at work. The measures used have been shown to be a valid way of dividing the subjects into equally skilled groups for the second part of the experiment, the comparison with simpler and shorter tests.

It was found that the number of questions that the subjects completed using the 'correct' or optimal method using system commands rather than making do with spaces was a better indicator of their overall score than the number of questions they answered using an 'incorrect' method.

The level of the commands required to carry out the optimum method within the menu hierarchy of the passage was found to have a slight, but not significant, effect on the subject's strategy; this smaller than expected difference may be due to the small number of deep commands used in this test that can be mimicked by using spaces; removing the results relevant to these increases the proportion of Level 4 tasks that make up the total number of tasks carried out in a non-optimum manner. This result indicates that even this valid, work based task must be carefully constructed in order to ensure that it is not possible for relatively unskilled people to perform well by using non-optimal strategies.

The next section of the experiment will examine the validity of shorter tests that involve less supervision, equipment and are easier to mark. The tests will be presented using different mediums - users will have access to the computer to test their ideas, or will be able to see a picture of the screen (thus showing the first levels of menus which may act as cues for recall), or will be presented with the questions only.

The result of these tests and the effect of the different mediums available will be compared with the measure of skill derived in the first part of the experiment.

Paper Tests

Introduction

The aim of the experiment is to find the lab test that has the highest correlation with actual performance while creating the fewest problems in administering them.

Coombs et al. (1981) adopted an approach to studying individual differences in learning computing skills that was based on selecting 'indicator' tasks that were designed predict the user's performance on tasks closer to the actual work tasks - 'target' tasks. Although they went on to describe qualitative details of the users performance on the course designed to teach the computing skill (in this case FORTRAN) there were no checks that the target tasks were valid tests of FORTRAN programming.

The selection of their target tasks was based on the hypothesis that subject's ability to handle complex sets of variables is determined by a small number of underlying skills. In this case they proposed that the skills were the ability to learn about a large number of individual operations and the ability to assemble the objects to fulfil a particular programming goal. Coombs et al. (1981) propose that although the users are taught the details of operations they are not taught how to assemble them.

Formally, the first test required users to correct syntactic and clerical errors in an otherwise correct program. The second test required them to assemble syntactically correct lines into a functioning program that would solve a given problem.

Main Hypothesis

The assumption behind Coombs et al (1981)'s work is that their target tests are good indicators of the user's skill and that people who perform well on the performance test should perform equally well on the paper and pencil tests. I.e. there will be a high positive correlation between each of the target tests and the performance test.

Other hypotheses

This experiment is also designed to test the effect of different mediums for presentation of the tests on performance. As argued in Chapter 4 the command names and menus used in application packages are aids to learning and performance. Different methods of presentation and access to the computer during testing will alter the amount of these cues that the user is presented with. When the tests are conducted entirely on paper without a model of the computer none of the cues will be present (with the exception of the details provided in the question, but these are constant across the media). When users are allowed to use the computer to try out their ideas all of the cues normally present in the task will be available. It is therefore likely that performance will be enhanced in the computer presentation condition.

1. Total score on all of the tests.

computer group>prompted group>plain paper group

This matches the amount of cuing present with each group's test medium. The larger the amount of cuing, the easier the test. The computer group will have the additional advantage of being able to experiment in order to work out the correct answer.

3. Items involving commands deep in the menu structure.

computer group>prompted group>plain paper group

The computer group is the only one that will be able to examine the deeper menus in order to select the correct command; in addition they will able to fiddle in order to find the correct answer. Items deep in the menu are reached via the surface menus; the prompted group will have an advantage over the plain group in that they have the higher menus displayed in front of them.

4. Items based on low level commands.

computer group > prompted group > plain paper group

See the explanation above. The effect in this case will be smaller because cues will be available to all of the groups.

5. Transfer to Microstar System

computer group = cued group = plain group

All of the groups should perform equally well on this test because they all receive the test in the same medium.

6. Test of ordering of commands

computer group > (cued group = plain group)

Being able to try out ideas will assist the computer group. The cued group and the plain group should perform equally well because the cues for this test are provided in the question.

Subjects

The subjects used in the performance test were used in this experiment. They were allocated into three groups balanced according to the measures described in that experiment. (See 'Categorisation of subjects')

Apparatus

Three sets of test questions were used. One set was presented as a question plus some text associated with the question, one set presented as a question plus the text displayed on a drawing of the screen in the basic menu state, and a third set was presented as a question plus instructions to load the text associated with the question onto the computer. The third group were advised that they were able to use the computer to work out or check their answers but their final answer was to be written on the sheet - what they did with the computer would not be marked. Although the first two groups conducted the test in the terminal room they were seated so that they were unable to see a Word screen; they could however see the keyboards of the computers. The third, computer, group were allowed to use the electronic help but this was not suggested to them.

Types of test selected:

The tests are based on those used by Coombs et al. (1981) and a variety of other tests commonly used.

Test 1

'Please indicate what you would do next (after the sequence of commands listed below) if you were working on the following piece of text and you wanted to ...'

load the file LETTER.DOC from the A: drive ?

Command sequence:

Type T for Transfer
Type L for Load
Next ...

Test 2

Someone is working on the paragraph below; you are given a description of what they are trying to do and a sequence of commands for carrying out that operation. Please fill in the missing command in the sequence.
(The cursor and the document are shown in the state they were in BEFORE the sequence of commands started.)

They are trying to underline the word 'prototype'.

Sequence:

Missing ...
Press F for Format
Press C for Character
Click on 'yes' next to underline
Press Return

Test 3
In section 3 questions you are given a document that someone is working on, a description of what they are trying to do and a sequence of commands that may be in the correct order to carry out that operation.
Please indicate whether the sequence is correct or incorrect.
(The document is shown in the state it would be before any commands are carried out.)

The aim is to save the document under the name PART1.DOC on the A: drive.

Sequence:

Type T for Transfer
Type S for Save
Type A:\PART1.DOC
Press Return

CORRECT / INCORRECT

Test 4

Please describe (as briefly as possible) how you would use the Word word processing package to ...

swap the second and third sentences over in the paragraph below.

Test 5

The following two pages briefly describe a word processing system. Please look at them now.

In the questions that follow you will be asked to describe how you think you might be able to use the package to carry out some tasks. You are not expected to know how to use the package - what I would like you to do is make an educated guess at the answer (because you already know a fair amount about computers) and write it down in the space below the question.
Please make some attempt on each question - any answer is better than none at all.

Selection of test questions

The test questions were selected according to the level of the answer in the command menus - each group of tests contains the same number of deep/shallow items

Questions were presented on both sides of the paper, one question per side regardless of the amount of white space. Answers were written on the question sheets. This was carried out in order to make the tests appear of equal length and to reduce the overall thickness of the document (final version amounted to about 20 pieces of paper). Presenting the questions on different sheets prevented the subjects from looking ahead or at their previous answers.

The screen imitation was designed to match Word 3.0 on the Nimbus as closely as possible (though in black on white). Compromises were made, the screen could hold slightly more text than the computer (in order to get the dimensions of the screen in proportion) and the menu was created in a rather ragged bold in order to emphasize the screen layout. Cursors, pointers and windows were included; document location markers were not. All imitation screens were shown with the 'Alpha' highlighted to indicate that the machine was in command mode.

Example of the paper only test :-

Test 1.1

Please indicate what you would do next (after the sequence of commands listed below) if you were working on the following piece of text and you wanted to ...

load the file LETTER.DOC from the A: drive ?

Command sequence :

Escape
Type T for Transfer
Type L for Load
Next ...

The appearance of the mock screen is shown in Figure 1.

Procedure

The original intention was to conduct the test in two large lecture groups, but due to individuals not turning up groups varied between 1 and 12.

Subjects were not allowed to go back to questions after they have turned the page (the contents of later questions may be the answers or be clues towards the answers of earlier questions).

There was a 15 minute limit for each group of questions. Subjects were instructed to wait after completing a section until either they had all finished the section or the 15 minutes was up, before moving on to the next section.

The subjects were not closely observed during the experiment except to monitor their completion of sections.

Instructions

Subjects were told that these tests were designed to measure the same sort of thing as the performance test but that they were much shorter and carried out on paper. All of their answers were to be as brief (1 or 2 lines) and clear as possible. It was pointed out that questions were on either side of the paper and that they were not to go back to a question once it had been attempted.

Figure 1 - The paper version of the screen and a question.

Test 1.1

Please indicate what you would do next (after the sequence of commands listed below) if you were working on the following piece of text and you wanted to ...

load the file LETTER.DOC from the A: drive ?

Command sequence :

Type T for Transfer
Type L for Load
Next ...

The paper describes early work on a prototype intelligent system whose purpose is to apply principles of induction and dialogue to interaction with a user to produce a full and consistant 'technical' specification of a task or product. The paper outlines the knowledge representation, induction and dialogue processes, the object based program implementation, and gives excerpts from a typical dialogue.

COMMAND: Alpha Copy Delete Format Gallery Help Insert Library
Options Print Quit Replace Search Transfer Undo Window
Edit document or press Esc to use menu
Page 1 {.} ? Microsoft Word:

NOTE - the reduction in size has degraded the image quality

Results

Numerical Results

Totals per level per question and overall total

The questions have been scored according to the criterea : correct, cheat (i.e. non-optimal strategy using spaces rather than commands), wrong and don't know as appropriate to the question types. The results shown here in Table 3 are grouped according to question type (left to right) showing the score for each classification of answer and grouped according to the level of the question - i.e. the level of the task in the application menu hierarchy. Means are shown on the right extreme of the table.

Table 3
Totals per level per question and overall total

	Q1						Q2						Q3					Q4					Q5	Mean
		N	Correct	Cheat	Wrong	Dont Know		N	Correct	Cheat	Wrong	Dont Know		N	Correct	Wrong	Dont Know		N	Score	Dont Know		Score	
C	L1	2	15	0	1	2	L1	3	15	0	4	8	L1	3	14	11	2	L1	10	77	0	L1	70	17.17
	L2	3	14	0	2	11	L2	2	15	0	0	3	L2	0				L2	10	60	2	L2	49	18.25
	L3	2	10	0	0	8	L3	3	17	0	1	9	L3	3	13	12	2	L3	0			L3	40	15.00
	L4	3	11	0	0	16	L4	2	12	0	0	6	L4	4	30	4	2	L4	10	55	3	L4	71	17.17
S	L1		15	0	2	1	L1		18	0	5	4	L1		20	6	1			88	0	L1	82	20.17
	L2		16	1	3	7	L2		15	0	0	3	L2							76	1	L2	56	19.25
	L3		12	0	3	3	L3		14	0	3	10	L3		22	4	1					L3	48	18.00
	L4		15	0	3	9	L4		11	1	3	3	L4		22	10	4			68	1	L4	71	16.00
T	L1		16	0	2	0	L1		14	0	9	4	L1		19	8	0			76	0	L1	74	19.00
	L2		15	1	0	11	L2		16	0	0	2	L2							71	0	L2	55	19.50
	L3		11	1	3	3	L3		16	2	3	6	L3		23	4	0					L3	50	18.50
	L4		12	0	2	13	L4		15	0	1	2	L4		27	8	1			53	3	L4	72	18.25

Media

From Table 4, shown below, of correlations it can be seen that both the computer and screen groups had significantly higher correlations with Part 1 (correct scores) than the text based test (r = .92, .85 and .23 respectively) and that the computer and screen groups had very high correlations with the performance test total scores.

The aggregate scores (totals of correctly answered questions balanced for a max of 10 marks per subject per question group) were 292, 328 and 323 respectively. These differences were maintained throughout the question types. The subjects in the computer group often left the last couple of questions in a set blank because they found it difficult to complete the entire set in the time given. It was impossible to discriminate between this type of blank and those arising from lack of knowledge however an examination of the data indicates that this is probably the cause of the difference in total scores.

Table 4

Table of correlations between paper and pencil groups (q's correct) with performance test groups (q's correct).

	Group	Q1	Q2	Q3	Q4	Q5	Total
Presentation	Correlation						
Computer		0.97	0.85	0.75	0.7	0.46	0.92
Screen		0.79	0.64	0.9	0.35	0.2	0.85
Text		0.54	0.05	-0.12	-0.13	0.3	0.28

Question types

In all three groups Q5 (transfer to a new system) shows the lowest correlation with the P1 results.

All of the tests in the computer group correlated highly with P1 with the exception of Q5. The order of degree of correlation is Q1>Q2>Q3>Q4>Q5

The same result holds for the screen group though the correlations are lower except on Q3 where the correlation is .90, the highest of all of this groups results. The former trend applies to the text based groups correlations (which are far lower than those of the other two groups) with the exception of a proportionaly much higher correlation on Q5 which is of the same order as the correlations of the other groups for this question. This result is to be expected since this question is presented to all of the groups using a picture of the MicroStar screen. By far the highest correlation is between computer group Q1 (supply the next command in the sequence) and the results of the performance test.

As described above, the total score per question type remained at about the same level for different question types and different media though those for the computer group were somewhat lower probably due to their inability to complete all of the questions in the time given.

<u>Levels</u>

There do not appear to be any differences between any of the groups with respect to the different levels of questions; there are some very small differences on some questions but no over all pattern.

This is surprising because the computer based group has access to all of the levels in the application while the paper only group has no cues at all; the cues should have led to better performance of the computer group on the Level 4 questions. A possible explanation for this result is that the levels of cues does not map well onto the types of difficulties that subjects have, as was indicated in the first part of the experiment where subjects did not 'cheat' to a high degree on the Level 4 tasks.

Qualitative Results

Problems with administration

In the laboratory where the experiment was conducted subjects in the paper and screen groups were able to see the keyboards of the computers; although it was originally thought that this would not effect the subject's performance there were instances where subjects could be seen to be examining the keyboards for the names of keys and functions. This problem could be easily reminded in future tests.

Some of the subjects in the computer group used the electronic help system. As mentioned in the description of the performance experiment, this behaviour is acceptable if the system is included as an integral part of the word processor.

A more serious problem is that some subjects in the computer group were unable to complete the tests in a section within the 15 minute limit. This was probably due to their experimentation with the computer. Some subjects, mainly the Text group, often had to wait 5 to 10 minutes before they could proceed with the next section. The time limit should be extended and subjects run individually.

All of the computer group quickly learnt not to save any altered text and to load the appropriate piece of text for each question, even though the instructions concerning these actions often led to confusion before the start of the experiment.

Some of the computer group were unable to load the experimental text, despite the instructions at the start of the experiment. Some required individual help at first.

A small number of subjects forgot that questions were on both sides of the paper - they needed to be reminded of this at the end of each section. In future the questions should be on one side only.

None of the subjects showed any distress, many were bored with having to wait for other subjects to complete a section of the task.

Problems with marking

It was difficult to weigh up the answers to Q5 (Transfer to MicroStar) - in general 8 marks were awarded to adequate answers (showing an approach that might work), less than 8 marks for incomplete of inappropriate strategies and more than 8 marks for answers that suggested more than one approach to the problem.

The most problematic questions to mark were those that needed a knowledge of the spelling checker - many of the group had not been taught about this feature and so were forced to guess about it when their general level of ability often suggested that had they received the correct tuition they would have coped well with these questions.

As stated earlier, answers were marked as 'cheating' if the subject ignored the sequence of commands presented and suggested a completely different strategy.

The short tests were faster and less ambiguous to mark than the performance test - the text did not need to be examined for the use of spaces; the optimal strategy in any given situation was clear.

Strategies

The Text Only Group tended to write down their answers very quickly whereas nearly all of the computer group tried out all of their answers. The latter strategy frequently led to situations that they had not anticipated; this was usually followed by attempts to rectify their mistakes and thus they wasted time.

Conclusions

The degree of validity of these tests with respect to the performance test is computer>screen>>text based on correlations of .92, .85 and .23 respectively. An interesting result is that although the size of correlation varies if the questions that the computer group missed due to lack of time are taken into account the groups have very similar scores both in total and throughout the question types. This would suggest that the different media do not have a consistent effect on each subject. Out of the question types the pattern of the magnitude of correlation was Q1>Q2>Q3>Q4>Q5 - the difference in size of correlation shown by the group total was reflected in the question correlations with the exception of Q5 where the groups were more evenly balanced. This was because each of the groups received this question in a screen based format. Q1 therefore appears to be the best indicator of performance on the large editing task.

The level of question type, as based on the depth of the information in the Word menu structure required for the answer, does not appear to have an effect within or across media. This suggests that 'level of question' is not a determinant effect due to media type - contrary to the original predictions.

There are some practical aspects of the test that could be improved on, the only major one being the timing of the group tests to ensure that the computer group have sufficient time to answer all of the questions and the other groups are not waiting for too long. If the test were to be administered again it would be useful to see the effect of running all 5 questions as a single block to be carried out as quickly as possible.

There were very few problems in marking any of the questions - Q5 was the only one that required some judgement but a criteria for doing this was easily formulated.

General Discussion

Theoretical results

The performance test reveals the strategy that users would use when given the opportunity to use their knowledge of the system in a practical context. The shorter tests force strategy and thus provide a measure of knowledge of commands rather than preference. The highest correlation between P1 performance and P2 performance was the computer group (r = .92) and in particular Q1 (r = .97) which tested their ability to fill in the next step in a command sequence. Next highest was the screen based group (r = .85) then considerably worse the text group (r = .23). It is surprising that recall of commands would be the best indicator rather than the tests that were supposed to measure 'higher' aspects of the skill such as ability to transfer to new systems, ability to choose and recall the complete command sequence. This result is probably due to the poor match between what the questions are supposed to measure and what they do actually test. Dreyfus & Dreyfus (1986) and Rasmussen (1989) in their descriptions of hierarchy of skill both describe self reflexive aspects of skill where subjects are able to discuss their strategies - this type of skill is not the same as the skill necessary to perform the task and is the reason for these tests that have a poor correlation here with actual performance.

Another interesting result was that the 'level' of a question did not have any effect within or across groups; this is suggests that it is not the source of the effect of the different mediums on performance nor the factor that made questions difficult to answer.

Practical results and difficulties

The short tests were faster and less ambiguous to mark and also seemed to cause less distress to the subjects. They also required minimal supervision whereas the performance test required the experimenter to be on hand to help subjects out of dead ends and prevent the total loss of the document, which did happen in one case. Once the timing problems of the short tests have been remidied these should be faster than the performance test.

Summary

This chapter has examined the validity and ease of application of different tests of word processing skill with respect to a test of their ability to edit a document. It was found that the most appropriate medium for the presentation of the tests was the where subjects had access to a computer and the opportunity to try out their answers. The tests that most closely approximated the scores on the editing tasks were those that presented the question and required an answer in the same form that the user would carry out the task on the machine. This result matches the concept of the existence of several aspects of skill - from the ability to perform a task to the ability to reflexively describe ones actions.

The experiment also revealed problems with the selection of tasks to make up the tests. The use of levels of cue did not aid the selection of tasks; levels did not determine which tasks were difficult or easy or the strategy that users would adopt to perform the tasks. A further examination of the nature of skill and a means of selecting aspects of work tasks is required in order to provide a foundation for producing improved training strategies and measuring their effect on the development of skill. These aspects of skill will be described in the next chapter.

CHAPTER 4

DEVELOPMENT OF A MODEL OF SKILL FOR THE DESIGN OF MEANS OF ENHANCING AND TESTING SKILL

Introduction

Operational research solves problems through the following sequence of steps (Ackoff, 1956):

1. Formulate the problem
2. Construct a mathematical model to represent the system under study.
3. Derive a solution from the model.
4. Test the model and solution derived from it.
5. Establish controls over the solution.
6. Put the solution to work.

This strategy closely matches the approach adopted here. In Chapter 1 the subject and argument of this thesis was stated; in Chapter 2 the relevant literature was used to expand the details of the problem, to describe methods of approaching the problem that have already been tried and to argue where these solutions have been deficient. It concluded with the issues that this thesis will attempt to resolve. Chapter 3 was a further evaluation of the problems of the area; it examined the practical problems of measuring skill and the selection of those parts of the task to be used for testing skill. Although results of the experiment were useful for the selection of the method of applying tests, the method of selection of those parts of the task to be used in the tests was found to require a further development. In particular a model of skill leading to an operational definition of measures to be used for test task selection was needed. This is also required for the derivation of training solutions for casual users. The aim of this chapter is to provide such a foundation.

It has been found that casual users tend to forget how to perform previously learnt tasks due to their intermittent use of the tools. They perform a wide variety of work tasks and thus need to be able to bend their system knowledge to the new situations. They may even need to learn to use several systems. Thus any savings in subsequent learning that can be encouraged through the method of their initial training is desirable. They tend to make incorrect assumptions, generalizations and use inappropriate analogies which lead to errors. If their attempts to solve new tasks through these processes could be enhanced, performance would be significantly improved. Satisficing is common, due to attempts at exploration which are often haphazard, undirected and unsuccessful, error ridden and fear provoking and thus seen as an uneconomic and unattractive effort. They prefer to spend as little time as possible learning to use the system so that any training materials provided should aid self exploration and educated guessing as much as possible. Only when they first start to use the computer, the honeymoon period, are they prepared to put effort into learning to use the machine. Training cannot prescribe how to carry out every work task or how to cope with every eventuality, so users should be helped to solve problems for themselves. Evolutionary rather than compensatory help would therefore be the most appropriate form of training.

The purpose of this chapter is to establish criteria for the selection of measures of skill, as explored experimentally in Chapter 3, and to examine means of improving the problem of skill acquisition by casual users as described in Chapter 2.

Although the validity of different mediums and specific types of test as measures of the skills used to edit an entire document was found, the tests were selected on a pragmatic basis in order to cover a range of approaches.

In order to formulate a training solution for casual users a model of the skills that need to be acquired needs to be formulated. In order to select tests of these skills the model should be referred to and applied to the task to be studied in order to pick out aspects whose performance by subjects indicates the acquisition of skill.

A model to represent the learning system under study (task, user, manual) will be constructed which will provide a basis for the solutions that will be developed in Chapter 5.

Training requirements

The characteristics of the user, their tasks, their present learning and performance characteristics and the current training provided have been described. These have given a picture of the problem to be investigated but have not described the nature of a fitting solution.

If the efficacy of the solution is to be tested the outcome must be observable. Stating that the user must 'know how to' use the system is not a useful approach because 'knowing' by itself cannot be tested. The criteria for the success of the technique must be in terms of the observable behaviour of the users. Thus from the above account the users must be able to show:

1. Longer retention of the skills that they originally learnt.
2. Increased performance at carrying out similar work tasks.
3. Increased performance at solving new problems that differ significantly from the original tasks.

These criteria will be elaborated in the discussions of the nature and measurement of skill.

The definition of a skill

The interpretation of a single instance of behaviour is a difficult and bold act. Repetitions of the same act under the same circumstances only serve to confirm that the person has the capacity to do this. It is the variation of action under the same conditions or vice a versa that informs far more about the person than any number of repetitions. For instance, it would be premature to describe a person as highly skilled at chess if they were able to reproduce the opening moves of a master, even if the moves perfectly combated the actions made by the opponent. So much more is revealed if they were unable to repeat the sequence in the next game, or reproduced it despite its flaws as a suitable response to a different set of opposing moves. Skill is a label correctly used to describe behaviour, but not behaviour that is an isolated episode.

The term skill is used loosely to describe a variety of behaviours. One person may describe another as skilled at tossing a coin by merely being able to repeat the action on demand. A second judge may include the ability to toss a coin while carrying out some other action or perhaps carrying out the coin tossing with a particular flamboyance. It would not be correct for the second judge to say that the person was unskilled at coin tossing if the performance lacked the required verve, merely that the person did not meet the criteria specified in his definition. For the definition of skill is circular and ideosyncratic. In order to be able to define skill sufficiently rigourously for it to be measured the actions that go to make up that skill must be tightly specified. When that is done the label 'skilled' may be dropped and replaced by a precise operational definition of the task. The substance of that definition depends on the objectives of the author.

In this situation the objectives are to increase the user's ability to perform despite long periods of not using the machine, to increase their ability to transfer their knowledge between similar tasks and to improve their ability to solve new problems. The definition of computing skill could have included the ability to describe their actions verbally, to train others to perform the same tasks, to explore new systems efficiently and to describe the inner workings of the machine. However these abilities are not the skills that casual users need. The three points listed form the basis for more precise definitions which will include particulars of the tasks that they will be expected to perform and the criteria for success at those tasks.

The measurement of skill

While stripping the secondary behaviours away from the user's use of computers has been useful in order to produce a definition of the required skill, a useful test of the efficacy of a training solution would have to occur within their normal work environment. Within this situation the usual pressures to produce finished work, to carry out other tasks in parallel, use other systems or not use computers at all for periods will apply. If the user is able to perform his work better under the new training regime in these circumstances, not just in the laboratory, the solution can be said to be effective. It is for this reason that methods developed in the isolation of the lab using a restricted set of tasks must also be tested in the field.

There is a danger of producing laboratory tasks that are so refined that they test skills and problems that do not occur in the user's normal working environment. These problems are examined in a study described in full in Chapter 3 and in brief immediately below. Further problems of measuring skill are elaborated in Chapters 7 and 8.

A study was conducted to examine the effect of testing using pencil and paper instead of a computer and of testing skill by asking users about their behaviour. Their performance on these tasks was compared with their ability to edit a normal document with the word processor. The cues available from the computer were reduced from a condition where they could test out their answers on the computer, to being able to see the screen but not use it, to seeing a paper mock-up of the screen and finally presented with the test questions only. The answers required varied from descriptions of how they would perform a task to completing task sequences to sorting commands into the correct logical sequence. It was found that the tests where the user had full use of the computer and those tests that most closely resembled the actions they would normally perform on the computer produced scores that correlated highly with the work task scores. Those tests carried out with paper and pencil alone and those that required the user to reflect on their knowledge of word processing produced the lowest correlations. Laboratory tests of skill should therefore approximate as closely as possible the actions performed in the tasks at work and should be carried out on the same computer.

Transfer and Problem solving

The current ethos in the learning literature (the literature since Card, Moran and Newell (1983)) is to break a skill into a group of production rules (as in an AI program) that will guide the person's behaviour according to the stimuli received. Transfer is the amount of learning time saved (or required in addition) in learning a new task after learning a prior task.

In the literature the amount of transfer between two skills is held to be proportional to the number of common rules between the skills (Polson,Bovair, & Kieras, 1987); though exceptions to this principle are frequent. Anderson (1987) describes skill acquisition as the compilation of specific rules for that skill from weaker rules which are used initially to solve the problem that demands the formation of the new skill. In other words problem solving is the repeated use of previously acquired rules that are not quite perfect for the current problem that eventually lead to the formation of a new optimum set of rules. Some fundamental limitations to this idea were pointed out by Gick and Holyoak (1980) who show the difficulty subjects have in perceiving the common elements of the task - this is very different from the results of studies looking at transfer in more 'basic' skills. The results of the study (Tetzlaff, 1987) which examines the effect of learning the same sets of production rules but in different orders show that transfer cannot be predicted by a simple count of production rules. Also, since transfer is measured in a high level cognitive task such as those targeted in this thesis by examining the user's performance on a new task such a measure is also testing their ability to solve a new problem.

Thus there is a strong link between problem solving and the mechanisms of transfer - the descriptions are intertwined. Dreyfus & Dreyfus (1986) believe that human problem solving cannot be reduced to the level of logical rules - they argue that intuition and other non-mechanical processes are essential for skilled behaviour. Therefore not only are the definitions of these two measures inseparable in terms of the underlying theory (which is in itself sketchy) but there are also fundamental disagreements between major theorists.

Throughout this thesis the approach has been to treat the computer user as a 'black box' whose behaviour can be modified by changes made to the environment. The changes suggested in this thesis are made after consideration of theories relevant to this area of psychology; similarly the measures of behaviour used in the experiments though grounded in theory are operational in nature and designed to guage changes of behaviour that indicate the acquisition of the particular skills being studied. So each measure needs an operational definition for each experiment.

One can imagine tests specifically set up to check whether a subject has acquired a particular production rule from the text, but in concrete terms these tests will be the same as those not quite so intensely ground on a particular view of the inside of the 'black box'. Since we are unable to categorically say whether a particular performance is due to creative problem solving or the transfer of old skills to a new problem it is sufficient to measure the user's ability to perform new tasks as this is the practical indicator of level of skill.

Such a 'lay man's' criterion for expertise at a skill avoids the theoretically difficult areas called problem solving and transfer. However the definition of measures will need to be refined in order to match the model of skill used as the basis for the development of the training strategies.

This description of skill has outlined the behaviours that will be encouraged in this study, the necessity of a definition of skill in terms of observable actions and has provided a sketch of the main points to be considered when testing skill. In order to progress with the description of skill and its measurement the model of skill used in this thesis needs to be specified.

It is not proposed that the internal processes of the model or items of knowledge described are real and can be tested; rather the model that uses them allows the experimenter to make hypotheses and draw conclusions about training conditions and skill. If these prove to be accurate the model is no more real but it can be regarded as more useful.

Model of Skill

It has been argued above that in order to define a skill the behaviour must be thoroughly described. This is a prerequisite of any systematic approach to training (Annett,Duncan,Stammers, & Gray, 1971; Stammers, 1971; Shepherd, 1989) and for IT tasks Shepherd (1989) recommends HTA as an effective method.

The process of HTA produces a description of the actions necessary for a given task to be completed without the analyst needing to take psychological considerations into account. Others, such as TAKD (Johnson,Diaper, & Long, 1984) require effort to be expended on describing the task knowledge of the user or the psychological difficulties involved in performing a particular task. Such information is not required in this study; the selection of HTA as a method of analysis will be discussed in more detail in Chapter 6.

Whatever the method used a task analysis will produce statement of the tasks to be conducted and a description of the operations to be performed in order to carry out those tasks. In HTA operations need to be iteratively re-described in sufficient detail for the particular training context. Annett et al (1971) have argued for the P x C stopping rule that states that operations should be re-described until the probability of the user not being able to perform the operation times the consequence of failure on that operation is negligible. Shepherd (1989) recommends taking the analysis as far as the key stroke level in the context of training for IT tasks; this criteria will be used here. The description of a task consists of the list of key strokes that must be made in order to complete the task, and nothing more. Although the analyst may create plans to describe operations and names for sub-tasks and organize the plans into hierarchies none of these entities exist in the task or can be categorically asserted to exist within the mind of the user. See Figures 2, 3 and 4. They may represent the analyst's insights about useful training strategies or may merely be the product of the trainers favourite way of breaking down complex tasks e.g. Shepherd & Duncan (1980). It is the task of the trainer is to present these key

Figure 2

A task analysis of the general task, 'examine the remainder of a letter'

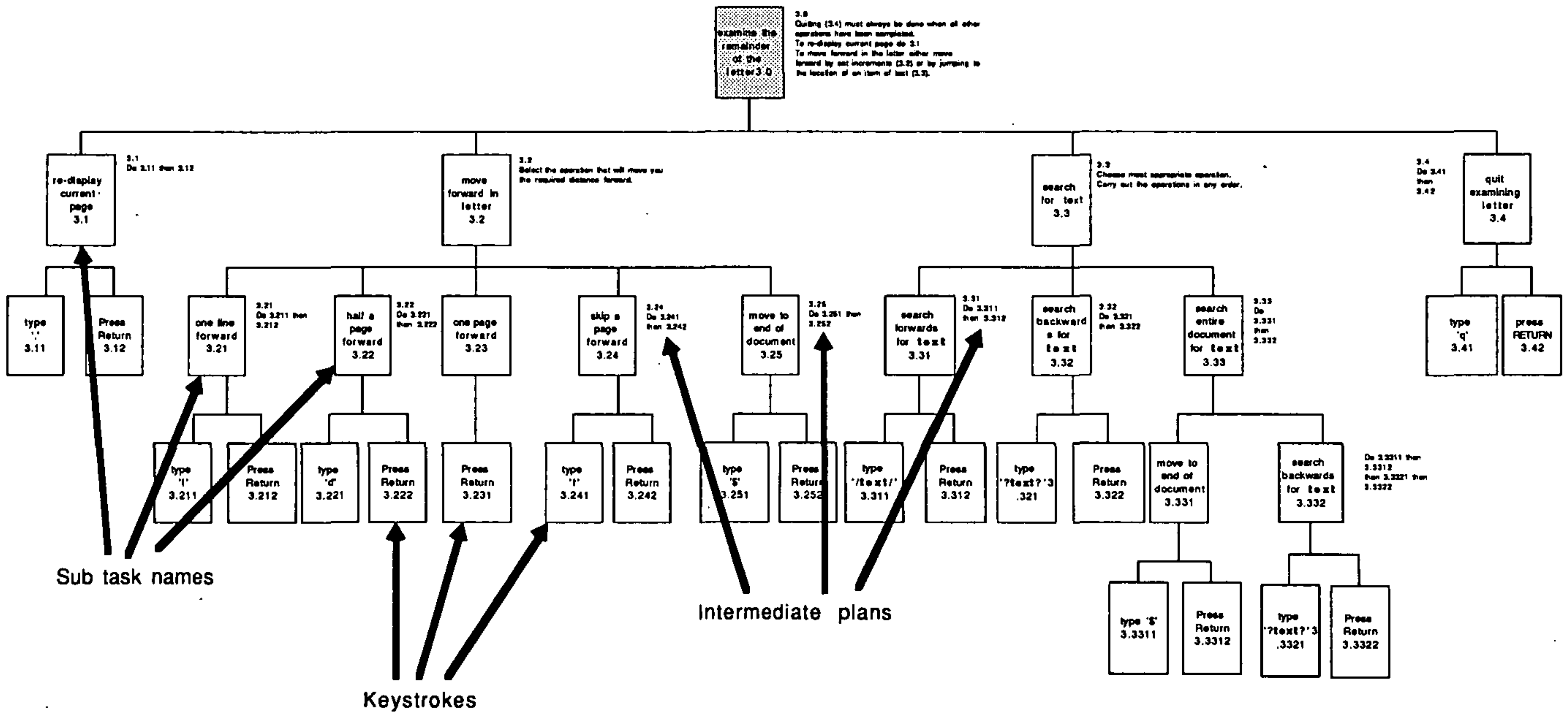


Figure 3 - Analysis of a task from plans to keystrokes

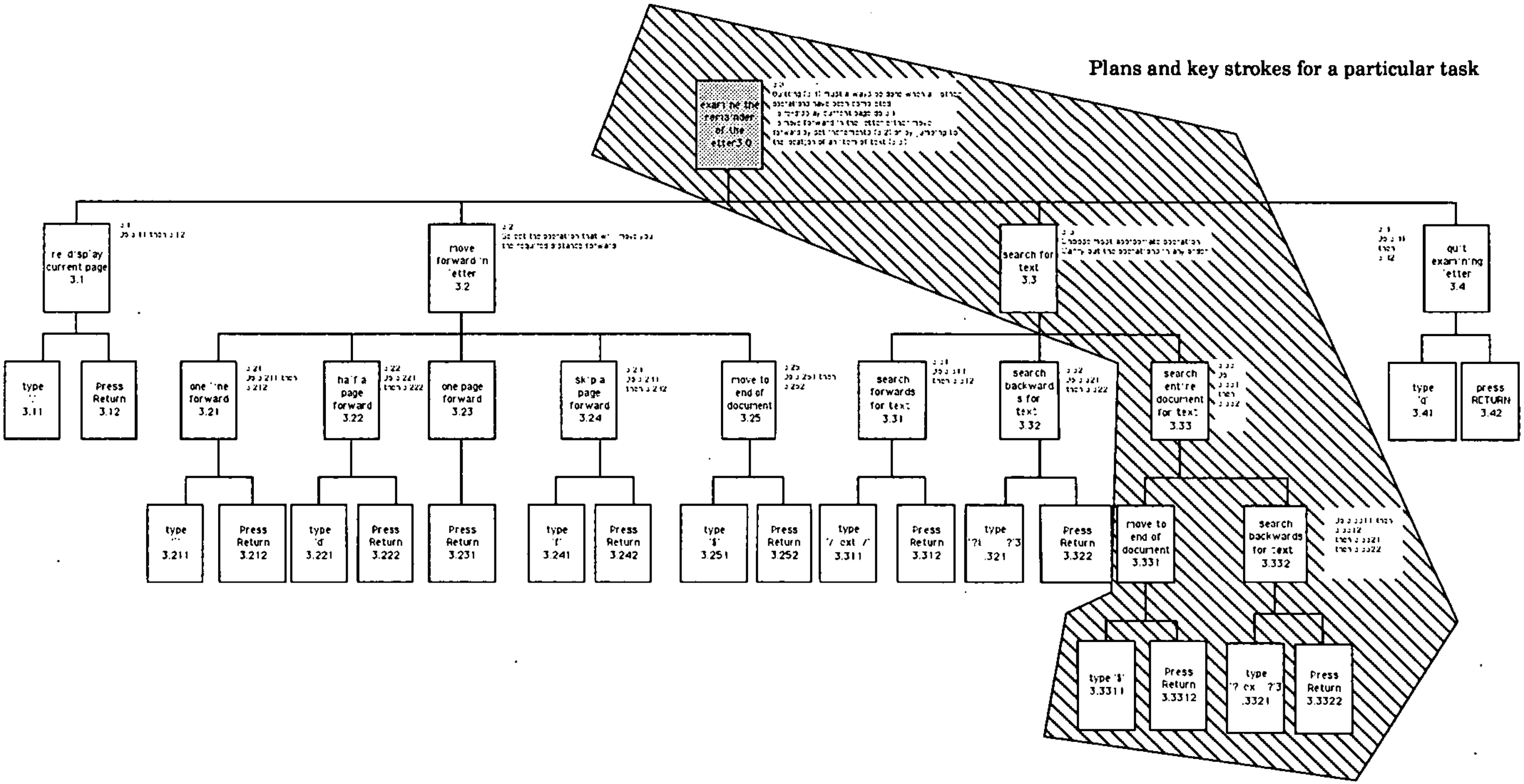
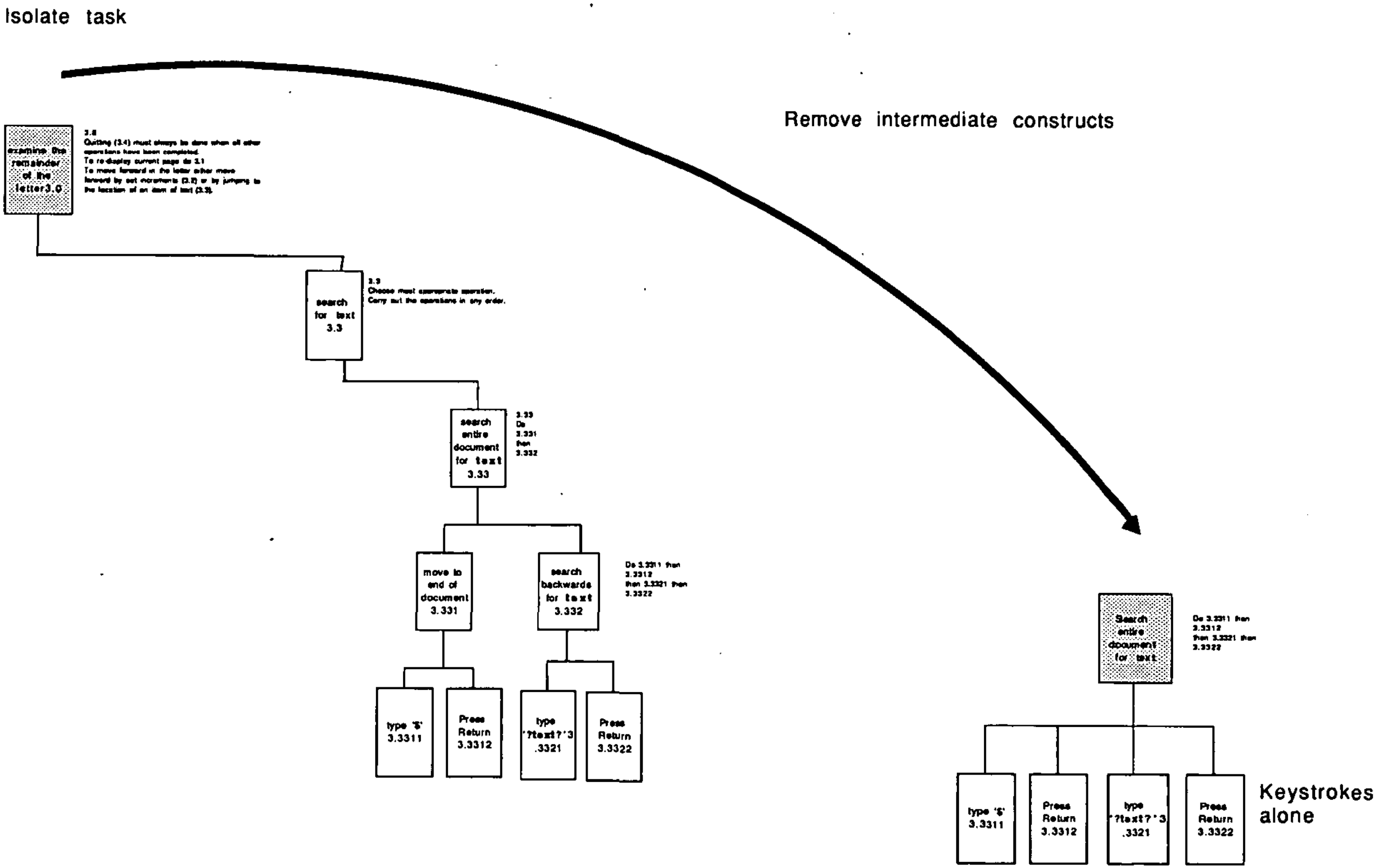


Figure 4


Isolate task
Remove intermediate constructs
examine the remainder of the letter 3.0
search for text 3.3
Choose most appropriate operation.
Carry out the operations in any order.
search entire document for text 3.33
3.33 Do 3.331 then 3.332
move to end of document 3.331
search backwards for text 3.332
Do 3.3311 then 3.3312 then 3.3321 then 3.3322
type '$' 3.3311
Press Return 3.3312
type '?text?'3 .3321
Press Return 3.3322
Search entire document for text
Do 3.3311 then 3.3312 then 3.3321 then 3.3322
type '$' 3.3311
Press Return 3.3312
type '?text?'3 .3321
Press Return 3.3322
Keystrokes alone

stroke sequences in such a way that the user learns to perform the task skilfully, based upon his model of skill and training strategies (Shepherd, 1985). The model of skill used in this thesis is presented below after a description of the relationship between the structures and labels within the HTA of a task and training. The training strategies will be developed in Chapter 5.

Supposing that the key strokes that control the package do not contain any indications of the functions that they control. For instance, the key stroke to load a file is commonly the letter 'L'; the designer has selected 'L' in order to help the user to remember which key to press for 'Load'. This approach works providing that the user has been told that the function is 'Loading' a file; if the function is described as 'Reading' a file 'L' would no longer be a useful key stroke design.

Function	Function Name	Key stroke Useful ?
Load	Load a file L	Yes
Load	Read a file L	No

Although the key stroke and the function of the key stroke remains the same, the name given to the function by the designer determines the ease of learning the key stroke/function relationship. Ausubel (1961) points out that the link between words and concepts is an arbitrary one, so learning the name of a function and associating it with the concept that it represents should be just like learning any nonsense pair. However in the case of assigning names to functions designers commonly use names that already have meanings for the users in order to 'bootstrap' the learning of the new function. 'Load' already has many meanings for people - to load a cassette, to load a lorry, to load an electrical system. In choosing this word to denote the transfer of bits from a file to memory designers hope to associate the new function of loading a file with the common usage of the word 'load' by analogy.

Although people can learn artificial grammars of nonsense syllable combinations if they are exposed to sufficient examples designers try to aid learning by replacing the nonsense syllables with useful meaningful names that imply the function and sometimes the syntax of the commands.

A second method used by designers to aid the learning of their package is to organize the commands hierarchically to imply a grouping of command functions. Thus in 'Word' the menu command 'Transfer' is used to group commands that act on files - e.g. 'Load', 'Save', 'Merge'. This ploy is also used by technical authors who divide manuals into chapters and sections according to common elements of commands. By learning the grouping of commands users can save time and learning effort rather than learning lists of 'unrelated' commands because they have an organizing framework to hang the details on (Graesser & Nakamura, 1982). This is one reason why menu driven systems are easier to learn than command line systems. A second reason is that the commands are present in the interface, acting as information about potential system functions that the user may apply.

Instead of organizing commands into functional groups it is possible to pick out the organizing features of the system, present these then describe how they can be applied to the use of the commands. For instance, in word a central feature of the use of the package is the action of highlighting a piece of text that is to be the subject of the next command. It is possible for the author to organize commands into functional chapters and for each command state that text must be highlighted before issuing a particular command e.g. highlight the text before pressing 'Delete'. Alternatively the author can state than text must be highlighted in order to make it the subject of the next command then go on to describe those commands that make use of this feature. By pointing out the organizing feature explicitly users may learn to apply it to other commands without needing to be told to do so.

It should be becoming clear to the reader by now that any addition to the description of the key strokes needed to perform a task (which in itself may include inherent additional information about the system functions) is an elaboration that may enhance skill acquisition. Hypotheses about the usefulness of an elaboration for enhancing skill can be based on two areas of theory - models of skill performance and acquisition and models of the effect of elaborations on memory. The latter will be described in Chapter 6, the former is presented below.

From the HTA of the package use a list of key strokes to perform each of the user's tasks can be formed. This is what the users will eventually need to be able to perform. From the description of casual users the changes in behaviour that are needed have been extracted :

1. Longer retention of the skills that they originally learnt.

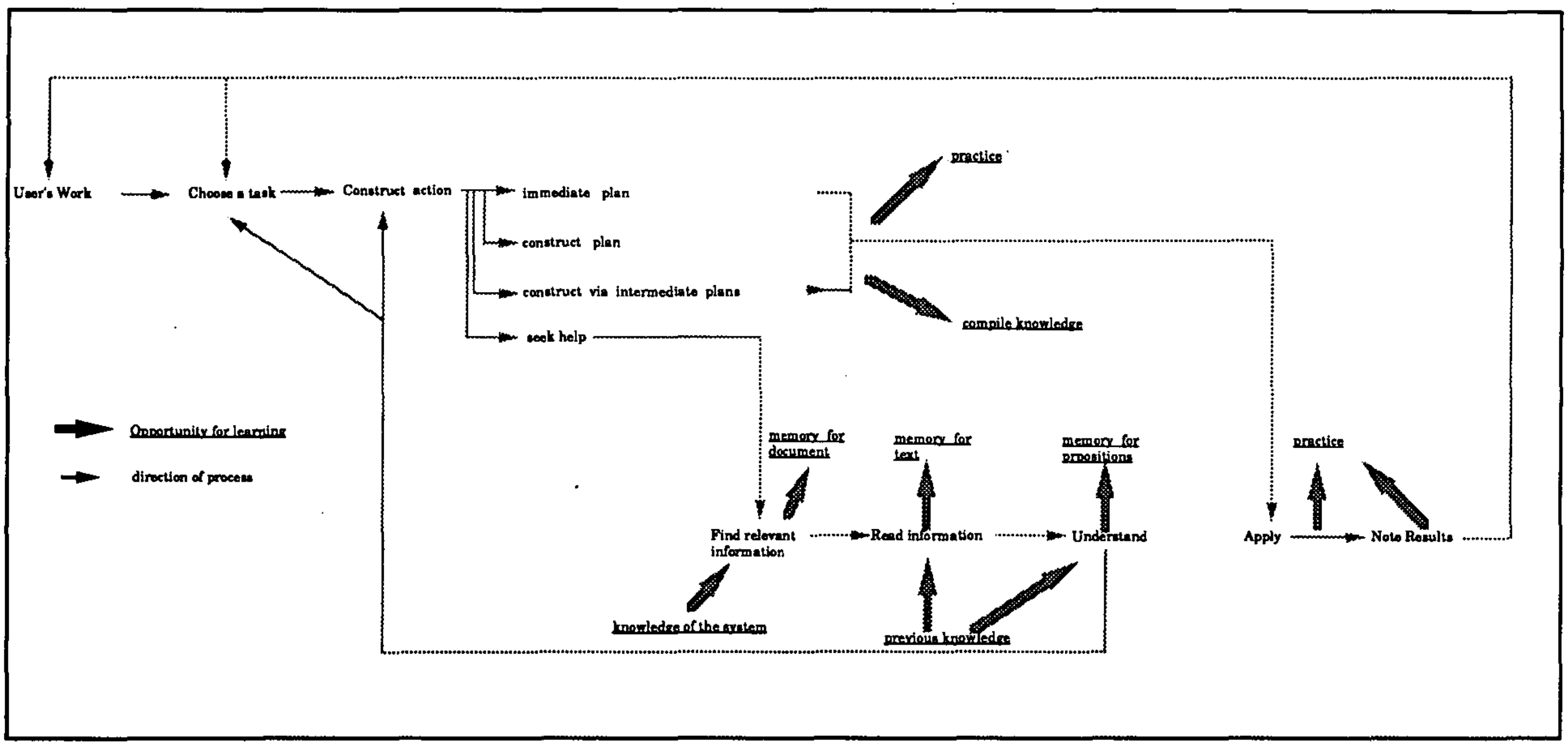
2. Increased performance at carrying out similar work tasks.

3. Increased performance at solving new problems that differ significantly from the original tasks.

The model of skill described here is that developed by Andrew Shepherd (Shepherd, 1990). Although the full form of the model describes the process of carrying out a task from the sampling of the environment to monitoring the effect of the results it is the way that the user constructs the means of achieving goals that is pertinent here.

When faced with a task the user has four possible ways of constructing an action. They may know the sequence of operations necessary for the task and will carry them out. They may know the prerequisite sub-tasks and be able to work out how to assemble them because of knowledge of their superordinate plan to enable them to carry out the task. Or they may be able to break the task into a number of problems and solve each one in turn by using their knowledge about the reasons behind the plan. Finally they may not know enough to solve the task and will have to either abandon it or seek help. See Figure 5.

Figure 5- Manual Use Model



Casual users need to be able to perform a variety of tasks rather than just carry out a limited number of routine tasks. If the latter were the case then it would be a useful training strategy to teach the users the key strokes needed to perform the tasks then they could recall them when necessary. However, when faced with a new task the user needs to be able to construct a response. According to the model they can do this by using knowledge of sub-goals or deeper knowledge of the reasons behind sub-goals. Chapter 6 will describe ways of promoting these forms of knowledge.

Knowledge of sub-goals and reasons behind the operation of the package may be useful to the user once acquired but it may be that teaching the user to perform tasks by rote is a better way to train the user because they may develop the deeper knowledge more effectively by monitoring their own actions. For instance, they may learn the deeper knowledge but quickly forget it because it is not used whereas the knowledge of the key strokes is repeatedly reinforced. Again, this issue will be discussed in Chapter 6 when describing the selection of elaborations for training.

Development of the training strategy.

Casual users are not prepared to spend time attending a formal training course. This is lamentable because such a training situation allows for the control of the presentation of information and opportunities for practice. These sequences give the trainer the opportunity to reinforce learning and help users to learn from their attempts. In this way the trainer can ensure that component skills are acquired and that the user is able to apply them to a progressively wider range of tasks.

A trainer would also be able to control the difficulty of the tasks that the user attempts - for casual users the difficulty is determined by the work they must perform; unless their manager is far sighted enough to progressively load the user they will have to cope with tasks as they occur.

The only time that casual users will take an interest in training materials is during the initial 'honeymoon' period of use (Eason, 1976) and when they find they are unable to carry out a task.

Since the user will not attend training courses he must be trained from a distance.

Distance learning resources include audio tape, visual material (including video, slides and film) and documents (paper and electronic). The first two are not suitable because of their cost, the difficulty of presenting them at the work site and their linearity. Sitting through a tape based course or video course is likely to be just as unattractive to the casual user as attending a formal hands on training program; possibly more so because in the latter personal and human attention is provided. Although technical authors may be frustrated by user's tendency to jump from one section of a manual to another it is this feature, along with low cost and easy accessibility, that makes documentation an acceptable medium for training for these users. Because electronic documents obscure the screen during use and users prefer to read long documents in paper form, a paper based manual is more suitable than an electronic one (Cohill & Williges, 1985). For these reasons this thesis aims to examine the potential for increasing computing skill, as defined above, by enhancing the information provided in computer manuals.

The core of the documentation will be based on the description of the tasks necessary to attain work goals on a application package taken from a HTA of the use of the package. These are the minimum set of instructions that describe how to carry out each task. The method of presenting these so that the skill is learnt will be described in Chapter 6.

Conclusion

Skill has been operationally defined, as have the training requirements of casual users. The practical problems of measuring skill have been described in Chapter 3.

HTA has been selected as the method for analyzing the task. It was pointed out that the structure of the hierarchy and the labels of the sub-goals are elaborations on the basic task description. The way that these elaborations can aid training was illustrated but a task analysis is not a training strategy; a model of skill is needed as a basis for designing training. The model of skill described by Shepherd (1990) has been briefly presented; its use in aiding the selection of elaborations to the task analysis will be described in Chapter 5.

Finally the training requirements of casual users was discussed again and it was concluded that the most appropriate medium for the training was in the form of a paper manual.

CHAPTER 5

REVIEW OF THE MEANS OF ENHANCING TEXT USED FOR SKILL ACQUISITION

Introduction

In Chapter 4 the training requirements of casual users that were teased out from the relevant literature in Chapter 2 were dissected in order to examine them concerning the model of skill used in this thesis. This is based on the Shepherd's (1990) operating model of skill that describes how users compile actions to fulfil work goals. HTA can be used to describe these goals and the keystrokes necessary to meet them. The model of skill is a basis for constructing elaborations between the goals and keystrokes that can be used to aid the training of casual users.

On re-evaluation of the training needs of casual users and the opportunities for training them it was argued that the most appropriate medium for training computer users was text in the form of a manual. This chapter will describe the issues surrounding the use of text as a medium for training these users and will show how the model of skill along with a theory of text processing was used to produce a number of hypotheses about the best way to present the description of a task to a user in order to promote learning. The production of different types of text that embody these hypotheses will be described in Chapter 6.

This chapter will describe the progression in the literature from studying text as an abstract stimulus to text that is used for a particular function. A model of the use of text in this training context will then be described. This will be used as a framework for describing the devices in the literature of reading and memory research that are held to be means of increasing the probability that a particular piece of text will be retained in memory by a reader. It will be argued that while these can be applied to abstract pieces of text they do not enable the trainer to select which pieces of text should be elaborated in order to promote skill. It will be argued that recall of text is not a sufficient goal of a training strategy and that a suitable solution will provide a means of promoting skill and not just recall.

The development of the study of elaborations to text from aids for recall to aids for skilled performance will be critically described. From this literature (the memory and reading literature that was described earlier) and the models of both skill and manual use methods of elaborating text for skill acquisition will be produced.

Text is used in many different ways; the design of a particular piece of text should be based on the function of the text. Diehl & Mikalecky (1981) classify the use of documentation as using to do, using to learn and using to review/assess. Wright (1983) classifies the behaviours associated with text use as searching, understanding, applying. Diehl and Mikaleky (1981) also characterizes the features necessary for each type of document use and describes appropriate styles. Simple readability indexes are of little use for technical writing, as argued by Kern (1985), Diehl and Mikaleky (1981) : adjusting the index does not make the text easier to use. They propose that the degree of use depends the purpose of the reading, the familiarity with the text and the importance of the information for the job. These factors relate to texts used in an array of jobs, not just for learning or reference. They classify the purpose of the reading into 'reading to do', 'reading to learn' and 'reading to assess'. As will be described later, these classifications can be further broken down and are associated with different reading styles. They suggest that many of the features that make text easy to learn may make the text difficult to use; this is a suggestion and not the result of experimental work.

More concrete justification for using job based tasks comes from a study of macro reading behaviours (Pugh, 1977). Pugh found that skilled readers adopted different reading strategies in response to different types of questions about a text. These included scanning and search reading (answering 'How?' and 'What?' questions), skimming (deciding how to or whether to use a text), receptive reading ('What is the author saying ?') and responsive reading ('What do you think about what the author is saying ?'). It may be the case that different results for the same mathemagenic behaviours have been caused by the adoption of different reading strategies by the readers in response to different types of questions presented by the experimenters. Waller (1979) also notes the need for text designers to start looking at the actual use of texts by users rather than at the results of standard reading tests.

A pointer to a solution to this paradox is their recommendation that if the author knows that the text will be used repetitively for reference he/she should provide efficient access structures. They suggest that written information will be used more if it is seen to be more efficient, exact or direct than information gotten from other sources. Reading to do essentially requires concise, very well indexed and referenced information while using to learn requires summaries, introduction to topics, self test questions. The two styles are at odds. It is possible to write 'learning' manuals that are well referenced and so can be later used as doing manuals if the reader is prepared to sift his way through the additional text. However, casual users tend to spend the minimum amount of time necessary to find information on reading the text and are likely to be frustrated by, and ignore text that has been inserted as an aid to learning. By their nature they will not read manuals from cover to cover as many authors intend but will try to pick only what they think is necessary.

The problem here is how to produce a text that is sufficiently terse, quick and easy to use for the casual user and yet constructed so that skill is quickly developed.

The first step in constructing such a text is to outline a model of how text is used by casual users.

Manual use model

This model is very closely tied to the model of skill used in this thesis adopted from Shepherd (1990). In this case more attention is paid to the act of seeking help and the points in the process where skill may be enhanced.

An observer of a computer user at work would see a cycle of action from the selection of a piece of work to be carried out on the computer, a series of actions to complete the work and then back to the demands of the office. Within the cycle there may be pauses for thought, references to manuals and other sources of information, key stroke sequences, messages from the computer and other public actions. The purpose of Figure 6 below is to *make sense* of those observable actions by interposing cognitive activities likely to occur based upon personal experience and selecting relevant theories from the cognitive science literature.

The *observeables* in the diagram are the user's work, seek help actions, reading, apply, the computer's response and the sequencing of these actions. Unless the user's work is a rigidly prescribed sequence of keying actions on the computer (which occurs very rarely), the first private action is the decision to carry out a particular part of the work using the computer (Eason, 1981).

The user chooses a goal from his work task(s) and attempts to construct an appropriate action. This may be through the recall of a sequence of keystrokes, or constructing a plan of sub-goals whose associated keystrokes are known, or by constructing intermediate goals, or finally by seeking help then selecting a goal.

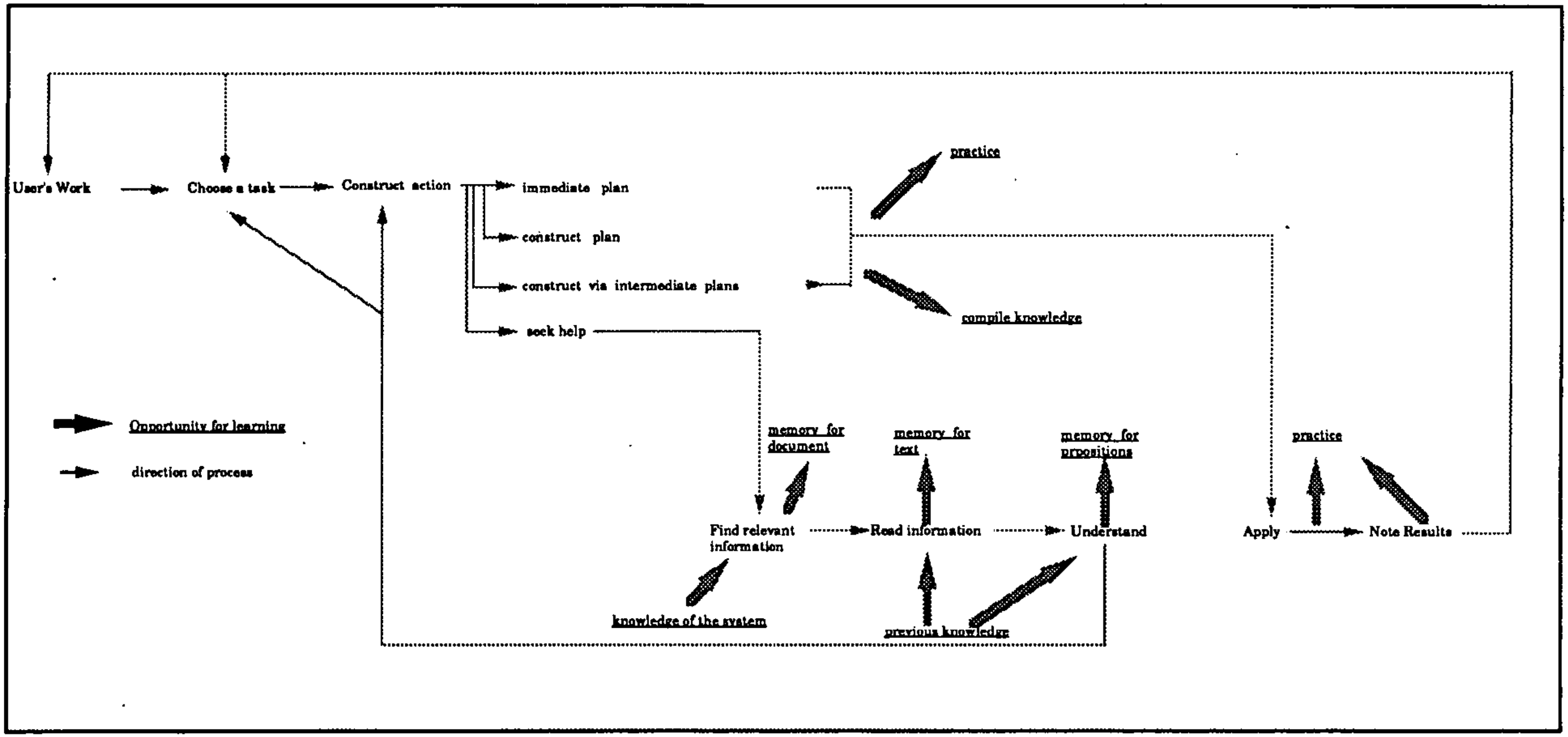
Factors determining their choice of action include their knowledge of the functionality of the system, their understanding of the mapping between the system functionality and their work goals and their assessment of their ability to use the system

The process of recalling key sequences or constructing sequences may enhance recall for these items. Constructing a sequence of operations from sub-goals may lead to compiled knowledge (Anderson, 1982).

If the user is unable to construct an action that they are prepared to implement they may decide to abandon the task (if they are at liberty to do so (Eason 1981) or seek external help. The source of help will depend on the facilities available to the user - local experts, manuals, job cards - and their preference for the source of help due to their previous experience with the different mediums. The information that the user will search for is dependant on their understanding of the relationship between the system functionality and their goals, their appreciation of the information that they require to solve the given problem and their tendency to satisfice.

The action of seeking help has been modeled in two ways. It is seen as a method of acquiring plans for action in the skill model. The

Figure 6- Manual Use Model



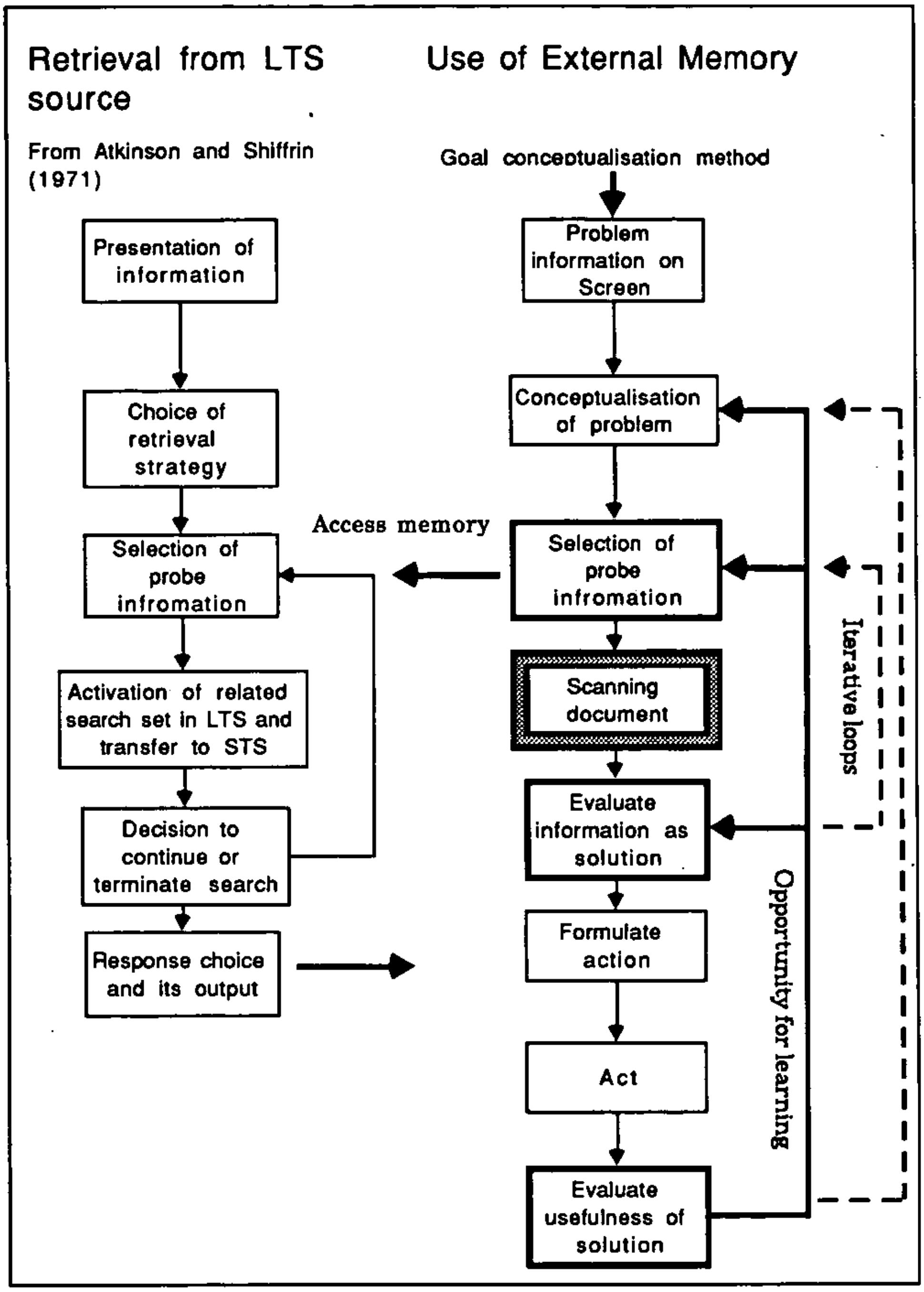
second model is designed to draw a parallel between Atkinson & Shiffrin 's (1971) model of the retrieval of an item from the long term memory system and the retrieval of information from an external memory - see Figure 7.

In the skill model as the text is read, it can be stored as an episodic memory, and it can be processed by the reader so that their knowledge of the skill of computing can be enhanced. The acquired information is then used to aid the user to construct a new goal. While reading the document the user may also read information not directly related to his current problem; this may be stored in memory. The user will also form memories of the location of the information (Rothkopf, 1971) and of the structure of the document . The information selected as relevant to the current goal is fed into the goal decision system as before. If the new information allows the user to act then the results of the actions are noted and evaluated. Even if the act is not useful the process of making that particular response has been practiced and may be more easily recalled in the future. A decision about the next goal to be reached is then made.

In the external memory model, the goal selection process and the information about the problem displayed on the screen is used to form a conceptualization of the problem. From this the user decides what information he needs to look for, the probe information, then scans that document. The process of scanning involves the use of the index, the structure, headings and typography of the text, as well as periods of reading the text to locate information . There are opportunities here for remembering parts of any of these items. When information matching the probe information is found it is evaluated as a solution. If it is not useful a new goal must be selected, leading to the selection of new probe information. If however it does appear to be useful an action is formulated, carried out and evaluated. As stated before, this process is another point in the model of skill and manual use where successful actions can be reinforced.

The model has identified processes where memory for the skill can be cultivated: when actions are recalled, when actions are constructed, a manual is scanned, text is read, text is processed and when the user evaluates his actions.

Figure 7 - Use of external memory



A logical analysis of the potential for intervening in these areas will help to focus in which way that text should be manipulated. Text could be produced so that the user only recalls operations and does not construct them or alternatively constructs the majority of his actions. This could be accomplished by manipulating the information content of the text. Episodic memory for the text and text structures acquired during scanning and could be strengthened. The likely effect of this on skill seems uncontrollable and of doubtful use. The user may remember pieces of the text but whether the user could apply it and to what extent it could enhance skill is dubious. Episodic memory for the text read could aid skill if the user could perfectly recall the appropriate areas when needed but this would be a difficult task and of limited use since the text could not describe all of the situations that the user might encounter. The information that the user gains from the text, the additions to his knowledge of the skill of computing, is an area that can change the user's ability to formulate actions. The study of the acquisition of information from text will be taken further.

Responses from the system after carrying out an action, if any, are interpreted by the user. The results of the action are judged and stored with a memory for the plan. Expected, correct results should tend to reinforce the action sequence used for that task. Repetition of that cycle, as used in the training of users, is thought to increase the likelihood of that skill being recalled in the correct situation. If the user's action did not accomplish the current goal the effect of the systems response on learning is dependant on:

> the sense the actor can make of the results,
> whether the actor able to decide which part(s) of the plan have been effective,
> whether the actor able to decide why he was not effective,
> any changes in the system requiring new goals to be met.

The actor may then decide that the task is finished or that more work needs to be done or that he is unable to complete the work.

Whatever the manipulation carried out to the text the trainer still has no control over the sequence of learning that the user must accomplish in order to perform his work tasks. However the text can be manipulated so that when the user locates the information to help him to solve his particular problem it is presented so that his skill will be enhanced.

There is a wealth of literature describing ways of improving text (Hartley, 1974; Hartley, 1981). Unfortunately much of it is atheoretical; authors describe ways of changing the appearance of text and the effect in a particular situation (Coles & Foster, 1975; Frase & Schwartz, 1979). This leads to guidelines for designers such as '80 ways of improving instructional text' by Hartley (1981). Waller (1979) makes the following criticism of this method:

"If there is no genuine foundation for the process of systematic enquiry we are engaging not in science, but in what Liam Hudson (1972) called 'the cult of the fact'."

Any material added to a text adds information. Text added to a core text can be of three types. It can aid the recall of the core text, enhance the use of the memory for the core text or direct behaviour itself. The latter refers to information that is not part of the task analysis; it is new information that may provide a mental model of the system and/or assistance in error recovery.

Carroll's approach to this training problem - 'The Minimal Manual' (Carroll,Smith-Kerker,Ford, & Mazur-Rimetz, 1987) - addresses the coarse learning needs of the user in terms of their need for active learning, error recovery strategies and the particular problems of learning to use one particular system. The text that he adds to a core text is essentially of the third type.

This study examines the potential for enhancing behaviour by using the first two types of additional text, focussing the study on the issue of how to deliver a set of instructions that are sufficient for the user to learn to perform the task rather than on how to supplement the text with additional descriptive material. This allows for precise experimental manipulation of the text while keeping the information content standard throughout. That is not to say that supplementary information, such as that added to 'The Minimal Manual', is not beneficial to learning but such information is difficult to precisely define and constrain so that replication of its effects in different contexts is difficult, as are examinations of the basis for the effects.

Most of the results of the studies of means of improving text are equivocal with many of the methodological issues as yet unsettled (Walker, 1979; Faw & Waller, 1976).

While typography research (for a review see Macdonald-Ross & Waller (1975)) affects the appearance of the text and can improve some functions of text (Young & Wogalter, 1990) most readability research aims to manipulate the prose style of the text. There appears to be two main approaches; altering the length and number of words per sentence and a more grammatical approach that examines the deep structure of the text. Duffy & Waller (1985) describe research on the effectiveness of the readability indexes produced by the former as indicators of the goodness of fit of the manual to the user population. He concludes that firstly the subjects used for the creation of the indexes were of a different population from the target population and hence are of dubious validity and secondly measures of readability at the work place do not correlate with the readability indexes, nor do they correlate with indexes of actual usage of the manuals. He argues that this lack of correlation is partly due to the influence of other factors affecting manual usage. This is another reason for following a different line of approach than this.

Technical writers tend to 'write by the seat of their pants', using design strategies that have worked in the past (Waller, 1979 and interview with staff at IBM). Their methods of coping with particular situations should be the subject of psychological enquiry, a basis for the development of an understanding of the area. Pirsig (1974) advances that it is impossible to rationaly describe what makes a good text, what gives it 'Quality'. However he was describing a subjective experience. Instructional text is designed to produce a particular outcome and its success at this can be measured. The text can be designed according to rigidly defined rules based on hypotheses drawn from theories of reading, memory and skill. Although the experimenter does not have access to the internal or cognitive variables such as the encoding processes and the learning outcome he is able to modify the external ones - the instructional text, and measure them - the user's learning behaviour and task performance.

The internal variables, the theories of learning and text processing, will be described below along with specific hypotheses about ways of presenting text in order to promote skill. The methodology for turning these hypotheses into experimental material will be described in Chapter 7, while the means of measuring behaviour will be described in Chapter 8.

The problem with attempting to apply much of the memory research to this area is that it either deals with small parts of the process under study (Eg. methods of increasing the probability of retention of a particular word from a list of unrelated words) or it refers to manipulations that are of little practical value. E.g. associating a noun with an unusual adjective will increase the probability of retention of that noun [1].

The main difficulty with theories of memory for text is that they treat text as an abstract stimulus. Though the meaning of individual words is considered in some experiments, the words are studied in isolation and not as means of imparting information.

As stated by Garnham (1981) the semantics of the text is not the sole determinant of what will be remembered; theories that relate the content of the text to the purpose for reading are more likely to predict what will be remembered. Not only must the reader remember the text, he must 'remember' those parts of the text that lead him exhibit skilled behaviour.

Elaborations should be seen as an aid to the cognition of the user, a means of improving their problem solving, rather than just a device to aid recall by strengthening the trace for the core material.

'A basic characteristic of the declarative system is that it does not require one to know how the knowledge will be used in order to store it. This means that we can easily get relevant knowledge into our system but that considerable effort may have to be expended when it comes time to convert this knowledge to behaviour.' (Anderson, 1981 p206)

---

[1] The findings that are applicable to this area are discussed in the literature review, though specific instances that apply to the forms of text that are discussed at the end of this chapter will be used as arguments for the choices described.

Not only must the reader remember the text, he must 'remember' those parts of the text that lead him exhibit skilled behaviour. The semantics of the text is not the sole determinant of what will be remembered; theories that relate the content of the text to the purpose for reading are more likely to predict what will be remembered (Garnham 1981).

The manipulation of the text must therefore be applied to those parts of the text necessary for this - either the method of elaboration must entail a way of selecting these parts or a separate theory must be applied.

A more specific literature than 'memory research literature' needs to be described. Its subject should be the manipulation of coherent pieces of text that are to be used for a particular function. Two areas are described by this label - the study of elaboration and the study of mathemagenics. Mathemagenics are memory promoting devices commonly used in instructional text to promote learning. Elaborations are also memory enhancing devices but these arise from the memory, reading and skill literature instead of from an educational technology school which was the source of mathemagenic techniques. (Reder,Charney, & Morgan, 1986 p64) define elaborations as '... any information that supports, clarifies, or further specifies the main points of a text.'. The development of elaboration research will be discussed before describing specific methods applied in this thesis.

Elaboration Research

Elaboration research has mirrored the development in reading research from studying text as isolated experimental stimuli to be read because the experimenter says so, for comprehension or for memory, to text that has a function and is read in order to aid the subject perform the task described in the text. There has also been a progression from studying elaborations designed to aid the retention of isolated words and phrases to elaborations that should aid the retention of information and promote skill.

Fisher & Craik (1980a) found that complex, elaborate encoding produces a richer trace, but that this richness can be used to enhance recognition only when the test conditions permit a reinstatement of the original encoding text. Although such elaborations may enhance memory for text the method is unwieldy and enhanced recognition alone is of little use as an aid to skill.

Stevenson (1981) found that transitive sentences had a higher recall than intransitive forms; she concluded that this supports Craik & Lockhart (1972)'s argument about the effect of the depth of processing on recall. This experiment is a move away from the memory research that examined words in isolated in pairs, however the text studied by the subjects has no function for them and consists of disconnected sentences.

Stein, Littlefield, Bransford & Persampieri (1984) studied elaboration as an aid to knowledge acquisition. They designed material to simulate an unfamiliar domain - a number of facts that are apparently unrelated. Elaborations were designed to increase the distinctiveness of the memory trace, increase the distinctiveness while increasing the associative relatedness of the concepts or reduce the arbitrariness amongst the key concepts. Again these elaborations consist of decorative details that are not supposed to add to the function of the text - as far as the subject is concerned the contents are arbitrary. However the experimenter is attempting to simulate a domain of knowledge rather than isolated words. This is reflected in the overall conclusion, that the best elaborations are those that highlight the significance of relationships between the facts.

Allwood,Wikstrom, & Reder (1982) studied elaborations in situ - within a body of text. The found that in a free recall paradigm far more text was recalled from a summary of a text than the full elaborated version. They found this result perplexing because authors have always added details to their work, supposedly to make the text 'better'. This last qualifier is the key to the problem - as will be shown later - Allwood, Wikstrom, and Reder (1982) measured the ***recall*** of text but authors traditionally write texts to improve ***skill***.

Reder & Anderson (1982) attempted to find out what it was about the summaries that improved recall - the spaced practice or the concentration on the important parts of the text. They found that both had significant positive effects on recall and hypothesized that the function of details is to increase the credibility of the text and to increase interest - poor embellishments of the text will however cause detrimental effects on performance, particularly on the recall of facts.

In the same theme Carroll (1984) designed a minimal manual in which as much as possible of the non-essential text was deleted. In comparison to the full manual this version was found to be the more effective training medium. Although Carroll (1984) also embellished the text with material designed to help the reader overcome difficult learning problems it was the skeletal nature of the text that was chosen for further study.

Carroll (1987) went on to examine what type of minimal instruction manual was most effective as a training device. He studied skeletal, inferential (designed to encourage the reader to draw inferences form the text), rehearsal and lengthy versions. Of these the inferential manual proved the best way of training users.

Black, Carroll & McGuigan (1987) studied the effect of different types of elaboration, again with respect to performance aided by a skeletal manual. Instead of syntax versus concept elaborations they chose a full version of the manual, an inferential version (encouraging readers to infer information rather than directly read it) and a rehearsal version in which readers are given the opportunity to rehearse explicitly stated information. The task this time was to use a word processor to carry out word processing tasks - simple learning tests (one command recall), command sequence learning tests and realistic task tests. The time to carry out the training increased from inferential version -> skeletal -> rehearsal -> lengthy. In the simple learning tests the time taken to complete the test was inferential < skeletal < rehearsal < lengthy though there was little difference in time between the groups. In the tests of learning command sequences and performing more realistic tasks the order was all of the tests inferential < skeletal < rehearsal < lengthy. The lengthy manual consistently produced worse performance than the skeletal version, showing that a fairly mechanical reduction of the text can be advantageous and that the introduction of specific elaborations can enhance the effectiveness of the manual. As would be expected, the rehearsal manual produced the best performance on the simple recall of commands tests - rehearsal is an established means of strengthening declarative memory for items. It was suggested that the effectiveness of the inferential manual, particularly on the realistic tasks, was due to the mental model of the system generated by the subjects who used it. A number of types of inference were used in this study; subsequently Black, Bechtold, Mitrani and Carroll (1989) investigated the effects of the different types of inference separately.

In this case the experimental task consisted of theconstruction of a computer database and instead of the instruction being manual based it took the form of electronic tutorials. The tutorials developed by Black, Bechtold, Mitrani and Carroll (1989) encouraged general to specific inferences (GS), explanation to specific inferences (ES), specific to specific inferences (SS) and explicit instructions (E). Another difference was that subjects were not allowed to make incorrect responses to questions and a count of errors was made. The order of time to complete the test was GS < SS < ES < E and for errors SS > E > ES > GS. In the discussion of their results Black, Bechtold, Mitrani and Carroll (1989) did not suggest why inferences should be effective in aiding learning. However they did point out that the explicit condition led to a relatively short learning time when compared to the earlier results of Black, Carroll and McGuigan (1987). This was due to the way that the tutorial provided the answer to any questions on demand. They suggested that the explanation condition provided more material to be read and this led to longer reading times. This conceptual information may not have been required for such short term learning, hence the poor performance of subjects in this group. Although the specific inference condition did lead to the formation of inferences (the subjects proceeded rapidly through the test) but the inferences were too specific or incorrect to enable them to perform well.

Carroll's elaborations were not derived from a theory of skill or reading but from a series of experimental results. Post hoc suggestions were made to explain the differences between the conditions but not to explain the mechanism behind the advantage of inferences.

Reder, Charney and Morgan (1986) proposed that although minimal text aid recall, elaborations to a text are necessary in order to make sense of situations, to provide a context for the use of information and to aid problem solving in new situations. This theory was supported by testing the merits of of a normal manual versus a manual with examples, analogies, metastatements and definitions deleted. The task was to carry out exercises testing their knowledge of DOS operating commands after spending time reading and reviewing the manual. The elaborated form produced the better performance on these tests of skill. A subsequent experiment comparing manuals where syntax verses concepts are elaborated showed that it was the details of syntax that aided skilled performance.

Reder et al. (1986) have shown that although summary text is best for enhancing recall elaborated text produces superior performance on tasks that require skill.

The definition of skill provided earlier in this thesis outlined different types of tasks. In some the user needs only recall the appropriate keystrokes, others may be minor variations of previous tasks and some may be very different to those that he has encountered before. Computing skill requires a mixture of different psychological functions - from recall to problem solving - one particular form of elaboration may encourage one (e.g. recall) to the detriment of another (e.g. the ability to transfer knowledge to a new situation).

The aim of this thesis is to explore the differential effect of elaborations on the user's ability to perform different types of task from a theoretical position rather than a pragmatic approach. It has been shown that minimality can improve recall and that different types of elaboration have different effects on skilled performance. The model of skill will be used as a basis for selecting degrees and types of elaborations to a basic text.

Functionally there are three basic types of additions that can be made to text. It can be designed to enhance the recall of the information already present. Or the text can be supplemented with information designed to clarify the information already present by providing examples of its use and sub-goals to be attained. Finally, a series of reasons for the operations can be added, in effect general mapping relations, which should allow the user to apply the core text to a wide range of problems.

The first group will be called ***linguistic*** elaborations because they stem from work on single phrases and sentences. These devices include the use of adjectives (Craik & Tulving, 1975), highlighting (Coles & Foster, 1975), sentence structure and complexity (Fisher & Craik, 1980b), and devices designed to slow the reading of text or make it more difficult to access text.

The second and third group are ***content*** elaborations. This includes the addition of examples, explanations and general rules about the information in the core text. These stem from analyses of the function of the text and theories of skill.

Another group of manipulations, ***mathemagenics*** - which were mentioned earlier in the chapter, have also been added to text in order to enhance learning. These include advance organizers (Mayer & Bromage, 1979), review questions and summaries (Kalt & Barrett, 1973). The experiment by Mayer (1980) is an example of this type; it examined the effect of advance organizer information, concrete elaborations provided by the reader and comparison elaborations, again produced by the reader. Although mathemagenics have been found to be useful they will not be studied in this thesis because they are substantial additions that casual users are likely to skip when using a manual. Also they do not fit in with the paradigm used here - in the literature there is a wealth of operational definitions for each of them and their use involves the addition of extra information that cannot easily be defined and linked to a core text. The approach here is to rigidly define the manipulations of the text and to limit the information added to the text. The next section describes content elaboration in more detail.

Skill and Elaborations

A hierarchical task analysis of the use of a computer application program produces a list of keystrokes to match each task that the user may meet. The analysis is represented by the analyst as a tree structure formed of several sub goals and associated plans. This organization is for the convenience of the analyst so that the sequence of keystrokes for any particular task can be compiled from the analysis rather than a situation where a list of keystrokes for every task must be stored in a laborious serial document.

The model of skill describes the way that users can construct responses. They may recall key sequences, construct responses from already known sequences of sub goals or formulate a response by iteratively solving sub-goals. Their ability to perform tasks and the method that they employ depends upon their knowledge of the skill.

Strictly, an elaboration is any text added to the list of keystrokes for a particular task. The least elaborated text is the HTA with the labels of the operations removed. This form still adds a sequence of decisions for the user to make about the path to take through the HTA in order to compile a list of keystrokes; even this sequence may provide the user with sub-goals to attain and extra information about the package. The alternative to this is to produce a help system that supplies keystrokes to match the user's description of their task - this demands a sophisticated interpretation that is not the subject of this thesis.

The first form of text to be used for training will be the unelaborated HTA. This will serve both as a control and as a text that may promote recall of the appropriate keystrokes due to the opportunity that the user has for concentrated attention on the basic matter of the task without distraction from elaborations (Anderson and Reder 1980, 1982; Reder, Charney and Morgan, 1986).

The next level of elaboration is to add the names of the sub-goals that the analyst created when describing the task. These labels supply information through the metaphors implied by their use here and provide english-like tags for the user to remember instead of abstract symbols (See Argument in Chapter 5). The labels again depend on the analyst's personal style of creating a task description. This will be the basis for the second form of text.

When recall of facts is the measure of learning from a text, summaries of texts are found to produce less learning than the elaborated texts (Reder and Anderson 1980, 1982). This advantage holds over a number of retention periods and measures of declarative memory (Allwood et al., 1982). It occurs despite the commonly held assumption that the extra material will produce multiple retrieval routes for the facts and the possibility for reconstruction of the material. The poor recall from elaborated texts is explained by referring to the Total Time Law (see Cooper and Pantle, 1967) whereby retention is a function of the total time and attention spent on an item. Time spent looking at the elaborations leaves less time for attending to the facts which are to be recalled in the test. Therefore subjects in this condition may perform less well on pure recall tasks than those in the core condition.

However Reder, Charney and Morgan (1986) propose that although elaborations are detrimental to the recall of facts they will aid skilled performance because they help to clarify the use of commands, syntax and concepts. The clarifications help subjects to perform well in novel situations, thus displaying skilled use of the tool. Their experimental results supported this hypothesis. A further experiment showed that elaborations describing syntax rather than concepts were the cause of this effect.

Thus the plans in the HTA are elaborated in this manipulation because they are the descriptions of the syntax of the operation. 'Reasons for the plan' has been chosen as the elaboration because it attempts to show why the plan has been constructed - clarifying the use of the sub-plans ie the syntax. It is also a fairly concrete way of limiting the type of material that goes into the elaboration. This matches Conway & Kahney's (1987) idea of a mapping relation and current educational practice (Anderson, 1987); when combined with the formal definition of the actions for performing the task in the core text the user should be able to perform all three types of task better than the subjects in the other conditions. This difference should be particularly noticeable on the more difficult tasks - those least related to their original training.

It should also help them to remember the plan because its parts are now described by links instead of unrelated pieces.This will be the basis of the third text.

The next elaboration is based on Kintsch & van Dijk (1978) theory of the assimilation of parts of the text into a coherent memory for that text. They suggest that if important items in the text are placed at the start where they can act as memory anchors for a number of propositions and can be processed several times in the course of this addition, these important items will be retained well. If however they are placed at the end of the text they may be linked to one or two items still in working memory so there is less chance of them being retained because they will not be re-processed. This manipulation puts the general rules (the specific description of the sub-task) at the start of the 'passage of text' on that sub task, rather than at the end. The general information (the summarizing statement) is therefore more likely to be recalled in this condition than in the next condition (5).

The use of examples of the plans as a form of elaboration allows a comparison with the elaboration that uses the reasons for the plans. This comparison is derived from the papers by Anderson (1982, 1987) in which he describes the ACT theory of skill acquisition. He demonstrates the usefulness of training based on providing the learner with examples of how to solve problems rather than giving them a description of why they need to solve the problem in a particular manner. Black, Bechtold, Mitrani and Carroll (1987) also show the advantage of providing instructional material that leads them to solve problems by inference from general examples as opposed to using reasons describing how the system works. In both cases the experimenters provided the users with general templates for solving problems which the the users were to use on specific examples. In this experiment all of the subjects are given this general template (this is the core text) - in this manipulation the difference between reasons and examples is supported by the addition of the elaborations to the general rules. To cut out the general rules and replace them with examples and reasons instead of adding these would have produced manipulations that were very different in content to the other texts - instead of the slight variations used here. This would have led to problems in explaining differences in the results because of the large number of differences between the texts. It may be useful to replace the plans with either reasons or examples in a future experiment. However, this experiment is looking at the effect of additions to the core text, not replacements.

This manipulation allows a comparison with Conway and Kaheny's (1987) proposal that a deeper knowledge of the mechanism behind a solution is needed rather than just trying to use the surface features of problems to produce a solution by analogy. In my opinion Anderson's method will work well for isomorphic problems but not the more difficult. The relative success of the different methods depends upon how one classifies the problems.

This is the core text elaborated by examples of the operations which are placed after the abstract description of the operations and plans. The example is intended to show an appropriate situation for the use of that operation (a real task) and an example of the syntax. It will be the basis of the fourth text. Ross (1984) demonstrated how examples help subjects to choose the correct command in a test situation. Pepper (1981) found that students attended closely to examples in texts; though Conway and Kaheney (1987) showed that students must recall the appropriate rule and correctly and apply the example to the current situation; this use of analogy may be a more common means of problem solving - rather than a deep understanding of the area (Anderson 1987). This dovetails with Fisher and Craik (1980) who found that complex elaborate encoding produces a richer trace, but this richness can be used to enhance recognition only when the test conditions permit a reinstatement of the original encoding context.

Elaborations to the syntax of operations has already been justified in (3) above. In this elaboration instead of reasons for the plans, an example of the plans is given.

There are two main reasons for this manipulation. Firstly it switches the position of the general rule to a later place in the text. According to Kintsch and van Dijk's (1978) model this should decrease the likelihood of that rule being remembered. This is in contrast to the previous manipulation where the example was given last and the general rule first. Secondly, it allows a comparison with the use of explanations as an elaboration, as described in the second manipulation though in this case the different types of elaborations are both in the same position in the text.

This is exactly the same as operation (4) described above - examples of syntax - with the exception that the example is placed ***before*** the re-description of the operation, ie at the start of the listing of all of its sub-parts. This changes the position of the 'general rule' (formal description of the task) to ***after*** the example - thus the fifth text.

The sixth text is a control text. The studies of the effect of elaboration on skill learning and Anderson and Reder's work (1980, 1982) compare the results of their elaborations against performance using a 'standard' un-edited text. This study attempts to control the production of such a text - to use the commercial manual for the the target package as a control would be to introduce an undefined text; this would detract from the effort to define the stimuli used in the experiment. The main quality of the 'standard' text referred to in previous experiments is the bulk of information provided along with the core HTA. The standard text here will therefore be made up of all of the elaborations described above.

Zipf (1965) pointed out that people prefer to make the least amount of effort necessary in order to complete a task. Eason (1982) and Cuff (1980) have characterized casual users as people who want to put the least amount of effort possible into learning a computer system - they would prefer to get on with their work. It is hoped that by forcing users to go through chapter after chapter to get down to the minuscule details of the task they will try to guess, to infer what the sub plans of a task are. This forcing of cognitive effort on them should increase both the likelihood of details being remembered when they go to the trouble of finding them and should also increase the number of occasions where they try to work out what the details should be - thus improving their ability to form plans and remember commands. There is obviously a trade off between extra effort that will increase learning and too much perceived effort which will make them decide not to use the manual. It is not known yet how the manual will be viewed by the users.

It is likely that if the mechanism for increasing work load to find information has been balanced correctly these users will out perform those on the simple core text on recall questions; they may be slightly better than the core group at the intermediate problems because they know more of the procedures for performing tasks.

This method of increasing the cost of accessing help rather than attempting to solve the problem unaided will be applied to all of the elaborations in order to search for interactions between the cost of accessing help and the type of help provided.

Summary

This chapter has applied the memory and reading literature to the models of manual use and skill to produce a number of hypotheses about the way to present text as a training aid for causal users. These are based on an examination of the memory, reading and skill acquisition literature; in particular the studies that treat text as a tool to be used and measures of its effect being based on test of skill rather than recall. This form of experiment matches the way that manuals are used by casual users in the work setting.

Much of the relevant literature was oriented towards aspects of text that could not be practically altered in a working manual - for instance the structuring of sentences and use of unusual adjectives. Some of the methods of enhancing text were inappropriate for casual users because they involve extra work - e.g. supplementary questions, advance organisers.

Elaborations to text were hypothesized to be the most appropriate method of improving manual for casual users because of the successful application of them in previous studies, the fit between the explanations of their effectiveness and the model of skill used in this thesis and the possibility of practically applying them to manuals. Predictions of the effect of the elaborations on skill acquisition were developed.

The next chapter will describe the method of analyzing a tasks as basis for a method of embodying the hypotheses about training in texts. It will also describe categories of task designed to test the user's skill with a view to the development of an experimental design to test the hypotheses in the following chapter.

CHAPTER 6

THE PRODUCTION OF ELABORATED TEXTS AND TESTS OF SKILL

Introduction

The purpose of this chapter is to describe the production of the experimental materials for a laboratory study. It includes a description of the process and problems of analyzing the task to produce a HTA and a discussion of the particular problems encountered when analyzing the task for the purpose of this thesis. The necessity of creating a precise description of the method for constructing the core text and elaborations is then asserted followed by a description of the techniques of producing elaborations.

The latter part of the chapter describes the design of the tests of skill. The rationale for the particular elaborations used in this study was described in Chapter 5; the selection of tasks to illustrate the differential effect of the elaborations on skill is described here. The basis for this is the use of the knowledge description within the HTA to produce four different categories of tasks - the user's ability to complete these tasks illustrates different types of skill. The tasks are 1) the initial learning task, 2) a structurally similar task, 3) a structurally dissimilar task that is derived from shared knowledge with the initial task, and 4) a task that is unlike the initial task.

This is followed by a description of the measures used to assess both the process of learning to use the package and the performance of the tasks.

Hierarchical task analysis of the electronic mail package 'mailx'

The use of the electronic mail package 'mailx' (mailx Revision 62.1, Hewlett-Packard Company) was chosen as the task to be analyzed for the following reasons:

1. The use of the system functions was tightly constrained. Unlike more flexible systems, such as word proçessors, the commands can only be combined in a limited number of ways[1]. This reduces the number of tasks that the user can perform with the package and so reduces the size of the task analysis to manageable proportions for this project. Although the training and testing will only cover a proportion of the package's use it was felt that a full analysis of the task should be made so that a number of different tasks could be selected for the purposes of training or testing. This would remove the need for a re-analysis of the package use in order to accommodate the new tasks in the organization of the HTA.

2. Electronic mail useage and access to the University computers was heavily promoted by the University Computer Centre at the time of conducting of the experiment. This meant that each year the new intake would be likely to want to learn to use the package and would be likely to continue to use it throughout the year, thus making them suitable subjects for a field trial because the software was not site specific and was actively used by a large proportion of the members of the University. Because the package was not used to perform academic tasks, tasks that would determine the student's success on his course, it was decided that it would be ethical to experiment with the subject's learning of this package.

3. Learning to use the package was intrinsically attractive to a wide range of subjects; this was important because it aided the recruitment of subjects for experimentation. While teaching the use of electronic mail in a group situation it was observed that as soon as a pupil had learnt a task he would enthusiastically practice it by communicating with other members of the group. It was also appealing to the subjects because it was a skill that would be useful to them in the future, outside of the experimental context.

The reason for using HTA to analyze the task was described in Chapter 5. The process of conducting a hierarchical task analysis has

---

[1] In fact it is the contents of the mail messages, the text sent by the user, that produces the wide range of tasks that can be performed rather than the number of commands available within the package.

been described in several sources - for instance Shepherd (1976). The use of HTA for analyzing Information Technology tasks has been described by Shepherd (Shepherd, 1987, 1989). The particular problems encountered when using this method to analyze the use of the program 'mailx' are described below.

1. The analysis of tasks which required 'hidden antecedents' or produced side effects.

Once a user has finished reading a new item of mail the system marks the item for deletion from the system mail folder to the user's personal mailbox; the mail is deleted when the user leaves the mail system. This feature frequently confuses new users because the system does not indicate that the action will take place by a verbal message but by removing the 'new mail' flag from the message. When the user next reads his system mailbox he will find that the mail item has been deleted and that there is no indication of where it has gone to.

This side effect was overcome in the task analysis by adding to the plan for reading the document the option to use the 'hold' command that preserves mail that has been read in the system mailbox. The loss of read mail when leaving the system was pointed out in the plan for leaving mailx; the user is given the option of leaving using the exit command which leaves the contents of the system mailbox in its original state.

Examples of a 'hidden antecedent' are the addition of a signature to a message and the addition of a message from a particular folder to a new message. In the case of the former the signature must have been set up in a file before the mail system is entered. In the case of the latter the user must have opened the correct mail folder before starting to write the new message and attempting to insert a message from that folder.

These are overcome by warning the user of the antecedent at a level in the hierarchy that allows them to fulfil the pre-conditions without having to backtrack.

2. The problem of describing the performance of functionally dissimilar tasks that contained similar structures.

These include the ways of identifying different types of mail and the different methods of specifying the contents of a message. The solution was to configure the actions so that as many as was possible of the similar structures could be retained in the analysis while keeping the different functions in the same place as other similar functions in the hierarchy.

This is just an extreme case of the training strategy of making the hierarchy as compact as possible.

3. Structuring the analysis so that concepts did not need to be explained.

An example of this is the concept of a folder. Conventional email manuals explain the concept according to the way that the mail is stored, the way that it is retrieved and the differences between a folder and a mailbox. In this analysis the difference is maintained through the way that the two objects are used in tasks.

4. Limiting the amount of knowledge the users needed to have of the use of a text editor.

It is possible to add text to a message by using a screen editor. Learning to use this tool would be a major task in itself. Instead the user is instructed to add text line by line and to correct errors by the use of the erase key which is assumed as part of normal keyboard knowledge.

The same applies to the setting up of an alias or signature - in these cases the precise operations needed are described, avoiding the necessity to describe all of the functions of the editor.

5. Problems associated with the methodology of this experiment.

A feature of the style of analysis used in this experiment is the effort made to point out as many common tasks as possible. Doing so has meant that the numbering of tasks is not as linear as normal - task 4.0 may appear between 1.2 and 1.3. This is because commonly used actions or sub-tasks are labeled just once but may be referred to in several tasks. The different shapes used in the diagram are designed to point out the common tasks.

In order to aid the selection of tasks the use of email is presented graphically in full. However, the HTA was not compiled into text form until the tasks had been selected then only those areas of the analysis that were to be used were converted into a core text. The knowledge analysis (see description below) was used as a basis for the selection of some tasks for use in testing; this was not conducted formally for the whole task but by selectively analyzing those tasks that the analyst knew involved common knowledge.

Two programs were used to aid the analysis of the task. The first, Diagram maker, aided the analysis by automatically producing a neat tree structure from a text describing the task. The second, Microsoft Word 3.0, included an outliner that allowed successive levels of the hierarchy to be displayed, edited, then hidden again so that the analyst is presented with only the parts of the analysis that are to be changed and not the whole document.

The necessity of creating a precise description of the method for constructing the core text and elaborations

The construction of a base text

As stated in Chapter 4, the basic task description consists of the names of the user's tasks to be performed and the key strokes to perform those tasks. It was argued that this is the basis for forming a training program, any addition to it is a training technique. The methodology of this thesis is to construct a core text from the basic task description, then to add elaborations to it based on psychological theories in order to produce texts for experimentation. The argument presented in Chapter 5 was based on a model of skill and the use of text; the argument below is a methodological one, concluding that the method of producing the core text, elaborations and tasks should be clearly defined in order to isolate and control the manipulations of the training method.

For instance Sasse,Johnson, & Briggs, studied the effect of different training materials on performance; however the methods of training did not clearly isolate the 'procedural' and 'conceptual' factors - the conceptual group received a hierarchical tree on paper while the procedural group were individually taught by the experimenter. This lack of control over the presentation of the training hypotheses within text weakens the conclusions that they drew.

Consequences of not clearly specifying a base text

Any text should be treated as an artifact, an embodiment of a number of psychological and/or design principles. For instance, there may be variations in both presentation and content within and between consecutive issues of the same commercial manual for the same version of an application program. Carroll & Kellog's (1989) acceptance of the multiplicity of design issues embodied in one artifact may be an appropriate strategy for design 'from the hip', but an experimental evaluation of particular issues and the psychological principles behind those techniques must aim to control as many extraneous factors within the design as possible. So, just as one would not try and compare different type-faces for effect on reading speed when the basic texts they are embodied in are not controlled, so it is a mistake to compare the effect of elaborations on skill learning when the base text for the elaborations is not controlled and defined, the method of producing the elaborations is not reproduceable, and the tests of skill are not meticulously derived from a model of skill. Schumacher & Waller (1985) point out that although researchers are expected to apply rigourous standards of statistical analysis to their data, their stimulus materials are rarely subjected to the same scrutiny. This thesis aims to redress the balance by constructing a tight methodology for constructing texts.

Necessary qualities of a base text

The text must completely describe the task to be performed so that any instructions that need to be given can be founded on that base. It must be reliably produced from a task; any worker must be able to carry out the method of producing a base description of the task and the final result must be capable of replication by different workers. The base should be as language free as possible, tasks should be broken down into logically linked verb-object statements. One should not have to describe the task in terms of examples or explanations of concepts as these are forms of elaboration on the more fundamental verb-object pairs. Similarly the base text should not consist of named sub-goals.

Methods of specifying base text in the literature

The experiments in the literature examining the effect of elaboration have different theoretical focuses. They can be classed as:

- •experiments that examine sentence features and the retention of words.E.g. looking at the complexity of sentences or combinations of adjectives (Fisher & Craik (1980); Stein, Littlefield, Bransford, & Persampieri (1984)),
- •information elaboration e.g. forcing users to make inferences or adding information - Reder & Anderson (1982), Black,Carroll, & McGuigan (1987),
- •and training types that use devices such as advance organizers or analogies (e.g. Mayer, 1979)).

Each type has its good and bad points concerning how the text is specified. Except for the word-learning experiments, none of them formally specify a base text that is rigidly used as a basis for elaboration - most take a text then re-write, but apparently without precisely defining a method of re-writing, just a style and without starting from a base text that is derived from an analysis of the skill. Word experiments take single base sentences then modify them according to strict rules; however, in the experiments listed above the theoretical basis for those rules does not appear to be very exact. Similarly, the training experiments use the same basic texts and add devices to them, but the devices are not grounded in a precise theory or method - they are modelled on popular conceptions of the original research.

Reder & Anderson (1982) isolate main points from a chapter on a topic to create their base text, then embellish the core with details from the same chapter. Their tests consist of true-false questions and short-answer questions; the method of producing them is not described.

Reder, Charney, & Morgan (1986) use an official IBM manual as their core text and created an unelaborated version by deleting *portions* of the manual. The tests of skill were actions to be performed with the computer, but the basis for selecting the actions was not described.

Fisher & Craik (1980) chose sentences to represent an unfamiliar domain but did not describe the method of selection. Their elaborations on the sentence structure were specific and the recognition test had a theoretical basis. The same comment applies to Stein et al. (1984).

The comparison elaborations in Mayer (1980) were tightly defined but the use of advance organizers and analogies was as loose as is typical of these experiments - text is added that is supposed to correspond to Rothkop's description of organizers but is actually created without reference to any definition of an organizer. The presentation of the material and the base text (1 page of descriptions of format, appropriate time of use, example and description of example for each command) was very clear, though not based on an analysis of the task.

Black, Carroll and McGuigan (1987) used brief descriptions and elaborations to make clear their methods of re-writing the manual for a word processor. It would be difficult to reproduce the same texts given these instructions. The same comment applies to the elaborations used by Black, Bechtold, Mitrani & Carroll (1989) who do not base their elaborations on a theory of skill.

With such loose descriptions it is difficult to be sure that the different manuals contain the same basic information and that they differ only with respect to the specified variable. When comparing experiments by different authors it is difficult to tell exactly how an enhanced manual in one study differs from that in another. The selection of tests without reference to a theory of skill or adherence to a basic transfer paradigm makes the effect that choice of task has on the effect of the manual type unclear.

What is missing is both a description of what the text is supposed to be transmitting from author to reader, in elaboration free terms, and a description of the method of embodying that message in text - the method of elaboration. The nature of the medium will determine the message that is received. Unless both the medium and the message are tightly specified, experiments in this area will be nigh on impossible to exactly reproduce, the basis for these effects will not be understood because the causes are not sufficiently described, and the measures of retention and skill that are purported to be brought about will be without a criteria for comparison, for the experimenter will not have described what is there to be learnt.

Production of the Core Text

The HTA produces a description of task goals and a hierarchy of sub-goals, each of which has a plan to describe the performance of subsequent sub-goals and finally a set of operations (in this case key strokes) to be performed. The core text consists of the task goals to be performed, and the sequence of plans leading to the key strokes to be performed. Wherever possible the names of sub-goals and operations are omitted from the core.

The system of indentation (use of Levels in Word) and numbering used in the illustration of the HTA are not altered. Paging set is according to the default paging on Word. Numbers of sub-goals should be typed in bold; plans should be on new lines and marked by a bullet. Plans should be at the same indentation as their sub-goals. New operations start on new lines, plans wrap around onto new lines. The text is not broken up by blank lines.

Elaborations

1. Core condition

In this condition the goals and names of operations are withheld from the user - they are only supplied with the key strokes to be performed. This is the core text taken to the extreme - all elaborations and goals have been removed, so the bare facts should be recalled most easily because of the concentrated effort that the subject can apply to memorising them.

2. HTA

This condition is the core text plus the names of operations and sub-goals. The operation names should be those provided by the software designer, the sub-goals should be named after the the work-tasks that the user wishes to perform. Sub-goals should be fully re-described beneath the superordinate goals - the user should not have to skip back and forth between operations; instead they should be able to read through the goals sequentially.

The typography and presentation is identical to that of the core text. Names of operations are placed before plans, on a separate line. The names should be immediately after the number denoting the sub-goal.

3. Additional reasons

This is exactly the same as the control version except reasons for the plans have been added. The reasons for the plans are added immediately after the plan - at the same 'Level' in the text and headed by:

•Reasons

just as plans are headed. Strictly speaking, describing reasons can be a limitless job. In this analysis the reasons have been taken as far as the experimenter thinks is necessary in order to make the plan of the task obvious to the subjects. The reasons are given as concisely as possible.

All other typography remains the same as that in the core text.

4. Examples of syntax (general -> specific)

This is the core text elaborated by examples of the operations which are placed after the abstract description of the operations and plans.

After each description of an operation an example of that operation is given. It is intended to show an appropriate situation for the use of that operation (a real task) and an example of the syntax. The example should be specific to the plan for this operation and should not include details of sub-plans in order not to detract from the elaboration of this particular piece of text. It should include the information contained in the plans and refer to work tasks. The examples should be brief - one or two sentences. They should be headed by :

•Example

as in the description of plans but should be at the same level as the operation name. All other typography remains the same as that in the core text.

5. Standard Text

This elaboration consists of the core text plus all of the elaborations described above. It is designed to simulate a 'normal' text written by a technical author - it includes a number of different types of elaboration that act as distractors from the core text but provide controlled supplementary information.

<u>Tests of skill</u>

The core text and the elaborations designed to promote skill have been described. What remains is to describe how to test the user's skill. The experimental paradigm is to present the user with a sequence of initial tasks to be learnt with the aid of the manual. Time to perform each task, number of tasks successfully completed and access to the manual will be recorded. After a delay the user will be presented with some tasks designed to test the skill that they have acquired from the initial learning - the amount of transfer. The same measures will be applied but the user will be presented with different types of task in order to gauge their skill at using email. These tasks will be presented in a sequence designed to eliminate order effects - the mixture of tasks, sequence, delay from initial learning and access to help are designed to produce an experimental situation that closely matches the normal work of the casual user. While the method of categorizing tasks is described below, at the end of the chapter the practical issues of selecting the particular tasks for the experiment are described. The selection of tasks starts with a consideration of skill.

Operational definition of skill

According to the model of skill described in this thesis, being skilled means being able to :

1. Carry out tasks that have been accomplished successfully before.

2. Carry out tasks that are closely related to those that the user has already performed before.

The two tasks should differ only with respect to minor parameters, say the last command in the sequence, the instantiation of a variable. One would expect the user to be able to carry out all of the actions in the second task, given that he has mastered the first, with the exception of the dissimilar part(s).

3. Carry out tasks that are significantly different from those that have been encountered so far but which are made up from component sub-tasks that have already been used.

As operational definitions these descriptions are vague and imprecise. Their inexactness is due to 3 key phrases, one used in each. In the first definition the phrase is 'accomplished successfully'. This is meant to indicate that the user is able to carry out the tasks because he has learned how to repeat the actions - not because he knows why the sub-tasks should be arranged, simply that he knows the actions to be performed and their order; he can mechanically recall them. The phrase also implies having learnt the task to a criteria level of success. This level depends upon the nature of the experimental work being carried out and will be discussed later. In the second definition the phrase used is 'closely related'. This is meant to indicate that the user knows something more about the sub-tasks than just the key strokes and is able to carry out variations in their operations according to the rules of their syntax.

The final phrase, 'significantly different', in the third definition stems from the attempt to make the definition cover those activities that indicate a knowledge of the potential uses and interactions between the sub-tasks; although the tasks contain the same basic actions as those described in the previous category, their combination and instantiation should reveal a wider understanding of their potential.

The difficulty in honing the definitions into rules for the selection of test items lies in the need to refer to what the user must 'know' in order to be capable of performing the tasks. This can be overcome by using the information in the task analysis to plot out what a user ought to be able to do if they are able to perform a given task or group of tasks.

If a subject is able to carry out a task to a given criteria level of performance, without reference to external memory aids, one can say that they ought to know some of the information described within the task analysis describing the/those task(s). At the very least they know which key strokes to press; they may also know the variations possible in the sub-tasks and perhaps the concepts behind the selection of sub-tasks. The presence of this knowledge, or rather the ability to use it, can be tested by observing the subject's ability to perform tasks that the experimenter predicts they ought to be able to manage given the potential knowledge indicated by their earlier performance.

The task analysis describes three divisions of information; the key strokes to perform a task, the plans directing the selection of the key strokes and the reasons for the plans. The divisions closely match the different types of knowledge the user would require to perform the different types of task. Since this information has been formally compiled as part of the task analysis it is a ready made basis for selecting the categories of task to be used for measuring skill. An example is described below.

Suppose that a person is able to make the plain characters in a paragraph of text into italics using the 'Word' word processor.

Being able to do this means being able to highlight the whole paragraph by clicking the mouse and being able to change the format of highlighted characters by using 'ALT I'. The subjects will not be able to produce a different type of character format or highlight other sections of text unless they can guess successfully at different formatting commands using the ALT key and different ways of highlighting by using the mouse. They are very unlikely to be able to format by using the menu commands or highlight by using the function keys.

At the level above key stroke knowledge the action of highlighting a paragraph by dragging may indicate that they know that they can highlight letters, words, phrases, sentences and entire documents. They may also be aware that they can click on the menu to format characters, but will not necessarily know what they can do with the other menu items. Thus they may be able to highlight different areas of text, format them into different character types and may be able to muddle their way through other command line functions.

With a knowledge of the reasons behind the plans they may know about the centrality of the 'highlight/click command line' operation to the use of word, the use of 'Return' to signal the end of a command sequence, the use of right hand button clicking to jump down menus and the basic functions behind the other command line items. The user should be able to use most of the commands under the 'format' menu, and should be able to attempt use of other menu commands if they can work out the meaning of the menu names with respect to the work task. Thus they should be able to perform structurally dissimilar tasks that share common knowledge.

As well as illustrating the range of tasks users may be able to perform given success at one task this example also shows how the interface of a package can aid performance by supplying information about the systems' commands. Because the 'Word' screen provides clues about its operation, a user may be able to work out how to use many functions by considering their knowledge of the meaning of the menu names and their past experience with computers. So it is difficult to state with certainty that they will not be able to carry out tasks other than those specified. A more accurate statement is that it will be more difficult for them to carry out the other tasks without extra knowledge - they are likely to take longer and make more mistakes.

It has been shown that it is difficult to predict levels of transfer by comparing the production rules - Gick & Holyoak (1980), Dreyfus & Dreyfus (1986), Tetzlaff (1987). The same problem applies here - the psychological difficulty of learning new operations is not revealed by task analysis and can only be stated with certainty after testing. But since the same tasks will be used for each experimental group, so long as very easy and very difficult tasks are not selected this issue can be eliminated.

A worked example from the use of email is provided below. The particular task is that of jumping to the end of a document that the user is in the process of reading.

The analysis of the task of examining the remainder of a message that the user is reading is set out below.

3.0 examine the remainder of the letter
•Plan
Quitting (3.4) must always be done when all other operations have been completed as necessary. To re-display current page do 3.1. To move forward in the letter either move forward by set increments (3.2) or jump to the location of a particular piece of text (3.3).
•Reasons
The paging system is a separate system to the one in which mail items are moved between and processed, so it must be quit once the document has been paged through. One can move forward by set increments or jump forward and backward by searching the document for a piece of text.
3.1 re-display current page
•Plan
Do 3.11 then 3.12
•Reason
The command to re-display the current page is '.'. The computer will not act on a command until the user has signified that the command is complete by pressing Return.
3.11 type '.'
3.12 Press Return
3.2 move forward in letter
•Plan
Do Select the operation (move forward by one line, half a page, a page , two pages or to the end of the document) that will move you the required distance forward.
•Reason

You must decide whether to move forward by one line, half a page, a page, two pages or to the end of the document.
3.21 move one line forward
•Plan
Do 3.211 then 3.212
•Reason
The command to move forward by one line is 'l'. The computer will not act on a command until the user has signified that the command is complete by pressing Return.
3.211 type 'l'
3.212 Press Return
3.22 half a page forward
•Plan
Do 3.221 then 3.222
•Reason
The command to move forward by half a page is 'd'. The computer will not act on a command until the user has signified that the command is complete by pressing Return.
3.221 type 'd'
3.222 Press Return
3.23 one page forwards
•Reason
In most systems the stock default action (the action most commonly carried out) is indicated by pressing Return. In the paging system the default is to page one page further forward, so this is indicated by pressing Return.
3.231 Press Return
3.24 skip a page forward
•Plan
Do 3.241 then 3.242
•Reason
The command to skip a page forward is 'f'. The computer will not act on a command until the user has signified that the command is complete by pressing Return.
3.241 type 'f'
3.242 Press Return
3.25 move to end of document
•Plan
Do 3.251 then 3.252
•Reason
The command to move to the end of the document is '$'. The computer will not act on a command until the user has signified that the command is complete by pressing Return.
3.251 type '$'
3.252 Press Return
3.3 Search for text
•Plan
Choose most appropriate operation to move to the part you want to read. Carry out the operations in any order.
•Reason
The commands may be carried out in any order because you can search both forwards and backwards or search the entire document.
3.31 search forwards for **text**

•Plan
Do 3.311 then 3.312
•Reason
To indicate that the computer must search for a piece of text bracket that text by forward slashes '/'. The computer will not act on a command until the user has signified that the command is complete by pressing Return.
3.311 type '/**text**/'
3.312 Press Return
3.32 search backwards for **text**
•Plan
Do 3.321 then 3.322
•Reason
To indicate the that text must be searched backward for, bracket the text by question marks '?'. The computer will not act on a command until the user has signified that the command is complete by pressing Return.
3.321 type '?**text**?'
3.322 Press Return
3.33 search entire document for **text**
•Plan
Do 3.331 then 3.332
•Reason
The system will only search in one direction at a time. Therefore if you wish the entire document to be searched you must ensure that you are situated at one end of the document. Because there is only a facility to move to the end of the document and not the start you must start the search from the end of the operation and then search backwards.
3.331 move to end of document
•Plan do 3.3311 then 3.3312
•Reason
The command to move to the end of the text is '$'. The computer will not act on a command until the user has signified that the command is complete by pressing Return.
3.3311 type '$'
3.3312 Press Return
3.332 search backwards for **text**
•Plan do 3.3321 then 3.3322
•Reason
To indicate the that text must be searched backward for, bracket the text by question marks '?'. The computer will not act on a command until the user has signified that the command is complete by pressing Return.
3.3321 type '?**text**?'
3.3322 Press Return
3.4 quit examining letter
•Plan
Do 3.41 then 3.42
•Reason
The command to quit is 'q'. The computer will not act on a command until the user has signified that the command is complete by pressing Return.
3.41 type 'q' 3.42 Press Return

If a user is taught the key strokes to move to the end of the document, they will have learnt to press '$' (task 3.251) then press 'Return' (task 3.252).

If they are also taught the plan for this operation they will know plans 3.0, 3.2, 3.25 and 3.4 - this tells them that they can move through a document in a variety of ways, including in this case moving forwards, and that they must quit examining the document when they want to move on to another. Plan 3.2 should suggest that there are a variety of different ways of moving forwards. They will therefore be in a better position than the key stroke knowledge only group to attempt to quit the document before moving on to another and have some idea that they can move in other ways other than forwards and that there are different ways of moving in that direction.

If they are provided with the reasons behind the plans - 3.0, 3.2, 3.25 and 3.4 they should know that they can move in other directions through the message by searching the text, that they must press Return in order to finish issuing a command and that they are in a different system to that which handles the processing of mail. Thus they should be better able to search text, complete the use of commands and recognize where they have to change between parts of the mail program. These tasks are structurally dissimilar to the original task yet share common knowledge.

The categories of tasks listed above - initial tasks, similar tasks, common knowledge tasks - will be used to test the effectiveness of the text elaboration at imparting skill to the user. Performing the initial learning tasks and the test tasks may have a transfer effect on performance rather than due to the effect of the manipulation on the skill of using email. In order to measure this a fourth test will be added. This will have no similarity whatsoever to the email tasks. If the users perform well on this task despite not receiving specific training on it then the non-specific transfer must be the cause.

There is also the possibility that a particular form of elaboration may prove to be effective because of the interaction between the nature of the learning task and the training strategy. This effect will be examined by repeating the experiment but using a different application package.

Measures

A measure of how well the user performs the tasks is needed. This includes how long it takes to complete each task, how many tasks are attempted and completed, and a measure of the efficiency of their solutions - in this case a decision as to whether the least number of key strokes necessary to perform the tasks were used or not. In addition it would be interesting to know at what point users fail within tasks and what sort of help they request. These would be useful in a qualitative-post hoc analysis of the empirical results.

These measures describe the performance of the task. A second aspect of the analysis of the learning is an examination of the amount of help that users require. The progress from a high level of initial use of the manual to unaided performance is a good indication that learning has taken place. There should be an interaction between time to complete the task and help needed. It is likely that initially tasks will be performed slowly and the manual will be used for most tasks. If the manual is effective its use should drop off and the time to complete tasks should decrease. There may be more interesting interactions:

- users that spend the most time with the manual initially may eventually become the most skilled performers
- some users may choose to access the manual for every task, learning may not take place or the user may find it easier to use the manual than work it out for themselves
- users supplied with one set of elaborations may quickly learn to perform the initial task without the manual but may need to refer to it again when faced with the tests. A different group may need to keep referring to the manual for the majority of the initial tasks but may not need it for the test tasks.

Selection of test items

The test items should be selected from the HTA of the task according to the definition of skill as described earlier. They should be either tasks to be recalled, tasks for the close transfer test and tasks for the far transfer (problem solving) test.

The tasks should appeal to subjects - this will aid the recruitment of subjects and keep motivation high during the testing. However, the scope the user has for constructing an answer to the problems should be constrained as tightly as possible otherwise some users will receive more practice or construct more elaborate responses than others. For example, if users were told to 'send messages to each other' some would be zealous in their practice, others might skimp.

In order to make the experiment attractive and enjoyable it was originally planned to formulate a game in which email provided the means of communication. Candidates included pontoon, rummy, hang man, battleships and 'dungeons and dragons'. However none could be found that would meet the practical constraints described above. There was a second draw back to this approach in that such tasks are not a normal use of email - both the game and the email package would have to be operated in a distorted way.

It was decided that the tasks should be based on normal email useage but that the actual tasks to be performed should be prescribed by the experimenter rather than allowing users free reign. Prior experience of teaching people email suggests that learning to use this package is intrinsically motivating and the specification of exact tasks to be carried out allows the constraints described above to be adhered to. Email useage is broken down below :

Basic email tasks

sending mail
reading mail
composing the mail message
forwarding mail
saving mail

Areas of these tasks with potential for transfer

destination of the mail
method of paging through the mail

methods of addressing mail items
contents of the mail
  typed text
  imported files
  imported messages
  imported stripped messages

The actual tasks to be used in the experiment will be decided upon on the basis of the results of the pilot experiments. As stated earlier, it is difficult to predict ease of use of commands from a production rule analysis, or in this case examination of the task analysis.

The tasks are likely to be as follows.

<u>Initial tasks and tasks to be recalled</u>

Sending a typed mail message to a single user.
Reading a single mail item
Inserting a message into mail to be sent
Saving mail

<u>Close transfer tasks</u>

Variations on the destination of the mail, the means of addressing and paging through the mail, inserting a file into a message.

<u>Far transfer tasks</u>

Retrieving and saving information to and from folders
Searching mail for text
Combinations of close transfer tasks.

Summary

This chapter has described the task analysis of the email package to be used in the laboratory experiments. The need for precise definitions of the method of producing a core text and elaborations from the task analysis has been set out. The procedures for producing text used in this study have been described in terms of the definition of the elaborations to text.

The means of testing the skill imparted by the elaborations have been described - four categories of skill tests and a set of measures of performance. These include a record of manual usage and transcripts of their interaction with the computer system in order to assess how much the subjects need to resort to the manual - as an indication of the degree of learning - and to supplement the stark measure of time per task.

The practicalities of selecting tasks to be used in the learning stage of the experiment and to be incorporated into tests was described. It was decided to to use an electronic mail package as the task for the experiment; a list of sub tasks suitable for different tests of skill was created.

The next step is to describe the design and conduct of the experiment in order to reflect the nature of the applied problem and the model of learning.

CHAPTER 7

EVALUATION OF AN EXPERIMENTAL METHODOLOGY TO STUDY SKILL ACQUISITION FROM TEXT

Introduction

A number of manipulations of text and means of testing the efficacy of the texts as training mediums were proposed in the preceding chapter. The first section of this chapter will describe the presentation of the texts and tests to users; the second part will relate the practical implementation of the design and pilot tests of the experiment.

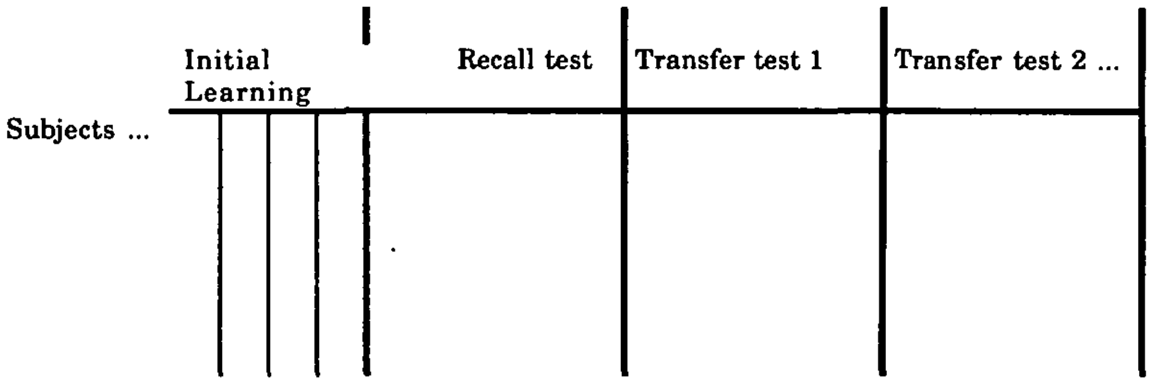
The final design of the experiment was reached after two main pilot studies; the first of these consisted of close examination of a video recordings of subjects carrying out the experiment. The second study was to have been the main data collection but observations that were made during this second set of studies and a preliminary examination of the results led to a re-evaluation of the design and the construction of the main study. The issues that arose during the pilots will be described at the close of this chapter as a prelude to the design of the major laboratory data collection.

Design rationale

Though many aspects of the study are typical of laboratory tests (e.g. tight control of the material to be learnt, the distractors encountered, the tasks to be carried out) efforts were made to replicate the sort of problems encountered by the casual user in a work setting. The guiding scenario was of a user who is presented with a series of simple tasks to be conducted in one session of interaction with the computer with the aid of a manual then later presented with the same tasks and some slightly different ones again with a manual available.

<u>Design of the study</u>

The process of using the manual to aid the execution of the first set of tasks is the initial training of the users. The subsequent tasks are the tests of recall and the ability to transfer to new situations. The design so far is as follows:



In this experiment concrete hypotheses have been proposed about the effect of different types of manual on the performance of the tests of skill. However there may be non-specific effects of the forms of presentation on computing skill. For this reason another type of test has been included in the battery of re-tests - a test of non-specific transfer. This test is designed to check whether one particular form of manual over another improves the person's ability to perform a task that bears some resemblance to the other tasks but which has such a tenuous link with the training material and training method that a specific effect is not expected. In concrete terms if the experimenter was examining people's ability to recall foreign words and the manipulations were designed to have effects on their ability to recall *particular* types of words a non-specific test would examine the subjects subsequent ability to recall lists other than those specific words. If one particular group's performance indicated that they were able to recall words *in general* better than other groups after the initial training then it is likely that this effect was a contributory cause of the group's performance on the more specific tests.

The difference between this test and the test of far transfer should be emphasised. The latter test was designed to be 'structurally dissimilar but sharing common knowledge' with the initial training task. The degree of similarity was specified and certain hypotheses about the performance of different groups on this test proposed. The non-specific test is related only in a very loose sense to the other tasks and performance is not expected to be improved by the training manipulations. Other than that the nature of the effects (degree, direction and cause) is unspecified.

In practice the definition of a far task does not work. It was intended to formalise the opinion of an expert at the tasks that if the subjects can carry out one task they ought to be able to carry out a second one because elements are the same even though the overall nature of the task is different. This intuitive judgement appears to rest on the person's understanding of the psychological similarities of the tasks rather than a logical analysis - i.e. not based on information revealed by a HTA. This difference is the one that makes it difficult to predict in advance the direction and degree of transfer between complex tasks - though there may be some similarities between the tasks other aspects of the tasks may overshadow the effects to an unknown degree.

A second problem with producing a tight workable definition of far transfer is that the user may not be aware of the tenuous similarities between the tasks and thus not be able to apply the common underlying knowledge. As an example consider the difference a user perceives after being trained to start the mail system and print an item between being told to print an item from the update folder and being told to open the mail system, change to the update folder then print an item. For these reasons it was decided not to use this test of learning.

In order to prevent an order effect the tests of skill in the re-test session was rotated. The main concern was that performing the first few tasks would lead to increased performance on the remaining tasks; thus if a group of tasks was presented last the performance would be to a high degree determined by the preceding actions. From the observation of computer users it appears that mixing the order of presentation of the different types of tasks more closely matches the situation faced by the casual user who is unlikely to encounter blocks of identical tasks.

In order to keep the remaining order effect constant for all of the different types of manual the same order of presentation should be used for each subject. The order effect could have been discarded altogether if sufficient subjects and time were available to run through all of the different combinations of test order but this was not practical.

The design so far examines the effect of the manual type on the learning of a particular package. This leaves the possibility that any effect is due to the interaction between some aspect of the package and the manual type. Ideally any text type that produces an effect should be applied to the design of a manual for a different package then re-tested with that package.

The design needs to be expanded in order to accommodate the easy versus hard versions of the manual. That is, manuals where it is easy to get the information manuals where there is a penalty for their use.

Introducing this variable doubles the subjects required for a full implementation of the design. This number can be reduced by carrying out the experiments where the underlying theory indicates that an effect is likely to occur first of all. A rational decision can then be made as to whether it is useful to carry out the rest of the design. For instance if it is found that making a form of manual that is designed to aid skill acquisition by aiding recall is not improved by making the information access more difficult (which is designed to improve recall) then there is no need to try the easy version of the manual.

A similar decision was made when perming the order of the tasks. Ideally each type of test should be one of the perms in the initial training task. However, the experimental hypotheses have been based around the presentation described above - a simple initial task followed by tasks with decreasing smaller extent of similarity. It was decided that the other alternatives, for instance training them on the non-specific test tasks first of all, would not be examined in this experiment because of the time and subjects necessary and because the transfer would not necessarily work in the opposite direction.

Construction of the manual and presentation of the tests

Rationale for using an electronic manual

It was decided to implement the manuals on a computer, the text to be displayed using a 'Hypercards' (Apple Computer, Inc) system. Although this created a scenario for the experiment very different from that which the user would have experienced using a paper based manual the benefits of doing so in this series of laboratory tests were felt to outweigh the disadvantages. The computer manual scenario entails a number of psychological features; some of these were directly observed during the experiment and will be described in the relevant sections of the experimental results, some are features already described in the literature which are likely to have a bearing on this particular experiment. For instance,it has been demonstrated that people prefer to read text from paper rather than the computer screen (Muter,Latremouille,Treurniet, & Beam, 1982).

Before comitting to an electronic presentation it was decided to examine the effect of different paper layouts of a core text based manual on ease of use and acceptability to the subjects.

Usability issues of a paper-based 'core text' manual.

Subjects were asked to locate areas of text that described how they were to perform a small number of short tasks within a core-text type manual. They were asked to describe any problems with the manual and their opinions of the manuals as a source of information both during and after the information searches.

Four subjects, undergraduate psychology students, were asked to use a variety of manuals in a walk through atmosphere. Although the manuals contained the same information their layout and tables of contents differed.

Each manual consisted of a full HTA of the use of the Unix electronic mail system. Tasks were allocated a different number and tab level within the page according to their position in the task hierarchy - See Chapter 7 for a description of this type of layout. In one version of the manual the tasks were laid out from top to bottom through the pages. In a second version only tasks of the same level in the hierarchy were presented on each page - e.g. a page would consist of tasks 1.1, 1.2 and 1.3 but not task 1.2.1. Tasks always included the numbering system that identified their position in the hierarchy. The first form of index showed the location of particular tasks in terms of the page where they could be found. The second index listed each task and its numbering label - this is a reiteration of the HTA without the plans or details of the tasks.

Some example tasks are listed below:-

How would you forward some mail that you had just read ?
While reading some mail, what is the command to display the next page in the letter ?
How would you add the contents of a message to the letter you are in the middle of writing ?
How do you delete a piece of mail ?

Subjects made the same comments, whatever the form of the manual or index:

1. They disliked needing to remember where each part of the task was and needing to navigate through the different pages. Typically they tried to leave a marker in each page that related to the task in hand.

2. They found the numbering system awkward - both because of the depth of the numbers (e.g. 1.2.3.2.1.1) and the lack of distinction between the numbers and the body text.

3. Subjects tried to flip through the manual rather than follow the numbering system in order to locate information. This proved to be very inefficient, because different pieces of information were not highlighted by section names. When the use of the numbering system was pointed out subjects were able to locate the target text with greater ease.

4. Some of the task names did not imply their sub tasks which made it difficult to locate the sub-tasks.

5. The method of following a plan by carrying out each of the sub-goals was not obvious to all subjects and required some explanation.

Many of these difficulties could have been overcome by turning the task statements into more verbose paragraphs and adding a system of text labels and highlighting in order to assist 'page-flipping' navigation. However this would have moved away from the controlled plain contents of core text back to a less controlled form.

One way to aid navigation through the tasks without using the numbering system or requiring subjects to keep markers in pages was to turn the core-text into hypertext. In such an electronic form controls could be added to the manual in order to aid the subject's movement from task to sub-task and vis a versa. This was one reason why an electronic manual was chosen as the means of text presentation. The second major reason, that of the facility to easily record which pages (and hence information) had been used is described late in this chapter.

When the experiment was first conceived it was believed that the method of training examined here could be studied independently of the medium of implementation of the text. By the end of this chapter it should be clear that this is not entirely true. Computers have been employed in psychological research in order to assist data collection, assist presentation of material and create a particular experimental setting. Aspects of this experiment were enhanced by employing a computer in the former two areas but because the presence of the computer was not opaque to the subject the intrusion into the latter area changed the milieu of the experiment.

The hypotheses of the experiment are focussed on the time that it takes users to perform different types of task. Recording the keystrokes with associated timings was one potential method of collecting both the time taken to carry out particular tasks and to keep a record of their actions for possible post hoc analysis of the data. Since the tasks were computer based it was decided to examine means of setting up the computer so that it produced a transcript of the interaction rather than recording each session of video then transcribing it by hand. Methods of doing this by inserting a device between the keyboard and the computer or by using software that recorded the interaction between the computer and the user were examined but the available solutions were found to be too expensive or produced far more information than was needed for this experiment. All that was really needed was a time stamp of the start and end of the task and a record of what the user had done in between - not a record of the timing of every key stroke. The macro listed below stamps the start and end of the use of the mail system and records the intervening interaction.

mailz macro

```
cat mailz
#! /bin/sh
date +%n%t%D%t%T%n
mailx $*
date +%n%t%D%t%T%n
echo "The message has been sent"
exit
```

The record of the end of the task is accurate but the time indicated for the start of the use of the mail system is not the same as the time when the task was first presented to the user (the time when he read the task description). These times are therefore used as a check in case of errors in the other method of recording the timing, described below. Making this transcript of the interaction is opaque to the user - though the command to start the mail system has been changed from 'mailx' to 'mailz' this difference does not affect the experiment and is explained to the potential email user after the completion of the experiment.

Times of performance alone when the task is aided is not necessarily a true reflection of the competence of the user even for tasks designed to illustrate different aspects of skill. Even when the times are supplemented with a transcript of the user's interaction with the computer the extent to which the user needs aid in order to complete a task is an aspect that needs to be examined. If a manual is particularly easy to use the user may record fast times for each question without necessarily having learnt any of the task. Alternatively a user may spend some time mastering a task with the aid of a poor manual but find that because of his efforts he doesn't need to refer to the manual again and thus produce fast times in the future.

The progress from a high level of initial use of a manual to unaided performance is a good indication that learning has taken place. There should be an interaction between time to complete the task and help needed. It is likely that initially tasks will be performed slowly and the manual will be used for most tasks. If the manual is effective its use should drop off and the time to complete tasks should decrease. There may be more interesting interactions:

- users that spend the most time with the manual initially may eventually become the most skilled performers

- some users may choose to access the manual for every task, learning may not take place or the user may find it easier to use the manual than work it out for themselves

- users supplied with one set of elaborations may quickly learn to perform the initial task without the manual but may need to refer to it again when faced with the tests. A different group may need to keep referring to the manual for the majority of the initial tasks but may not need it for the re-test tasks.

In order to record the extent of the manual use it was decided to use a computer to display the pages of text. A 'page full' of text is a possible metric but this ignores the substance of the information content. More useful is a measure of the depth of goals within the task hierarchy that the user needs to see before they can construct the necessary key strokes. For this reason 'pages' were limited to one level of goals. Subsequent re-descriptions were located on different pages until at the bottom of the hierarchy a page consisted of the key stroke(s) that comprised the operations necessary for fulfilling the lowest goal in the hierarchy. With such a heavily divided manual on paper it was be difficult for the user to navigate through the different parts that he needed to read; displaying the manual as a system with dynamic two way links between pages eased this problem.

The nature of the display system for these 'pages' will be described later, as will the meaning of the record of their use. All that needs to be said at this point is that the system recorded which pages were displayed and for how long.

The system that displayed the manual also displayed the test items to the user. The time taken for a task to be performed was measured from the time the task was first displayed to the time that the user indicated that they wanted to move on to the next task.

The design of the experiment was explained at the start of the chapter. The data recorded - transcripts of computer use, times per task and the amount of time that the pages in the manual used by the user were displayed for - have just been described. The next section describes the practical implementation of the different text types.

The display system

Hypercards (Apple Computer, Inc) was used to construct the manual display systems used in the experiment. The following terminology which will be used in the description of the electronic manuals needs to be explained.

NOTE These terms do not necessarily match the technical definition used in the Hypertalk (Apple Computer, Inc) programming language. They are used here to describe the display system's interface rather than it's internal workings.

Click

The action of positioning the pointer driven by the mouse over an item on the screen and pressing the mouse button once.

Buttons

An object in the interface that responds to clicking.

Page

Corresponds to an area of the screen approximately 16cm * 9cm or 80% of the screen area. It is here that the text of the manual is displayed.

Fields

A bounded area of text.

Jump

An action in response to clicking that causes the user to apparently physically 'jump' to a different page in the manual, actually by issuing a command for the display of a different page of the manual.

Presentation of the tasks

In each of the experimental conditions, with the exception of a small variation for the 'Core' condition, the first task was presented to the subjects in exactly the same manner. This consisted of a standard screen format containing a Title button, which when clicked would describe the buttons on the screen, a box containing the task description (the 'Question'), an Information button that could be used to open the manual (or in the case of the 'Core' condition signal that help was required), and a Next Question button that would cause the new screen containing the next question to be displayed.

Subjects were asked to carry out the task as described and that (if necessary) they could access the appropriate manual by clicking on the Information button. Once informed by the experimenter that they had successfully completed the task they were to move on to the next task by clicking on the Next Question button. Using this system they could attempt to carry out the question on their own but could open the manual at any time.

Generic description of the manual

During the first two stages of the laboratory work, the 'pilot' stages, the experimental materials were refined so that they corresponded with the aims of the experiment. The materials also differed according to the particular condition being run. In order to reduce repetition a general description is provided below followed by specific descriptions of the different designs.

The information in the manual consists of hierarchy of goals and associated plans normally found in a standard HTA. Each 'page' consists of a goal, a plan describing that goal in terms of a sequence of operations (which may be keystrokes or further goals) and a list of the operations described in the plan. Each of the operations which is further re-described in the manual (i.e. not the keystrokes) has a square box adjacent to it. By clicking on that button the user jumps to the page re-describing that operation.

The rest of the screen consists of buttons for navigation backward through the manual (To 'Go back a page' or 'Go back to the start (of the manual)'), an area where the question is always on view and a button which allows the user to go on to the next question. See Figures 8 and 9.

When subjects first open the manual they are presented with a page containing the top level of the Email HTA I.e. 'Use Mail System'. From that page they can access any part of the manual by clicking on the box next to the part that they wish to see re-described. To view information that is 'higher' in the hierarchy they must either go to the start of the manual and work their way down to it or go back successive pages up the hierarchy until they reach the location of the branch to the desired part of the task. Figure 10 illustrates the instructions given to subjects about navigation through the manual

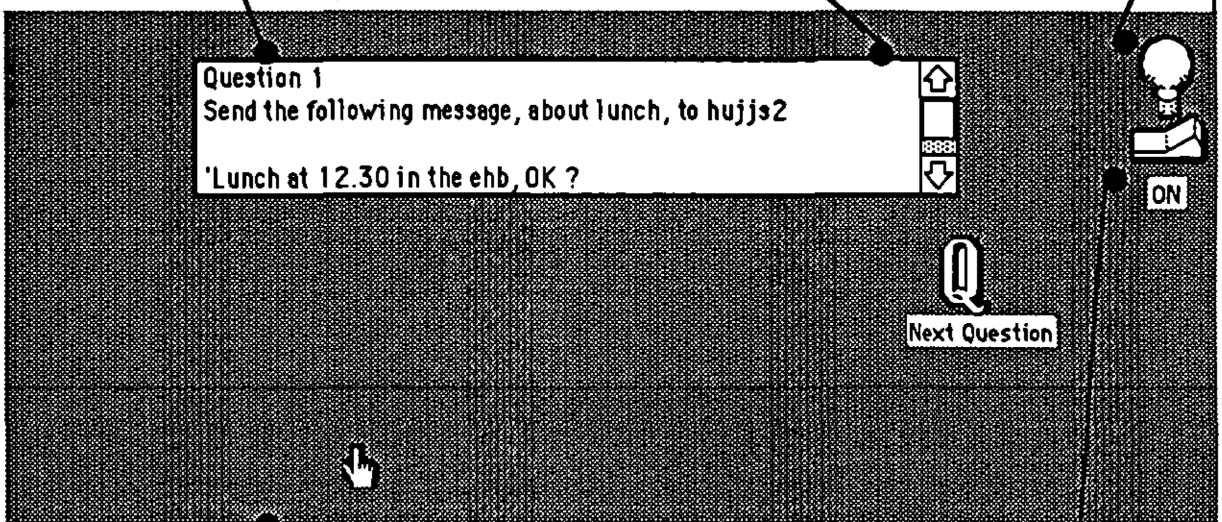
Figure 8 - Using the electronic manual

When you have opened the manual (See Start Here) you will see a screen that looks like this one.
(Note - on some screens bits and pieces like the 'Next Question' button will be missing.)

Curent Question
The text inside this box describes the message that you must send using the other computer

Scroll Bar
If you click on the arrows the question will scroll up and down - unless of course all of the question is in view already.

Light Bulb
At the moment it is off so the majority of the screen is dark.

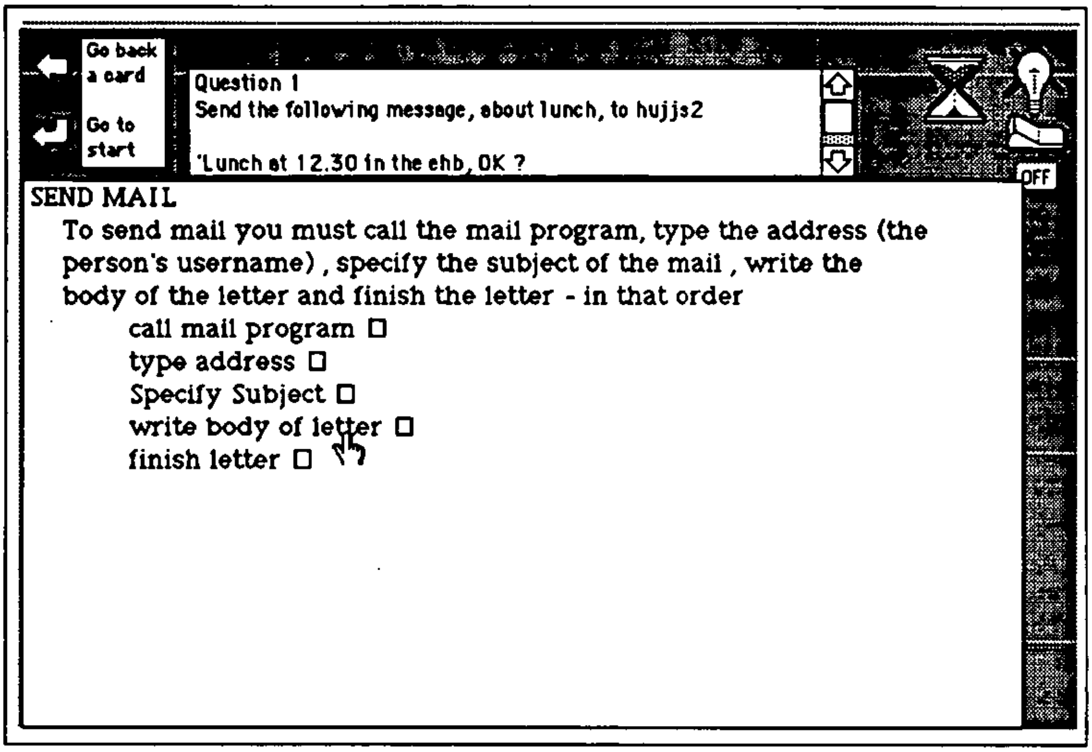


Next Question Button
Click on this to indicate that you have finished sending the current message and that you want to see the next question.
(Note - this will take you back to the title page of the manual and you will have to carry on with the new question. **There is no going back.**

Light switch.
Click on the switch to turn the light on. This will cause the rest of the screen to be lit up - you will then be able to see what is written in the manual.

Poiting Hand
'To 'click' on something - ie to indicate to the computer what you want it to do - place the finger of the hand on an object, using the mouse, then press the button on the mouse.

Figure 9
A typical screen showing buttons and light



START USING THE SYSTEM
To start using the system do operation 1 then do operation 2 in that order.
Operation 1 ☐
Operation 2 ☐

This is the first page of a hypothetical manual. In order to Start Using The System you should follow the plan (See Diagram 1).
In this case, you should do Operation 1 then Operation 2 in that order. If you already know what Operation 1 is then you should do it. If you can't remember what it is or have never known what 'Operation 1' is then you should click on the box next to the operation. (Make sure that the tip of the finger is inside the box) This will show you further information about that operation.

So to carry out Operation 1 you should Type 'Start' - you should always type what is inside the quotation marks.
Now you must go back to the previous page to see what to do next. To do this you should click on the arrow in the upper left hand corner of the screen - the 'Go back a card' arrow. (See Diagram 1) Note - click on the arrow, not the words.
The next operation to be carried out is Operation 2. If you don't know what to do you should click on the box

OPERATION 1
Type 'Start'

This is operation 2. As you can see, you must Enter the information. Again, if you don't know what this is, click on the box.

OPERATION 2
Enter the Information ☐

You will be shown further information about the operation 'Enter the information'. It tells you that you must press the Enter key - the large key towards the right of the keyboard. When you have done this, click on the 'Go back a card' arrow to see what to do next. This will show you the card 'Operation 2'

ENTER THE INFORMATION
Press Enter.

There are no more operations to be carried out in order to complete Operation 2. So go back a card. This will take you to the Start Using The System card.

Again, you will see that you have carried out all of the operations to finish the operation 'Start using the system'. You have now finished the task and should go on to the next question by clicking on the Next Question button. See Diagram 2.

Not all plans will be as simple as this. Some times you will have to decide which part of the plan you need to carry out. Sometimes you will have to repeat a sequence of operations.
The important thing is to read the plans carefully and follow the operations precisely.

Use of and navigation through the Electronic Manual

The general principle of the manual is that the user should follow the plans on the pages, starting from the top of the HTA, so that he selects the appropriate goals from those listed and attempts to carry them out in the order specified. If unable to carry out a particular goal, because they do not know all of the sub-operations, they should click on the box next to the problematic goal to obtain a page that re-describes it in terms of another plan and list of goals. If they are able to carry out all of these they should do so then click to go back a page and perform the operations to obtain the remaining goals on that page. See Diagram 3

In order to use the manual effectively the user must:

1. Follow the plans carefully.
2. Only carry out the operations that the plan appropriate to their goal specifies, in the order described.
3. Remember that once they have performed all of the goals on a particular page they should go back a page, to the plan higher in the hierarchy and complete those goals.

Data collection

The time that the question was first displayed to the user was recorded by the program as the start of their time to perform the task. The end of the task was demarked by the subject clicking on the 'Next Question' button. The start and end time of the display of each page of the manual was also recorded as was the name of the page. The name of the page corresponded to the goal that the page described.

Description of specific manuals

1. Goals, Light, Far Manual

This manual was used for the initial video pilot studies and was presented to the first group of subject in the subsequent study. A number of refinements were made to the material and the design of the experiment in the light of these two pieces of work. So rather than attempt to present the design as complete the variations in the suite of pilot studies will be described and then the results will be discussed.

The text used in the manual corresponds to the HTA condition - only the goals, plans and sub-tasks are described. The light stays on for 10 seconds at a time. 'Far Manual' refers to the transfer task - the particular wording makes this test more difficult than a close-transfer test. The distinction will be further described in the description of the results.

Difficult versus easy manuals

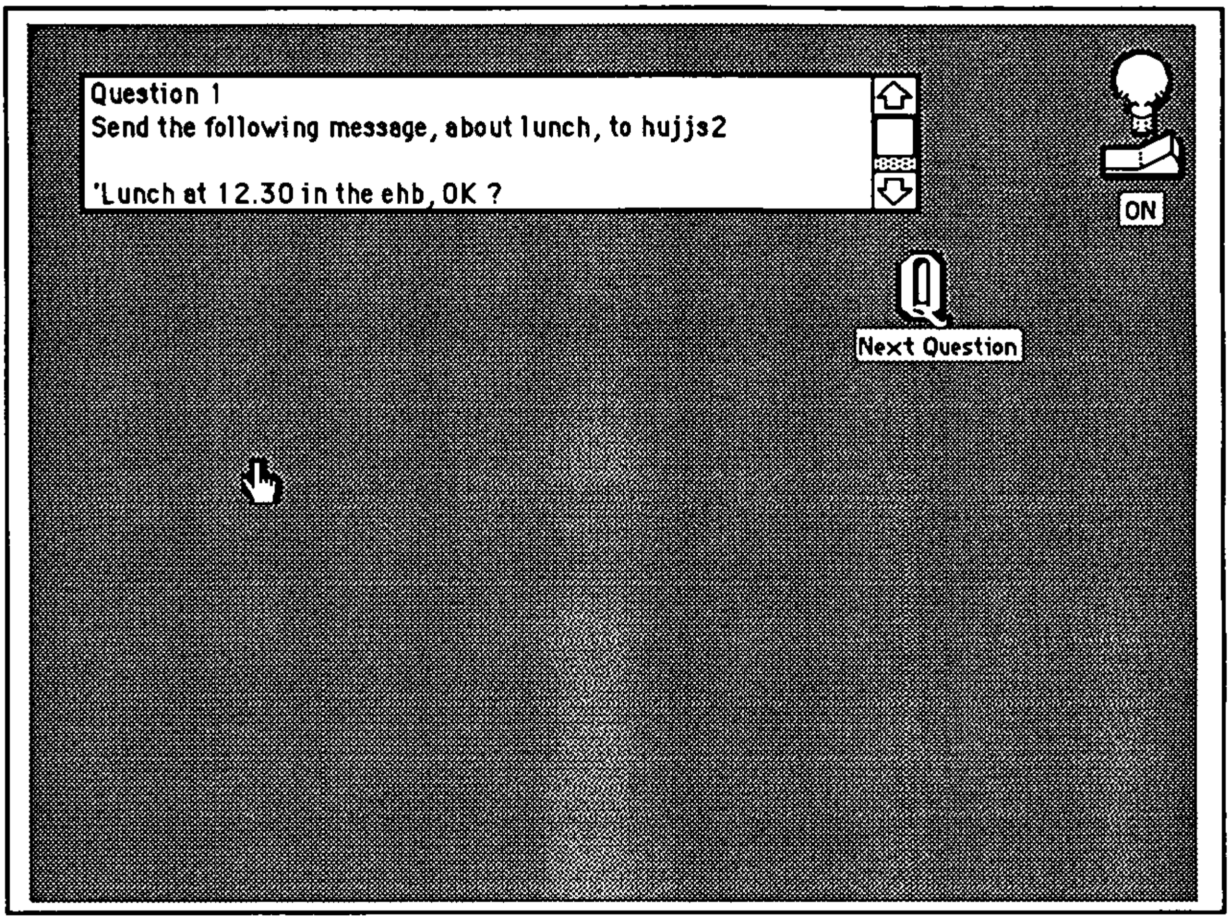
One of the experimental conditions was designed to make it more difficult for subjects to use a manual in comparison to another easier manual; by imposing a cost for referring to the manual, something that the subject would be adverse to doing within the experimental scenario. The manuals needed to be identical to the others in all other respects so although a cost could have been imposed by making one manual more verbose or perhaps requiring the subject to carry out a complicated routine in order to gain access to the manual these would have intruded extra variables into the experiment as well as the desired 'cost'. For example the difficult manual could have been produced in a smaller, less legible type face, thus making reading more difficult. But the extra processing required to read the text could have led to a better memory for what was read.

After examining different ways of loading the manual it was decided that the general form of the cost to the subject was that the manual would require more time for its use. In order to impose this a manual was produced with a switch fitted to it. Every time the subject wished to see the manual he would have to press the switch. By altering the time between the switch being pressed and the manual being displayed the cost of using the manual could be altered.

In order to make the principle of the manual use clear to the subjects the switch was portrayed as a light switch. When the switch was turned on a light was lit and the manual became visible. Because a penalty was to be imposed for each instance for manual use the light was attached to a timer, the penalty being the delay before the light came on. After a fixed time the light went off again and would have to be switched on before the manual could be used on a subsequent occasion. The physical analogy was described to the subjects in the instructions and displayed on the screen by making the button for the switch look like a light switch and placing an hour glass (egg timer) next to the switch on the screen. The sand in the glass ran out over the 10 second period that the light remained on. During that period the user could navigate through and read the pages of the manual but while the light was off all of the buttons for navigation and the text of the manual were blanked off. See Figure 11.

The decision to make the display period ten seconds was based on a compromise between a time that would be shorter than two successive uses of the manual (i.e. reading the manual, doing something, then reading the manual again) but long enough so that a large portion of a page could be read. Note that most screens contained only a few short lines of text. What was important was that once the switch was added to the display system a penalty time could be added to the switch so that the user in the delay condition would have to wait a few seconds after pressing the switch before being able to view the text of the manual. The consequences of introducing the switch are discussed in the results section.

Figure 11 - Light switch off



In this condition although the switch is present no delay is imposed.

Elaborations

This manual consists of the goals and plans derived from a HTA of the mail system and so appears the same as the generic system described above.

Task types

The task consists of the operations that the user must perform in order to meet the task description. The term 'question' is user to refer to the description of the task as presented to the user. The three generic questions used in this experiment were :

1. Learning question
Send the following message, about <subject>, to <username> '<message>'.

2. Close transfer question
'Forward the file <file name>, about <subject>, to <username>. Add the text '<message>' when sending it.

3. Non-specific transfer question
Create the alias <alias name> for the user <username>

The practice task was, as stated earlier :
'Instruct the computer to display the calendar of December 1990'. A description of the task and the rationale for using it is given below.

The label 'Far' has been applied to this condition because of the wording of the close transfer task - the question describes the operation of inserting a file as the higher goal of forwarding a file. The close, equivalent, task names the operation directly.

Tasks are presented on the 'title' page of the manual - the subject has the option to perform the task without the manual or to open the manual then perform the task with aid - see Figure 12

Video Pilot Study

The purpose of this study was to reveal the practical problems of implementing the experiment. In this respect it was treated as a question asking study (Kato, 1968); used as a means of finding out what instructions were required and which parts of the materials did not work or could be improved.

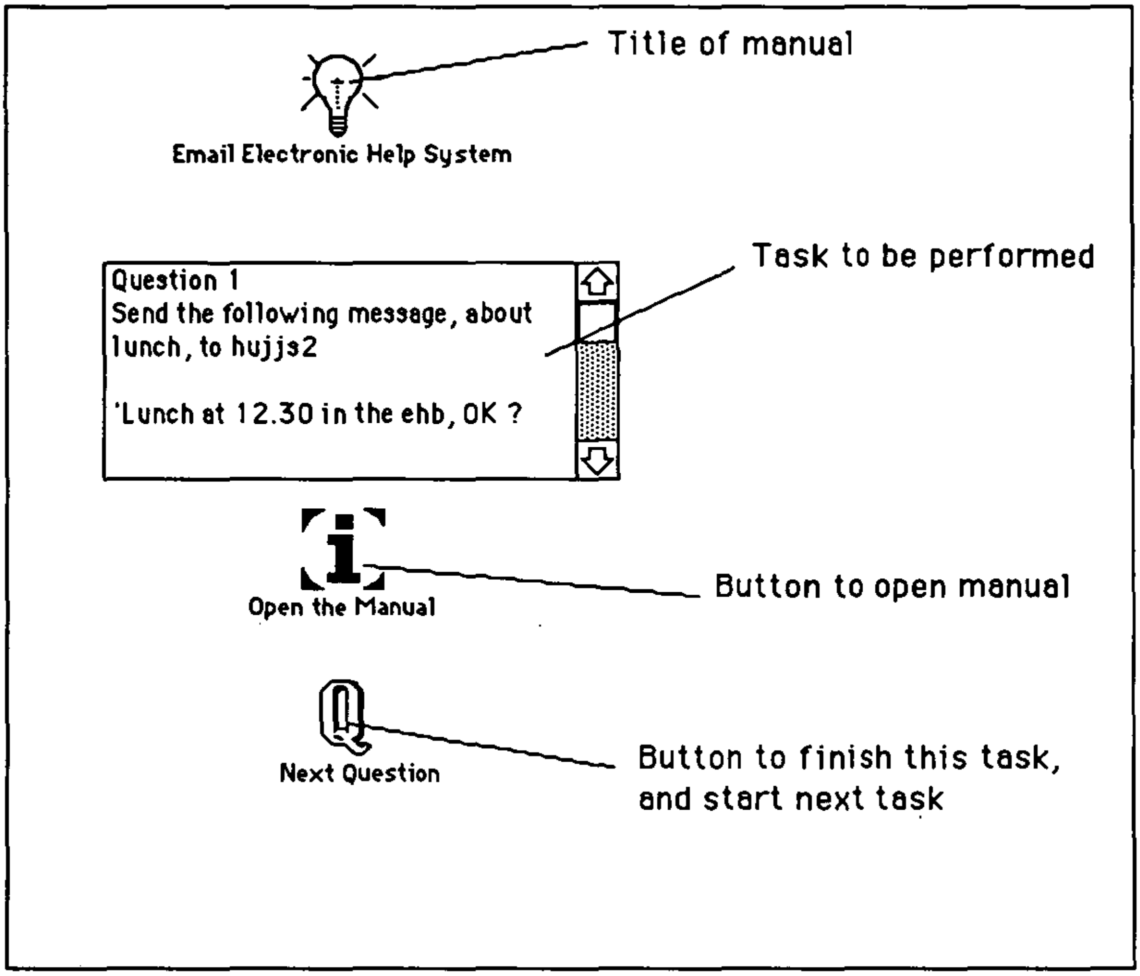
The trial was limited to the testing of the question display, the display of the the manual and the process of using the manual to aid the sending of simple messages (task 1) i.e. the major features of the electronic manual.

Subjects consisted of one human factors researcher, to provide a critical appraisal of the package, and two inexperienced computer users to provide an appraisal from the perspective of the subjects to be used in the main study.

The email tasks were to be carried out on an IBM PC type micro-computer (HP Vectra henceforth referred to as 'PC') connected as a remote terminal to a Hewett Packard Vectra microcomputer using the emulator 'Kermit'. The electronic manual was run on a Macintosh (Apple Computer, Inc.). In order to aid the video recording the Macintosh was raised to the same height as the PC screen and the two placed as close to each other as possible. By filming over the subject's shoulder it was possible to capture all of the Macintosh screen and the part of the PC screen used by the subjects while sending email. The video included sound recording.

Transcripts were made of those parts of the interaction with the system and conversation between the subject and experimenter that revealed problems with the experiment. These are described below.

Figure 12 - Display of the task



Subjects were told that the experiment was a test of the efficacy of the manual rather than of their skill and that they should feel that anything that was not clear to them or difficult was the fault of the experimenter and should be raised as a question in order to make it easier for others.

They were informed that they would be required to send four email messages, with the help of a manual if they wished. The Question Screen was explained (using diagram 5), as were the buttons and text on the manual screens (diagrams 1,2,3 and 4 were used for this). Navigation through the manual as an aid to performance was explained using diagram 3. This part of the dialogue will be explained more fully in the section describing the final version of the experiment where the script was finalised, through practice and the results of the preceding pilot experiments.

Results of the video pilot experiment and design changes.

Facets of the experiment that were not immediately obvious to the subjects (that needed explanation), parts of the material that needed refinement (including the interface to the manual, some of the contents and some of the typography) and inefficient parts of the data collection were revealed by this initial study.

The subjects required an overview of the experiment including details of which keyboard to use to send the mail, when they could use the manual, how many questions there were, the necessity of pressing the next question button when they had completed the task and the constraint of not being able make notes about the task. These items would normally have been included in the design of the general instructions without conducting this pilot but it was useful to have the transcripts as a basis for design in order to produce a set of instructions that worked *before* conducting the main studies.

All of the features of the interface seemed to be clear to the subjects, particularly the use of the light switch. The only item that needed to be explained was that they should click on arrows rather than their written descriptions in order to go backwards through the manual. The arrow

'Go back a card' was supposed to take the user to the card higher in the hierarchy where they had come from. During tests of the system the experimenter went linearly down the hierarchy then up again to the same page - all appeared to be well. However subjects explored up and down different parts of the hierarchy before attempting to go back to the higher levels; this process showed that the button took them to the last card that they had read - not necessarily always up the hierarchy. The function of the button was changed to meet the desired specification.

The use of the manual as an aid to performance, i.e. the navigation through the manual was initially found to be difficult for the subjects. They were sometimes unsure of where to go once they had competed a sub task and how to get back to a higher level goal By the end of the fourth task all had mastered its use, though in a debriefing the casual users voiced a preference for paper based manuals because they could actually see where they were in a document without having to remember which pages they had already examined. This point was particularly evident when subjects made errors or examined inappropriate parts of the manual - they got 'lost' and found it difficult to find the part of the manual that they felt was appropriate to their problem.

In order to help the subjects learn to use the manual an introductory task was created in addition to the diagram illustrating how they should move through the pages. This will be described at the end of this section.

A considerable amount of instruction was needed for the casual users on how to navigate through the manual. It needed to be emphasised that operations that were followed by boxes were not items to be typed but were goals that could be re-described by clicking on the box. Items that were to be typed were bracketed by quotation marks, which were not to be typed. Subjects needed to be reminded that when they had carried out all of the operations on a page, that they had fulfilled the main goal, they had to 'go back a page' in order to see what the next goal was. The importance of following the plans needed to be emphasised - the subjects had a tendency not to read them but to pick out items from the list of operations that they felt were appropriate to the task then carry them out in an order that they thought would work.

When questioned subjects stated that they did not always follow the plans because they did not understand them; they did not understand how they related to the task they needed to carry out. An example of this is their desire to create an alias when the task required them only to send a mail item to a username. The plan describing when an alias should be created was not clear so subjects assumed that they had to do this before sending any mail item, particularly because the operation 'Create alias' was placed at the top of sub goals of 'Use Mail System'. This problem was not limited to the operation of creating an alias - some of the names of the operations did not always clearly suggest their meaning, for instance 'Process mail'. The names are only labels, the meaning of which the subjects should learn from the underlying sub operations and the context of their use. It is therefore the plans that describe when to carry out the operations that should be refined so that subjects know when to carry out a particular operation because the plan referring to it makes this clear. However the basis for the design of the manual was meant to be limited to the training hypotheses being studied in the experiment. In this case it was to provide a hierarchy of plans and operations to structure the keystrokes to be carried out for a task - not to produce elaborate descriptions of the meanings of the labels in terms of the way that the system worked as opposed to how it should be used to meet goals. It was therefore decided to review all of the plans in the manual to ensure that the context in which goals should be carried out has been made clear but that attempts to produce more descriptive labels should not be made. It may be that some elaborations on this basic form would overcome the problem. In the case of the operation 'Create alias' the plan was re-written so that the user knew when it was appropriate to fulfil that goal.

Some aspects of the typography of the manual were not clear - periods and tildes were often missed by users; it was decided to point these out to the user if they started to type a line containing them. This differs from a decision to point them out as soon as the user displays a page containing them because the action of the experimenter pointing out their existence tended to suggest that they should be typed - which was not always the case.

The manual did not describe the effect of performing operations, e.g. it did not describe what would appear on the screen. This meant that subjects were not aware of when they had finished a task unless they checked that they had performed all of the goals. Even after doing this all of the subjects were unsure of their decision so it was decided to inform subjects when they had completed the task and also to inform them when they had carried out the task incorrectly.

Subjects had a tendency to look at the manual and not at the results of what they had typed. This behaviour had been seen in other people carrying out novel computing tasks and was not felt to be an artifact of the situation.

Familiarisation task

Though the way subject needs to use the electronic manual is the same pattern of behaviour that users ought to follow with traditional manuals it is unlikely that they have encountered such an explicit hierarchy within a handbook. The instructions for the experiment inform the user of these rules of behaviour but this is a lot for the user to remember; based on results of the video pilot study it was decided to produce an introductory task of the computer that the subject could use as an opportunity to learn and make mistakes.

To make the practice as valid as possible without teaching any of the email task it was decided to present a short task that the user would carry out using exactly the same equipment.

The task chosen was that of displaying the calendar of the month that the subject was born in - a simple facility of Unix that allowed them the chance to use the manual and experience how to follow a sequence of instructions, call re-descriptions of operations, go back a page in the hierarchy and go back several pages through the hierarchy.

Subjects were given as much help as they needed, whether requested or it became obvious from their behaviour - in both the use of the manual and using the PC. One of the objects of the exercise was to familiarise the subject with the general experimental set up and reduce any anxiety that they might have.

2. Pilot experiments

Although the set of experiments was initially designed as one coherent study the individual experimental conditions will be described in the sequence that they were performed because some of the insights gained in the study led to macro changes in the design of the material. The qualitative results that led to the changes will be described after each description of the material while the quantitative results will be discussed as a whole at the end along with more general qualitative results.

Subjects

Seven male and female undergraduate students were used for each pilot study. None of the students had used electronic mail before and had only a small amount of computing experience.

General Design

The materials and lay-out of the computers were as described above. Each experimental group comprised of seven subjects. The first, training, session took up to one hour; subjects were aware that a second, similar session would take place and were asked not to practice using email and not to write down anything that they had learnt. Subjects were given any aid required while carrying out the practice task but during the main learning tasks aid was restricted to the use of the manual though subjects were informed when they had performed a task incorrectly.

The second session lasted slightly longer and took place after a gap of two full days - i.e. Monday then Thursday. This gap was chosen as a compromise between a longer period where it was likely that subjects would forget or be unwilling to attend the second part and a very short period where 100% recall was likely. The initial instructions were repeated, as was the practice task though with slightly different parameters (a different month to be displayed). Subjects were debriefed at the end of each session, with a particular emphasis on any difficulties they had with the task, the manual and the display system. This allowed the experimenter to discuss with the subject the reasons for their behaviour during the session and for the subject to describe his perception of the manual and the task. After the second session the nature of the experiment was explained in detail to the subjects and any problems that they had in learning to use email were resolved.

1. Goals, Light, Far Manual

This is the same as the condition studied in the video pilot study, with the exception of some minor changes to the manual and interface, as described above.

The main details of the condition are:

1. The manual consists of goals and plans.
2. A 'light switch' is attached to the display of the manual.
3. The close transfer question is worded as below.

'Forward the file <file name>, about <subject>, to <username>. Add the text '<message>' when sending it.

Qualitative results.

Observations from this experiment that were seen in the other conditions will be discussed at the end of this section.

The main observation that led to an immediate change in the design was the difficulty that subjects had in performing the transfer task. It appeared that the subjects did not recognise that forwarding a file was subsumed by the action of sending mail or if they did when they examined the 'Send Mail' re-description and were unable to find an item that suggested forwarding a file they went back and looked through the Process mail and/or Create alias sections for long periods. Frequently they would come to a stop and ask for help, at which point they were advised to explore sections of the manual that appeared relevant to the problem. With perseverance all of the subjects found the relevant section, which was a sub goal of 'Write body of message' within 'Send mail'.

This was seen as a similar problem to the one revealed in the video pilot in that the names of goals did not suggest their function within the task and even the terms in the plans were not intuitively obvious to subjects. Rather than re-write the task analysis and manual in order to representing 'very frustrating'. Their comments all pointed to the way that the light went off as they were reading as the main source of their frustration. No mention at all was made of the delay before the switch worked though both groups found the extra mouse actions needed to work the switch irritating and an effort.

The addition of the switch alone appeared to be a penalty for manual use. It was therefore decided that the condition with a switch should represent the the manual + effort and that a manual without a switch should be the normal condition.

3. Goals, No Light, Close Manual

This condition again has the same manual but the light has been removed from the interface in order to produce an unpenalised form of the manual.

Qualitative Results

Though many of the subjects complained of frustrations with using the manual they were based on the need to flip through so many pages rather than go directly to the information they wanted. Frustration scores were lower than in the other conditions.

4. Core Manual, Close Task

The manual in this condition provides the user with the keystrokes that they need to perform for a given task. In order to record what information the user needs and to prevent them from just copying out a sequence of instructions it was important to withhold the details of the keystrokes until the user requested them.

Ideally a simulation of the mail task would have been written, including a help button for the user to request the next key stroke. As a programmer was not available this was not practical.

The practical solution was for the experimenter to supply the answer to the user's request for help when they pressed a button on the question display screen. Since the experimenter was already present during the experiment it was felt that this form of manual would not radically change the experiment with the exception of the way that the help was provided.

In order to maintain the type of help provided it was decided to reply to the subject's request with the number of keystrokes that the subject in the other conditions would see when they displayed the page relevant to that part of the task that contained the description of the necessary keystrokes. For instance, if they wanted to know how to type the message they would be told to type the text then press Enter. However, if the next line that they needed to type consisted of a number of these pages of operations they would be given only the keystrokes a page at a time.

In order to reduce the amount of interpretation needed by the experimenter in order to answer the subject's requests help was only given when it was specifically requested it by clicking on a 'What's next ?' button. The help was limited to instructions about what they would have to type after the position of the cursor. This meant that the help was literally limited to the next keystrokes but it also meant that it appeared rather unnatural to the subjects. For instance, if the subject made an error and the computer returned the cursor to the beginning of the line, instead of the experimenter informing them of what they did wrong they would be told what to type at the beginning of the line - often the subject had typed this correctly.

Subjects were informed that they would be told of the next keys to press and of the ramifications of this type of manual, as described above.

Qualitative Results

Subjects were surprisingly reticent about asking for help. They made far more attempts to work out how to perform the task through trial and error than any of the other groups. The latter in the majority of cases went straight to the manual when they became stuck. The subjects in this condition not only tried to guess what to do from the outset but when they attempted the recall and close transfer tasks their attempts illustrated an underlying goal structure that they had created. They were able to carry out the operation of starting the mail system and addressing the message then frequently hesitated, not sure of what to do next. Once the operations that described the insertion of the file had been provided they finished the task. The hesitation shown once the message had been started indicates that they were aware that this was a separate goal to inserting the file and finishing the message.

It took some time for subjects to settle on their strategy for using the 'manual'. Often this moved from concerted attempts to work things out for themselves to acceptance that they needed help and requesting it when they were stuck. Others asked for help frequently, just because they were unsure of the next action even though they knew what the correct one was.

Many subjects found the withholding of help irritating - i.e. supplying only some of the keystrokes necessary to make up a complete line, however they quickly adapted to the situation.

On the recall tasks this group of subjects performed with ease. However they found the transfer task particularly difficult because they were unsure where in the task they should start trying to insert a file. This information was observed in the subjects continual hesitation and muttering while starting the transfer tasks and their unsure behaviour compared with the recall tasks. This may be because they have learnt the key sequence by rote rather than by knowing the underlying goal structure.

Throughout these pilot studies, and the video pilots, it was noticed that some subjects copied their successful actions from the previous task directly from the screen. The implications of this behaviour now became apparent; by doing this the subjects were not using the manual that had been designed for their experimental condition - instead they were looking at the results of their previous actions as a source of help rather than the manual. If a measure of this type of behaviour had been made then this behaviour could have been acceptable. Indeed on many mail systems and other computing tasks it is possible to copy as the subjects did here. But in the conduct of this experiment it makes the training system that is being tested a weaker component of the experiment and is unacceptable.

In order to test the the effect of copying from the screen on the subject's performance it was decided to re-run one of the groups with their keystrokes from the previous task moved off the screen before they started the next one.

Of the conditions so far the one that this change was most likely to effect was the Light+2 condition, which was designed to encourage the user not to use the manual and therefore make them rely on other sources for aid - their own memory and clues from the screen.

5. Goals, Light +2, Close Task, display covered

This is the same as condition 2, Goals, Light+2, Close with the addition that the screen was cleared after the subject had successfully completed each task.

Subjects did not appear to have any more difficulty with the tasks. Such a difference may be seen in the quantitative results.

Qualitative General Results

These results apply across all of the conditions. They are listed in this section but will be drawn together and discussed at the end of the chapter.

1. Many people became 'stuck' - unsure of how to complete the task and unsure of how to use the manual to aid them. Help given to such subjects needs to be formalised so that it is not biased in any way.

2. Even though subjects appeared competent at a particular task they sometimes made surprising errors due to a lack of understanding of the concepts involved in sending email. They also had difficulty in understanding what the task question was asking them to do - they didn't understand what the question meant. Concepts that required explanation were - alias, username, address, inserting a file, the difference between a file and its name, creating an alias versus sending

mail, the difference between inserting text and inserting a file, the difference between the subject and the body of the message.

3. Some subjects did not watch what they were typing on the screen and so missed messages from the system or made slips in performance.

4. Subjects were unsure if the files mentioned in questions already existed or if they had to create them.

5. If subjects accidentally stumbled on the section 'insert a file' while sending a plain text message they were more likely to perform well on the transfer tasks than if they had not erroneously discovered this section.

6. Subjects frequently pressed 'Return' immediately after typing 'mailz'. This action caused the contents of the mailbox to be listed rather than initialising the sending of mail. This may be because they were used to pressing Return after each command in other contexts and because it was at this point in the manual useage that they first learnt to back track through the manuals - pressing 'Return' seemed the obvious next step at the time before they learnt how to back track.

7. Subjects were often hesitant about their ability to carry out a task and frequently looked to the experimenter for reassurance before typing a command. This behaviour became more evident after an error had been made, they tended to use the manual more than before even though apart from the slip they had been very competent at the task.

8. The non-specific transfer task was found to be harder than the other tasks because the system gave fewer feedback messages.

9. People don't realise when they have finished the task, so a message was added to the system to indicate when mail had been successfully sent and the experimenter told them as soon as the task had been successfully completed.

10. Most of the subjects efforts went into performing tasks for the first time, the process of discovering this seemed to be hardest, the next attempts seemed easier.

Quantitative General Results

The raw scores and tables of Means and Standard Deviations for the time taken for tasks, number of cards accessed and number of 'type' cards accessed can be found in Appendix 2.

1. Examination of the means and standard deviations reveals that there is a large variation within each group, for all measures, compared with the size of the mean. This makes it impossible to draw statistically significant inferences from the data and difficult to talk about general trends. Examination of the raw scores and the transcripts of the experiments shows that this variation is due to two main factors:

a. Subjects that performed considerably better or worse than the majority.

b. Subjects that made an error during a task due to either poor comprehension of the manual or slips in performance that led to confusion and substantial remedial action in order to correct their mistake. For this reason examination of subject's behaviour relative to their own base-line scores was no more effective at reducing the within group variation. Some of the reasons for the large variation will be discussed at the end of the chapter, as will means of reducing these.

2. The high variation within the groups extended to all of the task scores - both the scores for initial and final tasks and means across groups of tasks. It had been anticipated that differences between groups would have appeared throughout the tasks within a given task group - for example all of the times for transfer tasks of one group would have been faster than another experimental group. Instead, even if the high within groups variation is ignored, cross overs in score levels occur - for example one group may be faster than another on the first transfer task but slower on the third transfer task. Such a pattern decreases the potential for clear interpretations, particularly in combination with the high variation of scores.

3. Examination of the scores, in particular the differences on the non-specific transfer tasks and the consistently fast times for the core manual group, observations during the experiment and a re-analysis of the experimental design show that a major determinant of the results was the manual that subjects used during the second series of tasks. Ease of use of the manual, the amount of text to be read and the time to access the information all lead to versions of the manuals, in particular the core manual, that were inherently faster than others. This meant that the result of the second series were as much due to the method of acquiring information as the initial training strategy. The implications of this result are discussed at the end of this chapter.

4. Quantitative analysis was further hampered by the derivation of the differences between the experimental materials. The design changes evolved from qualitative observations made during the the experiments rather than specific attempts to design experiments based on testable hypotheses about the material. For instance, it was found that the light plus delay was not significantly different to the light alone as a penalty for manual use - according to the subjects descriptions. Alterations in the design of the materials were designed to improve the design of the experiment rather than prove the qualities of the two second delay as a penalty. Similarly, once it was found that subjects were confused by the far transfer task it was decided not to produce another manipulation based on the far task in order to add weight to the observation; and rather than repeat a trial using the light without penalty condition it was decided to try the light plus two seconds with a close transfer task. Instead of varying just one variable to study its effect two variables were changed in order to test the experimental materials rather than examine the effect on learning. Thus the manipulations did not lead to clear hypotheses that could be tested quantitatively, instead the the aim of the trial was the qualitative results discussed above.

5. Each group and each measure shows a similar pattern of results. Time to perform a task decreases with repetition, as does the use of the number of cards accessed and the number of 'type' cards accessed.

Differences between the groups with respect to this pattern are too insignificant to analyse.

Quantitative Specific Results

It is difficult to draw definitive conclusions from the quantitative data for the reasons described above. To the extent that this is possible the degree to which the quantitative data supports the observations made during the experiments is described below.

Light and far task condition

It was observed that people had difficulty with the transfer task; they became confused and could not find the relevant section of the manual.

With respect to the next condition to be tested, light plus two second delay plus close transfer task, subjects were slower on the transfer task (though they did not have to wait for the manual to be displayed), used more cards and more 'type' cards.

Although this would support the informal observation this result may be due to better learning by the Light plus 2 group because of the extra deterrent against using the manual. As stated above, these factors cannot be separated by this trial of the experiment.

Light + 2, Close task condition v. No Light, Close condition

As stated above clear differences cannot be drawn between the performance of this group and the Light, far task group. However, it can be compared to the No delay light, close task group. It was observed that subjects rated the No light condition less frustrating than the Light plus two condition. If subject's rating is a good indicator of the effect of the light delay on the subject's use of the manual then these results should reflect better learning of the material and better performance of the transfer task due to the extra effort put into learning rather than using the manual (See Chapter 6).

Subjects in this condition are marginally slower on the original learning task than those in the no light condition and better on the transfer task though these differences are outweighed by the within group variance. This result lends some support to the experimental hypothesis. However, from the general qualitative results it is clear that the manual penalty factor has a strong effect on subjects frustration rather than on their willingness to use the manual. It was decided not to continue this method of experimentation on cost of manual use and to concentrate on other training strategies instead. This decision will be further elaborated at the end of this chapter.

Core manual, Close task

Subjects in this condition were the fastest on all tasks; from observation this is because of the ease of access to information rather than due to learning. Evidence for this is that these subjects were considerably faster than other subjects on the very first task, not because they used less help because the records of the 'type' cards used show that these subjects needed more specific help than any of the other groups.

Light + 2, Close, Covered

The aim of this trial was to see if preventing subjects from viewing the record of their previous tasks as shown by the VDU would lead to a decrease in performance because they would be unable to copy their previous actions. Observation suggested that they did not find the task any more difficult though some did note that they weren't allowed to copy.

The results of this trial are to be compared with the previous light plus two, close task group. Contrary to expectations the covered group were faster than the group that could see their results on all tasks. The greatest difference ocurred on the first of each group of tasks. However, the large within group variation precludes the drawing of firm conclusions. Examination of the raw data shows that there are three subjects within the uncovered group with particularly slow performances; this may be the cause of the difference between the group's overall performances. A possible explanation for the improved performance is that the covered group was forced to attend to the learning of the material rather than the rote reproduction of it. If this were the only cause then the groups would have similar scores on the initial learning tasks - however the covered group is far faster and requires less help even on the first task. This suggests that the differences may be due to variations between the subjects ability rather than the experimental condition.

Summary

This chapter has described the design of the laboratory tasks, the implementation of the manuals in an electronic form, the specific tests used in the experiments and some initial studies. These have revealed problems with the tasks, instructions and design of the experiments. They have also provided preliminary results of the effects of the elaborations on user's behaviour and difficulties that the users encountered with aspects of the experimental design.

Several subjects did not understand what electronic mail was or what affect their actions were having on the messages. Evidence of this was their mistakes - e.g. trying to send a message by creating an alias, the difficulties they had in finding the parts of the manual appropriate to their tasks, and the difficulty of the far-transfer task.

In order to alleviate this problem a description of the electronic mail system in the terms of the goals of the questions and the general functions of the mail system. This mode of instruction could be applied to each form of the manual without providing additional procedural instructions.

Several of the subjects came to a halt during the experiment, stating that they needed help if they were to continue. It was decided to formalise the provision of help by limiting it to the instruction 'Search the manual for information relevant to your task.'. If subjects made an error they should be prompted in terms of the phrasing and goals of the question.

The light switch which was used to increase subject's desire to learn the task rather than use the manual did not work as expected. Unfortunately the light tended to go off while subjects were reading - this acted as a frustrator both when the light delay was short or long.

The sudden switching off of the light also disrupted the normal reading by interrupting the subjects in the middle of a passage. This disruption affected part of the learning cycle that was not manipulated in other conditions and so it was decided to remove this condition from the experiment.

The non-specific test of transfer was not correctly presented in the current design. The ease of use of manual describing the non-specific test affected the outcome of the tests - the presentation of information necessary for the re-test tasks needs to be balanced. This will be described in the next chapter.

Some subjects tended to copy from the screen instead of using the manual - a group has been tested to see the effect of clearing the screen before another similar question is attempted. It was decided to continue the practice of hiding their previous attempts at tasks after each one had been completed - even though the empirical results were not conclusive the observations and reports of subjects support this decision.

The next chapter will describe the implementation of these changes and examine the effect of the elaborations on skill acquisition on a larger number of subjects using a more rigourous and appropriate method.

CHAPTER 8

THE EFFECT OF GRAPHIC DESIGN ON ELABORATED TEXTS

Introduction

The previous series of electronic manual experiments highlighted several issues about the presentation of text. One of these highlighted an aspect of the experimental design that may have affected the results. Subjects reported features of the manuals that they felt detracted from their ease of use and hence reduced their usefulness as learning/ performance aids. The presentation of the material had been designed to be bland to avoid the need to balance the effects of typographical mathemagenic aids across texts whose contents provided different opportunities for the use of such devices. Subsequent examination of the texts by graphic designers exposed aspects of the presentation caused by the blandness of the design that would, according to 'common sense' make the texts poor aids to learning - described below.

The aim of this experiment is to test the effect of improving the typographical presentation of the material - both elaborated and unelaborated - on learning before continuing the examination of the manuals with large numbers of subjects.

Comments made by Graphic Designers

The following is a quotation that summarises the graphic designer's general attitude to layout as a means of improving the usefulness of a document :-

'I'm not talking of typography - just making it easier to use. What you'll find is that when people don't understand, it's your fault because all of the opportunities to help the reader have not been done. Confusing to read, hard to get the info. out. To make it a good manual it shouldn't need to be explained.'

They specifically requested :-

1. More pointers are needed to tell them how to use the text, what to read first, where to go next.. E.g. a title page to explain the nature of the text, to relax the reader.
2. More white space; less solid blocks of text.so that separate ideas can be visually picked out - to direct the eye, to direct interest. E.g.Half line spaces between operations.
3. Stronger titles to act as pointers/ markers.
4. Bullet points or numbers for picking out points.
5. Don't separate actions to such a degree.
6. If quotation marks aren't to be typed they shouldn't be used.
7. Allow better access methods between pages - e.g. jump back to a main menu
8. Make the link between the plans and the tasks clearer, perhaps in an introductory page.
9. Make the manual less formal; help them overcome computer anxiety.
10. Number the pages/screens.
11. Don't use '.', just say press the full stop key.
12. Keep the line lengths not too long and not too short - easier to read.

Previous studies have shown that much of the effect of specific typographical attributes on learning are context dependant (Waller, 1979; Schumacher & Waller, 1985) and that more work needs to be carried out in order to form general rules for applying these devices. Without such guides graphic designers rely on their experience of working with text and their appreciation of the purpose of the document to produce a document that 'feels right'. This produces a document that is usually pleasing to the eye, includes features to aid the reader in its use and helps them to access the core information in the document. If the purpose of the document is to aid skill acquisition then it will include spacing and highlighting to guide the reader to the text that the designer feels is necessary for the task. This design approach is underpinned by the assumption that it is the best way of presenting information in order to promote skill. However, it may be that the action of searching for information and reading auxiliary material - elaborations - promotes skill more effectively than jumping straight to the core material highlighted by the designer. The principle of 'cognitive effort' (Craik & Lockhart, 1972), which seems applicable here, has been demonstrated in a number of circumstances as a source of learning. Therefore, although contrary to commonly held beliefs it may be that making it easier to access information is not a good way of promoting skill.

The previous series of experiments showed that elaborations have different effects on the learning of classifications of tasks. Subjective reports, common design principles and psychological research indicate that the format of the text may also affect the learning of the task. The aim of this experiment is to examine the effect of quality of presentation, in terms of the devices employed to improve its typographic presentation, on two different forms of text - unelaborated and heavily elaborated text. There are two main criteria for the texts - that they aid performance while they are available and that they facilitate learning. The experiment will examine both of these using a two by two design - elaborated and unelaborated text, graphically and not graphically designed - in both contexts - with the texts present and after they have been removed.

The design matches the experiment conducted by Wright & Reid (1973) on a greater number of texts than used here, that varied in their layout considerably from each other rather than in the controlled manner used in this experiment. Subjects solve a number of problems using the texts as a source of information. The texts are then withdrawn, subjects then continue to solve similar problems. In Wright & Reid (1973)'s experiment the time spent reading the texts before a correct answer is given is taken as the measure of performance. Here the measure is total time spent before an answer, correct or incorrect, is supplied as it includes both thinking and reading time - the duration of all cognition relevant to the problem solving process.

It is hypothesized that elaborations in a text, whatever the design will slow the subjects response time while the material is present because of the extra material to be read. A better design should reduce the response time. The elaborations are unlikely to aid the performance while the material is present because the plain text is sufficiently simple for the subjects to comprehend and apply without requiring additional explanation.

For the reasons described in the previous experiments, elaborations to the text should aid performance in the recall condition. A good design may improve or worsen performance in the recall condition. Either it will improve recall relative to the poor design because the reader has been assisted in finding the information that they require or it will not improve or worsen relative to the poor design because the reader did not access the material that made up the elaborations. More concretely, the improved presentation of the texts as a result of the manipulation by graphic designers may aid problem solving by improving the visibility of the information salient to the tasks (Gilmore, 1991).

Conway & Kahney (1987) have shown that elaborations can aid skilled performance by providing the mapping relations between previous examples and the current problem situation. Thus elaborations are likely to improve performance here by helping the subject to link the basic information in the text to the specific test questions. Berry & Broadbent (1990) describe the way that elaborations in the form of explanation can aid problem solving, though the elaboration must be provided at the time of problem solving rather than before the task. This matches the form of presentation here, where the explanation follows the core information and it is available at the time of problem solving.

The materials used by Wright and Reid (1973) were so different in presentation to those used in this experiment that it is difficult to predict the results in this instance other than a general drop in performance in the recall section of the experiment. The specific details of the current design and materials followed by the results are described below.

Method

Subjects

Sixty seven undergraduate students took part in the experiment. All volunteered to take part in the experiment after it had been described to their lecture/sporting group.

Experimental Materials

The task used in this experiment is based on the travel agent task designed by Wright and Reid (1973). Fictitious material was invented so that subjects had no option but to read the written information to solve each problem; problems could not be solved from any previous knowledge. Subjects are required to allocate the most appropriate form of transport to travellers with different needs. Their decisions are supported by material that describes the problem space; in Wright and Reid (1973)'s experiment the nature of the presentation of this information was varied - simple English statements, a decision matrix, an algorithm and a prose passage. A Hierarchical Task Analysis of the task of allocating the correct mode of transport was performed resulting in a series of sub-tasks and plans. With the plans laid out on the same page the labels used by Wright and Reid (1973) for the parts of the tables and algorithms were inappropriate as goals for a HTA because they described conditions rather than tasks. For instance 'is time limited' is a condition whereas the goal in the HTA is 'decide about time conditions'. The text was re-labelled accordingly and is shown in Figure 13. This text was constructed without tabulation, highlighting, or the use of spacing with the exception of a blank line used to separate the sub-tasks from the plans. This matched the layout used in the electronic manuals. Graphic designers were then informed of the way that the material would be used and asked to re-design the presentation to suit this purpose - without altering the content. The resulting text is shown in Figure 14.

The original plain HTA text was then elaborated with reasons and examples as in the previous experiments, resulting in a third piece of instructional material - Figure 15. The final text, elaborated and graphically designed is shown in Figure 16.

A series of 30 problems were drawn up, again based on those used by Wright and Reid (1973). For each problem the subject had to specify the most appropriate mode of travel for the situation described in the question. The first 6 questions were simple forms where the terms used to describe the needs matched those in the HTA; these were used as warm up questions. Examples are given below. The remaining tasks were more story-like and required the subject to translate the needs into the terms used in the instructional text. The ordering of distance, cost and time information varied randomly across all problems.

Figure 13
Sample of unelaborated, plain design manual

Plan 1.0 Decide about time constraints on the trip.
If time is limited then decide about cost for a hurried trip. (1.1)
If time is not limited then decide about cost limitations on an unrushed trip. (1.2)

1.1 Decide about cost of the trip.
If your funds are limited then you should travel by Space Ship.
If you are unconstrained by money then travel by Rocket.

1.2 Decide about the cost of the trip.
If your funds are limited then you should consider how far you are travelling on your unrushed, cheap trip. (1.21)
If you are unconstrained by money then you should consider how far you are travelling on your unrushed, well financed trip.(1.22)

1.21 Travelling distance decision.
If you are travelling more than 10 orbs on an unrushed, cheap trip then travel by Satellite.
If you are travelling less then 10 orbs on an unrushed, cheap trip then travel by Astrobus.

1.22 Travelling distance decision.
If you are travelling more than 10 orbs on an unrushed, well financed trip then travel by Super Star.
If you are travelling less then 10 orbs on an unrushed, well financed trip then travel by Cosmocar.

Figure 14

Sample of unelaborated, graphically enhanced text

Plan1.0 Decide about time constraints on the trip.

If time is limited then decide about cost for a hurried trip. (1.1)

If time is not limited then decide about cost limitations on an unrushed trip. (1.2)

1.1 Decide about cost of the trip.

If your funds are limited then you should travel by Space Ship.

If you are unconstrained by money then travel by Rocket.

1.2 Decide about the cost of the trip.

If your funds are limited then you should consider how far you are travelling on your unrushed, cheap trip. (1.21)

If you are unconstrained by money then you should consider how far you are travelling on your unrushed, well financed trip.(1.22)

1.21 Travelling distance decision.

If you are travelling more than 10 orbs on an unrushed, cheap trip then travel by Satellite.

If you are travelling less then 10 orbs on an unrushed, cheap trip then travel by Astrobus.

1.22 Travelling distance decision.

If you are travelling more than 10 orbs on an unrushed, well financed trip then travel by Super Star.

If you are travelling less then 10 orbs on an unrushed, well financed trip then travel by Cosmocar.

Figure 15
Sample of elaborated plain text

Plan 1.0 Decide about time constraints on the trip.
REASONS
The most important decision you must make is about time. Time is either limited or unlimited. You should decide about your time needs first of all because it is no use thinking of cost a a form of travel if it just won't get you there in time. Time is the major deciding factor - it descriminates between the largest number of forms of transport.
EXAMPLE If you only have an hour, don't think of taking transport that will take more than an hour

If time is limited then decide about cost for a hurried trip. (1.1)
If time is not limited then decide about cost limitations on an unrushed trip. (1.2)

1.1 Decide about cost of the trip.
REASONS
Cost is the best way of dividing up the remaining two forms of transport. Cost can either force you to choose a cheap form or expensive form. Space ships are an old form of transport and are inexpensive, whereas Rockets are new and expensive.
EXAMPLE If you are in a rush and only have what happens to be in your pockets go by Space ship. If you are in a rush and you've got the company cheque-book feel free to go by Rocket.

If your funds are limited then you should travel by Space Ship.
If you are unconstrained by money then travel by Rocket.

1.2 Decide about the cost of the trip.
REASONS You must consider the price of your transport. Cost helps you to decide between a number of different forms of transport. Your funds can either be small or large and effectively unlimited.
EXAMPLE If you can take all week to get where you want to go but only have a few coins left from a drinking spree you'll have to budget for your foolishly limited resources.

If your funds are limited then you should consider how far you are travelling on your unrushed, cheap trip. (1.21)
If you are unconstrained by money then you should consider how far you are travelling on your unrushed, well financed trip.(1.22)

Figure 16

Sample of elaborated, graphically enhanced text.

PLAN 1.0 Decide about time constraints on the trip.

REASONS The most important decision you must make is about time. Time is either limited or unlimited. You should decide about your time needs first of all because it is no use thinking of cost a a form of travel if it just won't get you there in time. Time is the major deciding factor - it descriminates between the largest number of forms of transport.

EXAMPLE If you only have an hour, don't think of taking transport that will take more than an hour.

-- If time is limited then decide about cost for a hurried trip. (1.1)

-- If time is not limited then decide about cost limitations on an unrushed trip. (1.2)

1.1 Decide about cost of the trip.

REASONS Cost is best way of dividing up the remaining two forms of transport. Cost can either force you to choose a cheap form or expensive form. Space ships are an old form of transport and are inexpensive, whereas Rockets are new and expensive.

EXAMPLE When you are in a rush and only have what happens to be in your pockets go by Space ship. If you are in a rush and you've got the company cheque-book feel free to go by Rocket.

-- If your funds are limited then you should travel by Space Ship.

-- If you are unconstrained by money then travel by Rocket.

1.2 Decide about the cost of the trip.

REASONS You must consider the price of your transport. Cost helps you to decide between a number of different forms of transport. Your funds can either be small or large and effectively unlimited.

EXAMPLE If you can take all week to get where you want to go but only have a few coins left from a drinking spree you'll have to budget for your foolishly limited resources.

-- If your funds are limited then you should consider how far you are travelling on your unrushed, cheap trip. (1.21)

-- If you are unconstrained by money then consider how far you are travelling on your unrushed, well financed trip.(1.22)

1.21 Travelling distance decision.

Initial Questions

1. Distance is more than 10 orbs, cost is limited, time doesn't matter.
2. Time is limited, the distance is less than 10 orbs, cost is unlimited.
3. Distance is more than 10 orbs, cost is unlimited, time is unlimited.
4. Time is limited, distance is less than 10 orbs, cost is limited.
5. Cost is unlimited, distance is more than 10 orbs, time is limited.
6. Time is unlimited, distance is less than 10 orbs, cost is unlimited.

Examples of Story-like tasks

7. A local commuter with very little money who isn't worried about how long it takes him to get to work.

8. Someone trying to reach a far off holiday resort on a budget but no real need to get there in a hurry.

9. Someone who will make the same local journey regularly and wants to do it in style and without fuss or rush.

10. A merchant trying to cut costs on a long distance trip which has to be carried out very quickly.

11 Someone who wants to pay the least amount possible to travel to a local shopping centre to buy bargains at the sale. Because the shops are open 24hrs a day it doesn't matter how long the trip takes.

12. A Products Manager travelling to spy on a local competitor at an exhibition of goods. He's been told to spend as much as he wants but not to appear unduly worried so he ought to use a slow form of transport.

Procedure

An independent groups design was used, each subject used only one form of the test material. Subjects were tested in groups of between 1 and 10 using networked microcomputers to drive the experiment. The experiment consisted of three parts - the presentation of instructions, questions 1 to 18 where the test material was available and questions 19 to 30 where the material had been removed. The scenario for the experiment was that the subjects were learning a new job from the material but due to a shortage of resources the material may at some time be removed and given to another person. The instructions are given below:

INSTRUCTIONS

Imagine it is your first day as a ticket clerk at the local space port. It is your job to recommend to customers the most appropriate form of travel for their needs.
The company that you work for has a Green policy - so 'most appropriate' means that resources should not be wasted - if they are not in a hurry then they should not be put on fast forms of transport.

Instead of being trained to do the job you have been given the instructions overleaf.

Because so many staff are starting today you have been warned that the instructions will be removed soon so that someone else can learn the job.

Please supply the number representing the most appropriate form of transport in the scenarios displayed by the computer.

You may not go back to questions you have missed; the first answer that you press will be the only one accepted - so think before you press a key ! You are not allowed to make notes.

Thanks for taking part.

Subjects were presented with the instructions while the data recording program was started. When subjects indicated that they were ready to start they were provided with the test material and the computer was instructed to display the first question.

Timing started from the display of each question, which were displayed one at a time on the computer screen. Subjects indicated their answer by pressing a digit from 1 to 6 which corresponded to a form of transport. The list of transport forms and their numbers were displayed with each question; the forms of transport were ordered alphabetically so that it did not correspond with the order that they were mentioned in the test material in order to prevent subjects from using the choice order as a mnemonic. Subjects were not able to correct their answers as the program accepted the fist digit pressed then displayed the next question on a clear screen. The first six questions were of the simple form, the next twelve were story-like.

After the eighteenth question the program paused and subjects were required to call the experimenter who then removed the test material and re-started the presentation of questions.

As well as displaying the questions, the computer program timed how long it took for the subject to respond, recorded and marked their response, then wrote the information to file.

Pilot studies

The experiments ran without problems with the exception that two of the three subjects assigned cheap forms of transport when cost was not limited or fast forms of transport when time was not limited. These errors were prevented by changing the instructions to emphasise that the most appropriate form must be selected.

Results

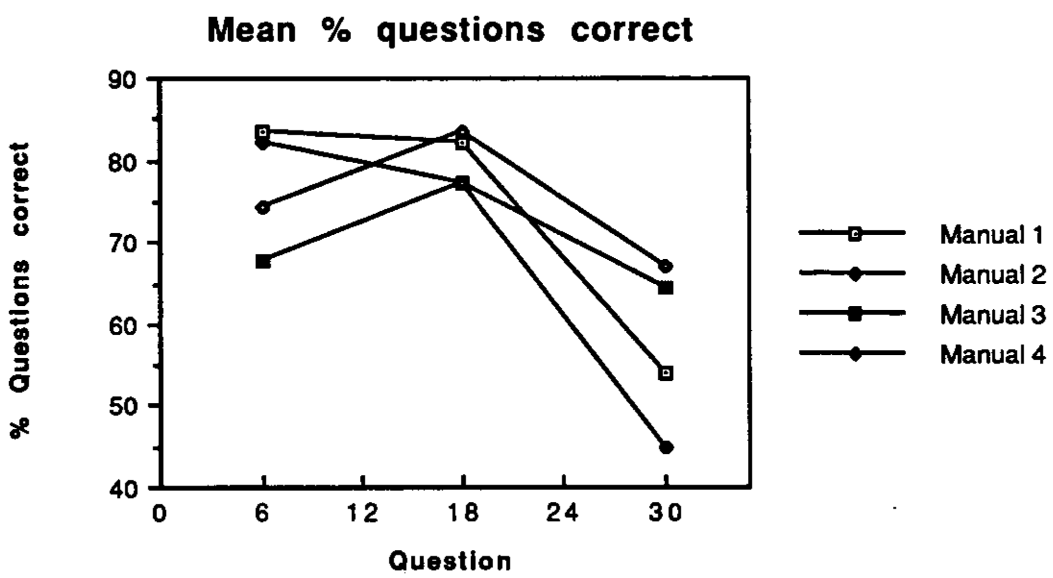
During the experiment an error in the test material was found. The information in one of the examples was contrary to the instructions in a plan. The results of the seven subjects that had used this material were discarded and seven more subjects used instead. It should be noted that six of the subjects did not point out the conflict; it is not known whether they were aware of it.

Two types of measure were recorded - the accuracy of the subject's answers and the time taken to respond.

The first six questions in the study were used as warm up questions; in both this study and judging from the results of Wright and Reid's (1973) work, all subjects spent some initial time familiarising themselves with the material before being able to use it effectively. Though the results for these questions will be discussed the primary areas of interest are for the scores for the next twelve questions while the material is present and for the last twelve questions when the material cannot be used.

Accuracy

The mean percentage of questions answered correctly for each group is shown here in Graph 3.



Note - the lines here are used to indicate the trend in the data and not an exact fit.

An ANOVA shows that there is a significant difference between the groups in the recall section of the experiment, but not in the sections where the text was available. The scores for the groups that used the elaborated material are significantly higher than those unelaborated in the recall section (A two way ANOVA confirms this $p<.005$) but the layout of the material has no effect.

The differences for each group in scores between the recall condition and the section after the warm up where text was available is highly significant with the exception of group 3 where the scores are more similar but follow the general trend of there being a lower score in the recall condition (using t tests $p< 0.01$).

Speed of response

Table 5
Time before any answer given

Question range	Mean response time (s)			
	Manual 1	Manual 2	Manual 3	Manual 4
1 to 6	61.4	56.4	80.4	85.6
7 to 18	32.1	30.7	30.4	29.3
19 to 30	20.7	15.1	17.3	15.9

Examination of times for subjects to respond to questions shows that the scores of the groups are significantly different ($p < 0.01$) on the warm up questions, post hoc tests indicating that the elaborated groups are significantly slower than the unelaborated groups . This is to be expected when the amount of extra material to be read is considered.

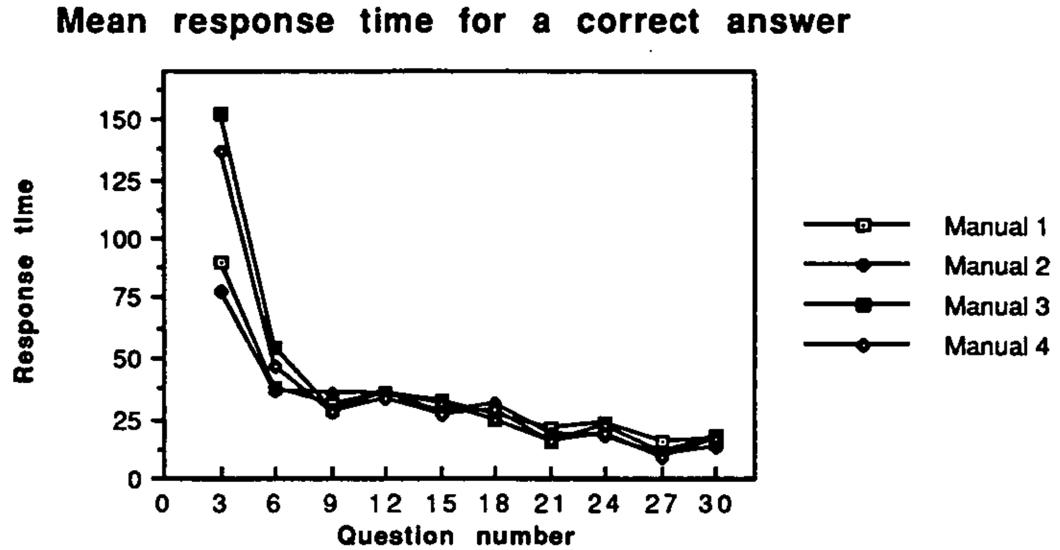
Examination of the speed for a subject to respond correctly to a question shows that this difference is reduced though the trend still exists. See Graph 4.

Across all groups subjects respond more quickly in the recall condition, however there are no significant differences between the groups in the speed of response (either overall or only for correct answers) in the recall section.

Non empirical results.

Many people expressed surprise when the texts were removed, and expressed reservations about their ability to perform without the texts. They confirmed that they had understood from the instructions that the texts may be removed but had not really anticipated the fact.

Graph 4



A small number of people had initial difficulties in using the texts but after reading the instructions again became more confidant.

Some subjects performed very quickly and finished with low scores. When questioned they stated that they had been using the material carefully. Conversely some subjects worked slowly and with high precision.

A small number of subjects used mnemonics to help them, in particular associating a form of transport with an existing transport form. Others attempted to use the position of the instructions on the page as an aid to recall though these frequently reported being confused or making consistent errors in the recall stage. Some tried to impose patterns on the order of the presentation of the choices - e.g. remembering that the fifth one was always the one used for short cheap trips.

Discussion

It was evident that subject's use of different strategies to aid them to perform the tests was not localised to one particular group whereas some clear patterns emerged in the empirical data.

Subjects took longer to read the elaborated texts than the unelaborated ones during the warm up questions. The layout of the texts did not affect the reading speed. After the first few questions the mean speed for any answer to be given and the mean speed for a correct answer was similar across all groups and slightly less in the recall condition. This may be caused by the subjects initially familiarising themselves with the texts during the first few questions, then accessing the information relatively quickly, whatever the text, for the rest of the questions. This could be confirmed by a small number of eye movement tracking studies. In the recall section they do not need to access the information in the text so the speed drops again.

Analysis of the scores shows that those subjects who used the elaborated text initially performed significantly better in the recall section than both of the unelaborated groups. While the text was present the scores were very similar. This indicates that though extra text did not help performance while it was present it did aid performance in the recall section - it improved learning.

Better layout aided subjects only during the initial reading of the material - after the warm up questions the presentation of the material did not have an effect.

These results illustrate one facet of the potential benefits of graphical design to the design of instructional materials. The tests produced without regard to graphic design were not as effective during initial use as those enhanced by graphic designers. The method of text design was not formally defined; human factors guidelines were not sought or quoted. It is unlikely that the polished presentation that they produced could have been attained by following guidelines alone; (Eco, 1976; Forrester & Reason, 1991) suggest that a good design solution may be inherently unpredictable rather the less formalisable proficiency of experienced designers is required. Carroll & Kellog (1989) approaches this issue from studies of the progress of HCI researchers towards innovative design. Although HCI can produce guidelines for the development of material it is the insight of designers that produces radically new and effective solutions to interaction problems. This experiment confirms that HCI produces a sound basis for the design of materials but also that graphic designers can apply skills to enhance the materials further. In the specific context of this series of experiments this study shows that the graphical presentation of the materials did not detract from their effect on skill acquisition, apart from during the initial period of exposure.

Subject's satisfaction with and enjoyment of the appearance of the materials was not measured in the experiment. During the debriefing of the subjects some expressed a preference for the graphically designed materials. The initial ease of use of the materials due to layout may be a strong determinant of whether a manual is used or not. Mahony & Gower (1991) p 268 argue that ' ... when users find something ugly it is likely to affect their comfort and ease of use.'. Subjects in the experiments described here are required to use the manual if they need help and so this factor does not play a significant part in the experiment. In the work environment people may reject otherwise useful manuals because of their poor aesthetic appearance. The difficulty of accessing information is likely to lead to the decision that using the manual takes more effort than it is worth (Eason, 1976, 1981). This factor will need to be examined in further work.

The experiment has confirmed that within the context of the series of studies described in this thesis the minimal amount of graphical design applied to the texts has not been a confounding factor in the results.

The usefulness of the application of the skills of graphic designers to interaction systems is discussed - in the current context the text produced by the designer is easier to use at first but this is a short term effect. The elaboration added to the text slows initial performance becasue of the extra amount of text to be read but improves performance on recall tasks.

The next chapter examines the effect of the elaborations on skill acquisition using a larger number of subjects with a design improved by the findings of the preceding studies.

CHAPTER 9

EXAMINATION OF THE EFFECT OF TEXT ELABORATION ON SKILL ACQUISITION

Introduction

The preceding chapters have examined the issues surrounding the means of improving manuals for learning by the addition of elaboration and methods for studying the effect of the elaborations on skill. Different types of elaboration were selected from the literature and incorporated into standardised texts. Through the development of a model of skill and a study of the validity of different skill measurement techniques a battery of tasks was constructed to examine the effectiveness of the manuals. Pilot studies of the design revealed aspects of the material, the method and the design of the experiment that needed to be reviewed. One of these, the graphical design of the texts, was examined in detail through an empirical study of its effect on performance. It was found that the design of the text had an initial effect on the speed of correctly completing tasks but the effect disappeared after the first few tasks.

This chapter is an improved evaluation of the elaboration techniques based on the findings described above. It aims to examine the texts in a more controlled manner than before while retaining the essential features of the applied context under study. It is expected that the careful design of the experiment should reduce the cause of variations in subject performance down to the effect of the different manuals and the normal variation between subjects. The specific hypotheses about the effect of the manuals are described in Chapter 6.

Subjects

Seventy undergraduate students aged between 18 and 24 took part in the study. All volunteered to take part because they wanted to learn to use electronic mail. None of the subjects had any previous experience at using electronic mail; all had limited computing experience with the exception of some who had developed word processing skills. Although subjects were selected as groups from particular lecture courses they were distributed evenly across the experimental groups.

Design

The design of this experiment is the same as in the previous pilot studies with the exception of the way that the manuals were allocated to subjects in the re-test section of the experiment. The pilot studies showed that the manual used in the re-test section had in some cases considerable effect on the subjects performance. This was because some manuals enabled subjects to obtain information more quickly than others; for instance the 'core' manual. This meant that any group using the manual would be able to answer questions quickly because they could obtain the information more quickly rather than because of their skill level. It was therefore decided to balance the use of manuals in the re-test section within each group.

After the initial testing of a group of subjects they were ranked according to their time to complete the last task as an indication of their skill level. In one or two cases where subjects made obvious slips in their performance on the last task the ranks were altered to reflect their ability. The group was then distributed evenly across five manual groups according to this measure of ability. The diagram below illustrates this change to the design.

Figure 17 New Design

		Initial-Learning				RE-TEST	Initial-Re-test			Close-Transfer			Non-Specific		
MANAUL	SUBJ	Q 1	Q 2	Q 3	Q 4	MANUAL	Q5	Q10	Q12	Q6	Q8	Q13	Q7	Q9	Q11
Core Manual						Core									
						HTA									
						HTA-Reasons									
						HTAExamples									
						HTA-Standard									

This allows the balancing of the re-presentation method across the groups but it also changes the nature of the experiment. Instead of the experiment being a usability test of a particular manual it is more in line with a traditional transfer of learning design where the stimuli are carefully balanced and rotated through all conditions. This change also alters the nature of the task that the subjects have to perform. Instead of becoming familiar with a particular type of manual then using it as necessary in the subsequent tests, 80% of the group have to familiarise themselves with a different manual in the re-test section as well as carrying out the tests. Since this task is balanced across all groups and there is no reason to suppose that one type of manual will assist subjects to adapt to another manual better than the others the effect of this task can be ignored for the purposes of analysing the data. However although it is also possible to pick out the cells where subjects do not need to learn a different manual in order to assess the effect of changing manuals.

Method

With the exception of the allocation of subjects to sub groups according to their scores on the initial learning tasks the method is the same as in the pilots.

Materials

The materials, including the tasks, the method of elaboration and the electronic manuals are described in Chapter 7.
The particular forms of elaboration used in this experiment are:

core manual,
hta only manual,
hta + reasons,
hta + examples,
standard manual.

The final version of the instructions and method of providing help to subjects is described in Chapter 7

Results

Ten of the subjects were not included in the final data set and were replaced by other subjects. Reasons for excluding them included subjects forgetting the second part or being unable to attend the second part, and subjects who felt that their performance was substantially effected by illness or a hangover

Qualitative Results

Subject's comments were very similar to the comments during the pilot experiment, with the following additions.

Some subjects reported that the examples made the situation clearer for them. Often the examples were not read once the subject had completed the task for the first time.

Some subjects felt that the reasons were useful for parts of the task, for instance remembering that 'r' stood for 'read', but that in many cases the reasons were quite trivial and were read only once, if at all.

None of the subjects complained about having to use a different manual during the re-test section.

Other observations of the subject's behaviour will be discussed later in this chapter.

Quantitative Results

Examination of the raw data showed that there was a considerable amount of variation within each group across all measures. This made it difficult to examine the experimental hypotheses using parametric statistics - a table of the means and standard deviations of the time taken to complete tasks is shown in Table 6.

Table 7 is a sample of the data, a table of the times for subjects to complete tasks with unexpected items highlighted. These included subjects with times for a task that were unexpectedly high or low for the group (e.g. 40 or 800 seconds instead of a more common value of 180 seconds), subjects that made unexpected high/low times for a particular task in relation to the rest of their performance and subjects whose results differed considerably from those of the rest of the group.

Although non-parametric methods of analysing the data or the application of a classification or filter method to the data could be used in such circumstances this would be inappropriate here. The function of such devices is to pull the data into the sort of pattern that can be analysed in this type of factorial design. However examination of the subject's use of the manual, the transcript of their actions and observations of their performance indicated that there were specific reasons for the high variation in the data rather than just a random variation that could be filtered out.

Table 6.
Means and Standard Deviations for times to complete tasks

		Initial-Learning				RE-TEST Initial-Re-test			Close-Transfer			Non-Specific		
		Q1	Q2	Q3	Q4	Q5	Q10	Q12	Q6	Q8	Q13	Q7	Q9	Q11
Core Manual	Mean	168.00	113.87	86.07	91.47	266.40	69.00	71.47	352.87	201.53	140.87	371.87	194.80	146.27
	SD	51.91	37.98	32.71	63.72	252.95	13.91	27.65	184.57	97.17	53.00	209.37	89.15	61.82
HTA	Mean	477.53	270.73	173.27	107.47	166.53	78.71	74.00	312.00	176.80	144.80	362.27	159.20	129.87
	SD	172.74	105.85	87.61	52.33	94.03	37.46	31.07	177.30	71.95	54.18	297.38	41.22	61.78
HTA-Reasons Manual	Mean	512.27	233.73	145.80	85.93	126.73	72.47	68.80	328.60	179.33	131.13	284.33	190.73	135.93
	SD	176.66	124.59	80.96	34.85	65.07	29.18	21.03	189.98	117.12	44.93	156.52	119.32	73.43
HTA-Examples Manual	Mean	512.47	275.13	148.53	108.07	128.40	67.07	71.47	272.80	147.13	118.60	252.47	148.40	127.73
	SD	114.96	122.78	107.96	84.44	75.23	26.48	34.19	119.85	51.63	42.06	91.29	29.43	55.75
HTA-Standard Manual	Mean	648.60	263.87	138.93	106.53	155.73	81.47	77.00	332.33	182.00	140.20	314.53	190.87	129.40
	SD	179.67	149.57	93.05	59.27	94.05	40.77	30.90	136.20	92.77	82.44	140.18	97.36	64.80

	SUBJ	Initial-Learning Q1	Q2	Q3	Q4	RE-TEST PRESENTATION	Initial-Re-test Q5	Q10	Q12	Close-Transfer Q6	Q8	Q13	Far-Transfer Q7	Q9	Q11
Core Manual	Sheila	280	56	53	40	Core	46	42	48	109	86	78	99	89	100
	Tracy	175	112	100	80		80	57	82	111	82	92	125	97	74
	Mark	163	115	145	266		102	81	74	184	176	98	151	94	69
	Carol	102	51	45	41	HTA	160	60	36	184	92	89	220	134	67
	Steve	123	120	70	72		86	67	83	589	186	134	521	236	161
	Gavin	382	197	109	96		447	80	56	564	166	136	276	170	122
	Bridg	141	100	78	50	HTA-Reasons	448	70	60	413	340	141	281	162	125
	Alison	185	147	79	72		85	80	96	433	222	187	227	169	119
	Julie	181	148	154	131		565	82	79	346	232	146	636	164	194
	Nick	109	89	53	51	HTA-Examples	252	54	54	156	158	135	242	178	163
	Kate	169	81	65	67		65	57	53	562	165	72	368	242	94
	Jane	216	104	74	209		83	91	75	675	213	187	374	193	225
	Tobias	112	128	68	46	HTA-Standard	401	62	50	256	150	139	465	192	226
	Sue	140	108	110	96		719	64	66	316	449	362	567	409	344
	Louise	171	152	87	62		176	66	148	413	865	225	673	353	207
HTA Manual	Paul	321	415	83	65	Core	72	53	47	100	114	108	131	65	50
	Lee	365	155	82	77		153	56	72	147	105	27	107	38	46
	Dierdr	675	423	362	233		178		180	314	209	192	158	151	111
	Lesley	211	231	92	69	HTA	44	29	38	145	67	72	185	141	195
	Vanessa	437	211	123	72		64	62	67	221	172	108	230	136	112
	Joelle	465	489	298	158		343	166	132	458	277	174	337	214	171
	Mand	394	227	97	63	HTA-Reasons	210	73	70	744	165	150	309	201	133
	Sheb	908	309	161	73		161	77	56	237	142	98	330	155	128
	Rolf	656	258	237	169		222	72	45	160	197	162	1201	165	148
	Richard	222	137	56	77	HTA-Examples	311	44	44	478	103	128	348	116	65
	Dhim	555	269	265	78		86	130	97	443	344	169	631	191	304
	Janet	465	434	186	126		411	142	120	380	327	297	344	234	158
	Vanes	412	155	166	134	HTA-Standard	314	74	69	218	194	146	321	171	127
	Tracey	564	276	150	115		79	65	62	413	182	169	656	196	134
	Karen	405	182	163	103		148	57	62	207	143	110	257	144	96
HTA-Reasons Manual	Mark	294	44	46	43	Core	89	64	53	86	100	72	61	67	51
	Neil	482	160	76	59		69	57	45	128	160	69	126	64	100
	Simon	458	341	164	79		77	63	49	122	90	94	108	63	67
	Darren	278	63	46	47	HTA	50	45	43	201	177	106	216	170	142
	Steve	630	269	101	73		315	70	101	186	112	78	226	119	121
	Louise	367	130	258	59		74	54	48	302	162	172	213	130	106
	Peter	451	273	162	69	HTA-Reasons	171	73	69	716	172	148	260	139	135
	Sue	457	244	79	86		136	66	88	312	114	107	283	162	80
	Cath	750	222	175	110		195	63	54	255	317	181	268	306	148
	Steph	455	182	81	67	HTA-Examples	75	62	68	237	96	94	315	185	152
	Darryl	882	348	185	105		162	127	82	219	199	302	239	243	322
	Susie	582	356	269	108		231	61	71	263	147	109	260	164	158
	Sarah	549	242	256	122	HTA-Standard	315	168	103	707	557	304	418	525	145
	Franc	314	108	71	84		78	47	56	405	113	133	370	196	51
	Adrian	725	486	232	172		47	101	96	481	284	171	722	301	341
HTA-Examples Manual	Andrew	377	62	41	36	Core	104	24	32	67	97	85	60	80	32
	Michelle	584	318	129	86		89	47	38	110	107	60	115	116	75
	Steph	716	349	336	262		96	94	89	141	144	126	121	150	97
	Sarah	472	283	90	56	HTA	58	36	29	409	104	75	192	127	250
	Robin	649	236	110	81		102	84	104	294	179	169	284	179	145
	Donna	696	697	325	256		251	154	175	282	391	302	348	155	118
	Robin	469	263	54	57	HTA-Reasons	64	48	65	227	100	74	363	143	119
	Lex	402	275	79	72		112	73	53	179	123	132	257	177	342
	John	545	230	78	59		69	63	80	406	176	111	233	118	113
	Nick	438	175	76	70	HTA-Examples	64	48	65	227	100	74	363	143	119
	Garet	459	328	175	76		72	62	53	314	162	122	300	165	129
	Alison	489	273	127	167		267	80	68	302	143	184	237	171	122
	Adam	514	162	41	59	HTA-Standard	177	46	44	115	126	120	330	139	86
	Graham	526	276	202	91		199	72	68	371	202	112	336	155	123
	Leann	449	289	231	116		140	86	80	450	163	133	326	209	136
HTA-Standard Manual	Colin	496	59	49	46	Core	68	41	45	115	91	62	86	83	61
	Russell	392	239	168	91		71	59	64	103	116	72	169	124	63
	Helen	761	310	336	160		136	74	63	234	123	119	128	131	82
	Ravi	522	135	69	65	HTA	65	56	46	367	117	112	241	123	50
	Parm	1017	480	183	90		280	85	91	265	215	124	282	292	209
	PJ	656	334	305	209		374	153	137	362	358	Ω	347	358	168
	Tracey	429	131	62	67	HTA-Reasons	55	61	61	366	167	140	220	136	41
	Louise	591	391	109	78		555	71	79	346	223	153	342	157	86
	Colert	852	910	394	234		374	187	145	486	266	506	373	228	217
	Sabina	709	164	96	70	HTA-Examples	91	64	67	229	119	75	289	129	89
	Vijay	762	100	77	76		165	58	63	291	46	63	345	169	116
	Sarah	651	83	64	80		66	50	49	407	148	164	406	167	90
	Haidr	576	405	94	73	HTA-Standard	191	71	58	306	176	78	467	186	127
	Clare	456	271	95	74		171	71	95	392	209	122	361	201	156
	Mona	636	466	128	165		117	281	88	646	422	196	846	459	365

Table 7. Unexpectedly fast and slow times (highlighted in grey)

The following instances of behaviour occured that were responsible for some of the unusual values recorded:-

1. The chance finding of the 'insert a file' page in the manual while attempting a different task. Knowledge of the command and its existence in the manual tended to make it easier for subjects to look up the command when they required it.

2. Accidents/difficulties when navigating through the manual led them to completely the wrong section. Frequently it took them a long time to realise that this had happened; occasionally they attempted to execute commands that had nothing at all to do with the current task.

3. Accidentally missing steps in a task leading to errors that they could not see the cause of. On some occasions this became particularly bad when subjects repeatedly followed the necessary sequence of sub tasks with the exception of the same sub task over and over again. This is similar to the spiral of errors found in other detailed studies of learning (Mack,Clayton, & Carroll, 1983; Carroll & Carrithers, 1984).

4. Not understanding what it meant to insert a file. This led to unusual searches through the manual or guesses at the point in the sub task sequence when the file could be inserted.

5. Extreme hesitancy after making an error. Some subjects who up until the point of a simple error were very competent and did not refer to the manual became hesitant and relied on the manual for every step.

6. Individual strategy in using the manual. Some subjects preferred to follow the manual for every step rather than try to memorise the task or rely on their own judgement. Frequently the subjects checked their knowledge before applying it because of a fear of making a mistake.

manual and aspects of the materials that were distorting the experimental setting. Although these factors have been removed the subject's behaviour and thus the data is heavily dependant on factors within the experimental setting that are part of the applied area under study, overshadowing any clear effect within the factorial design due to the experimental manipulations.

The task cannot be further constrained without losing the basic nature of the problem area. The most realistic test of a manual would be to supply subjects who wanted to learn to use an application package with manuals containing the different forms of elaboration and to measure their skill at a later date. Factors such as the individuals' access to other forms of help, the sequencing of the tasks that they need to learn and the subject's time and motivation to learn the package may overwhelm the effect of the elaborations. To use a single page text as the manual and to test the subject's recall of the material after one reading as a measure of its effectiveness for aiding skill acquisition is, as has been described previously, to ignore critical aspects of the situation. To have allowed the subject to have failed on tasks would not have helped - the subject may have tried to resolve intermediary failures or become lost in the manual, just as in this experiment where they have to succeed on each task.

This experiment has been designed as far as possible to exclude the major factors of the work setting that would undermine a factorial design. The task still contains opportunities and aspects where subjects can depart from the ideal path of reading information, assimilating it and applying it. The effect of the materials on the subjects behaviour when faced with such opportunities is just as important as the intended effect on the learning process.

For instance if the nature of the manual had ensured that they followed all of the plans carefully less errors would have ocurred, in particular the error due to missing out a step that subjects found particularly difficult and frustrating to correct. The manipulations examined here were not aimed at this outcome but may by chance have aided learning by preventing this sort of error in the cases where subjects would have otherwise made mistakes.

This is the second aspect to the problem - although the physical task of typing the necessary key strokes is relatively simple, the process of using the manual, understanding the instructions, applying them and learning how to perform the task is extremely complicated. As can be seen from the observations above, the descriptions of the results of the pilot experiments and the subjective reports of the difficulty of the tasks, this is not a straightforward choice - reaction time experiment. Whatever terms are used to describe the process there are a large number of steps in the process. At each step the subject may or may not follow the shortest path to the solution of the problem, depending on their previous experience and the nature of the materials they are supplied with. Because some of the paths take far longer than others (e.g. repeating the whole task to correct a mistake) and the path taken by the subject through the possible alternatives is not strictly determined by the starting materials the pattern of results observed in this experiment occurs.

If a subject strays from this theoretical optimum path the consequent behaviour will not always be slower or less accurate. Although time consuming at the time the accident may lead to an increased level of performance through the acquisition of knowledge that may be useful in later circumstances or help to resolve conflicts in poorly assimilated knowledge.

Since the applied problem cannot be further simplified and a factorial design is not an appropriate method of studying such complex behaviour an alternative procedure has been formulated. Detailed examination of the subject's behaviour and subsequent debriefing may be used to find the 'happy accidents' during the manual and application package use. Analysis of the nature of the materials leading to the accident and the antecedent history of the subject should lead to the selection of both the materials and the necessary antecedents for more successful learning to occur. Once these have been incorporated into materials the test can be repeated, leading to more learners having positive learning experiences. This is detailed iterative piloting rather than one shot experimental design.

An analysis of one type of critical instance in this experiment has been made in order to examine the usefulness of this approach.

Re-evaluation of the data

Wright & Monk (1989) argue that an account of the use of a system does not provide sufficient information for a full analysis of many usability problems. During the experiment described above a large amount of information was collected about the users behaviour as well as the time it took them to correctly complete a task. This included a record of the information that they accessed in the manual (the cards), the instances where they accessed information about the keys they needed to type (the 'type' cards), notes made by the experimenter during the experiment and notes from the debriefing of subjects.

One critical incident that was recorded several times was the accidental finding by subjects of the method of inserting files into messages. 'Accidental' refers to the action of accessing this information in the manual while engaged in a task that did not require it. It was noted by the experimenter and stated by several subjects that knowledge of the existence of this information aided them in finding the information again and carrying out the task it described. The converse, that some subjects who did not stumble upon these pages before they needed to find them frequently found it difficult to locate the necessary information when they needed it.

The positive result of this critical incident is that the subject should be able to perform the close transfer task, inserting a file in a message, more quickly that other subjects because they did not need to search for so long to find the information that described this procedure.

The 975 files containing the records of the subject's use of the electronic manual were searched for the string 'insert' - this is found in the name of the card describing the procedure for inserting a file. The subjects who accessed this card during either the first four tasks or the first re-test (none of which required the use of this card) were picked out. Subjects who looked at this card only on later occasions were judged not to have had this particular critical incident because by that time they would have been required to use that card to perform a task.

Results

Tables 8, 9, and 10 contain the means and standard deviations for critical and non-critical incident subjects in each manual group for each question, showing the times per task, number of cards and number of 'type' cards examined. Subjects in the core condition could not accidentally access an 'Insert file' card and so were not included in the analysis.

Seven incident subjects were found in the HTA Manual condition, 2 in the Examples Manual condition, 6 in the Reasons Manual condition and 5 in the Standard Manual condition. It is not possible to isolate the incidents to one particular type of manual condition.

The incident subjects are slower on the initial tests, recall tests and close transfer tests in the Reasons Manual, Examples Manual and Standard manual.

They also access more cards on during the same tasks in the Examples and Standard manuals.

More 'type' cards are used by incident subjects in the Standard Manual condition on all tests except the non-specific test and to some extent by the HTA Manual subjects on the initial learning tasks.

In all other condition and task combinations the non-incident and incident subjects performed in the same way.

All of the subjects had the same level of performance on the non-specific transfer tests.

Discussion

The results are the opposite of those predicted. It was expected that exposure to the 'Insert a file' information before it was needed would help the subjects to re-locate the information and perform the tests. Instead they performed more slowly on the tests where this information was required in the Reasons, Examples and Standard Manuals. They also looked at more cards (including 'type' cards) in the Examples and Standard conditions for these tests.

Table 8
Means and Standard Deviations of Times of critical- and no critical- incident subjects

			Initial learning				Re-Test								
							Initial re-test			Close-transfer			Non-specific		
			Q1	Q2	Q3	Q4	Q5	Q10	Q12	Q6	Q8	Q13	Q7	Q9	Q11
HTA	Incident	Mean	571.83	293.33	187.67	111.67	163.17	85.83	72.17	391.33	189.33	142.00	525.33	177.83	139.33
Manual		SD	188.31	97.83	69.43	52.11	76.09	39.66	31.10	211.40	46.65	30.93	407.95	30.37	19.47
	No incident	Mean	414.67	255.67	163.67	104.67	168.78	73.38	75.22	259.11	168.44	146.67	253.56	146.78	123.56
		SD	137.62	113.96	100.81	55.43	108.81	37.51	32.87	138.60	86.62	67.30	132.32	44.31	79.56
Reasons	Incident	Mean	696.50	346.25	191.75	119.75	159.00	110.50	96.00	423.00	288.00	163.75	441.25	298.75	209.75
Manual		SD	143.06	112.71	67.26	43.87	78.52	33.39	9.56	224.71	192.59	59.13	206.14	173.04	99.65
	No incident	Mean	445.27	192.82	129.09	73.64	115.00	58.64	58.91	294.27	139.82	119.27	227.27	151.45	109.09
		SD	137.98	104.96	81.65	22.35	59.25	8.26	13.73	174.71	40.44	34.60	90.10	67.70	39.97
Examples	Incident	Mean	593.50	450.00	227.50	211.50	284.00	107.00	121.50	292.00	217.00	193.00	242.50	163.00	125.00
Manual		SD	147.79	250.32	142.13	62.93	4.24	38.18	75.66	14.14	104.65	12.73	7.78	11.31	9.90
	No incident	Mean	500.00	248.23	136.38	92.15	104.46	60.92	63.77	269.85	136.38	107.15	254.00	146.15	128.15
		SD	111.06	60.53	103.52	77.01	44.10	19.75	20.13	129.11	35.47	31.41	98.48	30.96	60.14
Standard	Incident	Mean	753.00	381.00	234.40	153.40	215.00	114.00	107.00	341.80	221.00	195.60	302.20	220.00	174.20
Manual		SD	214.38	107.00	103.62	70.63	61.57	52.77	34.98	97.63	64.90	115.81	104.66	73.98	54.81
	No incident	Mean	596.40	205.30	91.20	83.10	126.10	65.20	62.00	327.60	162.50	112.50	320.70	176.30	107.00
		SD	143.90	135.19	33.20	37.64	95.64	21.58	13.85	156.67	101.23	45.32	159.86	107.73	59.36

Table 9
Means and Standard Deviations of number of 'Cards' of critical- and no critical- incident subjects

			Initial learning				Re-Test								
							Initial re-test			Close-transfer			Non-specific		
			Q1	Q2	Q3	Q4	Q5	Q10	Q12	Q6	Q8	Q13	Q7	Q9	Q11
HTA	Incident	Mean	44.83	31.33	17.50	4.33	12.00	2.67	1.67	30.17	14.50	9.67	44.67	24.00	15.50
Manual		SD	14.70	6.15	9.85	7.09	16.48	5.13	4.08	15.41	6.12	5.75	30.28	3.41	6.12
	No incident	Mean	30.78	26.78	13.33	4.00	12.67	0.67	0.67	17.50	10.67	9.17	31.17	21.83	14.50
		SD	2.77	9.38	11.19	6.08	11.64	1.03	1.03	6.47	4.93	3.76	5.86	1.60	8.62
Reasons	Incident	Mean	33.25	21.25	6.00	1.50	8.25	3.25	0.50	15.33	12.00	8.00	30.75	28.25	19.00
Manual		SD	6.18	6.70	4.55	2.38	12.28	6.50	1.00	10.97	10.95	7.62	4.57	18.15	14.28
	No incident	Mean	30.64	13.09	7.27	1.27	6.00	0.00	0.00	19.50	6.38	5.88	29.13	21.25	11.75
		SD	10.74	12.30	9.47	3.00	6.26	0.00	0.00	7.62	4.50	6.64	4.32	9.18	9.10
Examples	Incident	Mean	36.00	30.00	16.50	16.00	28.50	6.00	8.00	26.50	20.50	12.50	31.00	24.00	17.00
Manual		SD	11.31	14.14	20.51	14.14	2.12	8.49	11.31	4.95	16.26	4.95	1.41	0.00	1.41
	No incident	Mean	28.69	18.77	8.23	3.08	3.90	0.50	0.20	16.30	6.40	5.50	27.50	18.90	14.60
		SD	2.66	10.52	11.59	7.70	7.28	0.85	0.63	7.01	3.84	4.12	3.98	4.53	8.17
Standard	Incident	Mean	31.60	23.40	16.60	9.00	18.50	9.00	7.00	32.00	23.50	16.50	34.75	27.75	22.00
Manual		SD	2.61	10.29	14.24	12.45	9.98	10.89	8.08	17.94	11.90	11.90	7.97	9.46	2.71
	No incident	Mean	29.40	10.60	1.50	0.20	7.13	0.00	0.00	16.63	7.25	4.50	28.75	18.50	5.63
		SD	4.70	11.15	3.75	0.63	11.72	0.00	0.00	6.39	4.13	3.85	5.20	6.32	6.05

Table 10
Means and Standard Deviations of number of 'Types' of critical- and no critical- incident subjects

			Initial learning				Re-Test								
							Initial re-test			Close-transfer			Non-specific		
			Q1	Q2	Q3	Q4	Q5	Q10	Q12	Q6	Q8	Q13	Q7	Q9	Q11
HTA	Incident	Mean	12.00	7.50	3.50	1.67	2.67	0.33	0.17	6.33	2.67	1.67	13.67	6.67	3.83
Manual		SD	3.10	2.07	2.43	2.66	3.78	0.82	0.41	2.42	1.63	0.82	8.89	1.37	1.72
	No incident	Mean	7.78	5.67	2.44	0.33	2.44	0.11	0.00	3.89	2.67	2.00	9.44	6.00	4.00
		SD	0.97	2.50	2.88	0.71	2.55	0.33	0.00	1.17	1.22	0.71	1.88	1.00	2.87
Reasons	Incident	Mean	7.50	3.75	0.25	0.25	1.00	1.00	0.25	3.50	2.25	1.33	9.50	8.25	4.25
Manual		SD	2.08	2.36	0.50	0.50	2.00	1.41	0.50	1.73	2.63	2.31	2.65	4.99	4.79
	No incident	Mean	7.09	2.55	1.45	0.18	1.45	0.09	0.00	4.00	2.18	1.36	9.18	5.55	3.00
		SD	2.77	2.70	2.46	0.60	1.57	0.30	0.00	1.73	1.99	1.12	2.18	3.24	2.68
Examples	Incident	Mean	7.50	6.50	3.50	4.00	7.00	1.00	1.00	6.50	4.50	2.00	9.00	6.50	4.50
Manual		SD	2.12	2.12	4.95	2.83	0.00	1.41	1.41	2.12	2.12	1.41	1.41	0.71	0.71
	No incident	Mean	6.54	3.77	1.38	0.31	1.15	0.31	0.00	3.77	1.77	1.62	8.62	5.62	3.77
		SD	1.05	2.55	2.10	0.85	1.86	1.11	0.00	1.09	2.09	2.43	1.80	2.29	3.30
Standard	Incident	Mean	6.80	4.60	3.20	1.80	3.40	1.20	1.00	6.20	5.00	2.60	9.20	6.60	4.40
Manual		SD	0.84	3.21	3.96	2.49	2.97	1.79	1.41	2.59	2.45	1.82	2.28	2.30	1.95
	No incident	Mean	6.70	2.10	0.10	0.00	1.00	0.00	0.00	3.30	1.50	1.20	8.40	5.00	1.30
		SD	1.70	1.91	0.32	0.00	2.00	0.00	0.00	1.49	1.18	2.10	1.84	1.76	1.42

A possible reason for these unexpected results can be seen in the results of the incident subjects on the initial tests. The results are the same as those for the close transfer task - all incident subjects perform worse that the non-incident subjects. This indicates that the subjects encountered the 'Insert a file' card because of their strategy of using the manuals - examining more cards than the non-incident subjects. This strategy also leads to greater times to complete task. The result does not extend to the non-specific tests to the same extent because all of the subjects had difficulty in remembering this task and had to resort to the manual frequently.

The model used in this thesis does not explain why this effect is limited to the Reasons, Examples and Standard condition. Since one aspect of elaborations is that they aid the application of knowledge to unfamiliar problems by helping subjects to understand the new problem it may be that subjects are encouraged to explore in these manuals because they are able to understand the text more fully.

The model of the user and the factors that may affect their behaviour which was used to construct the different manuals has not led to clear results. As discussed above this is due to critical incidents in the learning process that were not anticipated in the model or in the design of the experiment. It is not possible to exclude the incidents form the experiment without losing the applicability of the studies to the actual work situation. An examination of one type of critical incident revealed a particular strategy used by subjects, contrary to the effect predicted by this thesis. A more sophisticated model of the manual use for learning is required and a means of studying the effect of materials on the pattern of behaviour and critical incidents. Pattern matching using neural nets has been used (Finlay & Harrison, 1990) to study deviations from models of paterns of interaction. This method would be a useful approach to take to the further study of this area.

CHAPTER 10

DISCUSSION

The central theme of this thesis has been to capture the complicated nature of the problems of the casual user, to use a model of skill to sift the literature relevant to problem of their training and to produce ecologically valid studies of the means developed to improve skill acquisition.

Although the source of the descriptions of casual users is the same as those used by many other authors in the area its treatment according to this central theme distinguishes this review from previous reviews. The review focuses on the training requirements of the users as indicated by their use of computers at work and on the aspects of their behaviour that makes the current design of manuals unappealing and unsuited to their problems. Carrying through these qualities of the applied problem to the criticism of previous studies led to a criticism of many of the experiments because authors chose to constrain their designs rather than incorporate the features.

That is not to say that the studies conducted in this thesis have managed to match all of the aspects of the casual user's situation. In all studies the user has no option but to attempt the computing tasks (unless they retire from the experiment), the only source of information is the manual supplied by the experimenter and the subject is expected to work until the task has been successfully completed. From the results of the final study it can be seen that the exclusion of these factors has reduced the number of critical incidents affecting the results, though in this case the results still reflect a large variation of behaviours by the subjects.

The validity of different tests of skill was examined by comparing the results of a battery of tests with the scores attained on a typical work task. This approach revealed practical problems of testing and an awareness of the different aspects of the materials and types of skill that need to be considered when measuring skill. The study led to the opportunities for subjects to 'stray' from the expected behaviour, derived from the model of skill, - critical incidents or 'happy accidents'. It is proposed that the factorial paradigm is an inappropriate approach to the study of such a complicated applied area.

The effect of a critical incident encountered by some of the subjects is evaluated. The particular incident is the accidental finding of instructions in the manual relevant to tasks to be performed later in the experiment. The model of skill used in the thesis led to the hypothesis that such an incident should lead to enhanced performance; examination of the results of subjects that encountered such an incident indicated that these subjects performed less well than non-incident subjects. Analysis of the use of the manual showed that they encountered the information early because of their strategy of heavy reliance and exploration of the manual. It is suggested that this was the source of the difference between the results of the two groups.

The development of a more sophisticated model of the system and the examination of the results of further studies for patterns of interaction caused by the nature of the training materials is recommended. This will allow the continued experimental study of this applied area without the loss of the features that characterise it.

The thesis has established that the category of users described in the literature as 'casual users' exists and that they exhibit the behaviours identified in anecdotal reports.

It has brought a rigour in the design of texts, tests and experimental method to this area of study and developed techniques that can be used by other workers for further research.

The studies have provided original descriptions of the behaviour of subject's use of manuals for this type of task.

The results and discussion of both the main study and study of the effects of graphic design of texts on learning provide a strong argument for the need of a designer of texts and tasks rather than the slavish application of human factors research.

REFERENCES

Ackoff, R. (1956). The development of operations research as a science. Operational research 1956, 4, pp256-295, 4, 256-295.

Allwood, C. M. (1986). Novices on the computer: a review of the literature. International Journal of Man-Machine Studies, 25(6), 633-658.

Allwood, C. M.,Wikstrom, T., & Reder, L. (1982). The effect of format and structure of text material on recallability. Poetics, 11, 145-153.

Allwood, C., & Wikstrom, T. (1986). Learning complex computer programs. Behavior and IT, 5(3), 217-255.

Alvari, M. a. W., I. (1986). Managing the risks associated with end-user computing. Journal of Management of Information Systems, 2(3), 5-20.

Anderson, J. (1981). Effects of prior knowledge on memory for new information. Memory and Cognition, 9(3), 237-265.

Anderson, J. (1982). Acquisition of cognitive skill: Compilation of Weak-Method Problem Solutions. Psychological Review, 94(2), 192-210.

Anderson, J. R. (1987). Skill acquisition: compilation of weak-method problem solutions. Psychological Review, 94(2), 192-210.

Annett, J., & Sparrow, J. (1985). Transfer of training: a review of research and practical implications. Programmed learning and educational technology, 22, 116-124.

Annett, J.,Duncan, K.,Stammers, R., & Gray, M. (1971). Task analysis No. 6). HMSO.

Atkinson, R., & Shiffrin, R. (1971). The control of short-term memory. Scientific American, 225(83).

Ausubel, D. (1960). The use of advanced organisers in the learnng and retention of meaningful verbal material. Journal of Educational Psychology, 51(5), 267-272.

Ausubel, D. P. (1961). In Defense of Verbal Learning. Educational Theory, 11, 15-25.

Barrett, G.,Thornton, C., & Cabe, P. (1968). Human factors evaluation of computer-based information storage and retrieval system. Human Factors, 10(4), 431-436.

Berry, C., & Broadbent, D. (1990). The role of instruction and verbalisation in improving performance on complex search tasks. Behaviour and Information Technology, 9(3), 175-190.

Black, J. B.,Bechtold, J. S.,Mitrani, M., & Carroll, J. M. (1989). On-line tutorials: what kind of inference leads to the most effective learning ? CHI' 89 Proceedings, 81-83.

Black, J.,Carroll, J., & McGuigan, S. (1987). What kind of minimal instruction manual is the most effective. In CHI + GI 1987, (pp. 159-162).

Briggs, P. (1990). Do they know what they're doing ? An evaluation of word-proecessor users' implicit and explicit task-relevant knowledge, and its role in self-directed learning. International Journal of Man-Machine Interaction, 32, 385-398.

Broadbent, D. (1978). Language and ergonomics. Applied Ergonomics, 8, 15-18.

Bruner, J. S. (1966). Towards a theory of instruction. Camb. Mass.: Harvard Press.

Card, S.,Moran, T., & Newell, A. (1983). The psychology of human computer interaction. Hillsdale: Lawrence Erlbaum.

Carroll, J, . Mack, R. (1982). Metaphor, computer systems and active-learning. IBM Research Report.

Carroll, J. M. (1987). What kind of minimal manual is most effective ? In Proceedings of CHI+GI '87, . Toronto:

Carroll, J. M., & Kellog, W. A. (1989). Artifact as theory nexus: hermeneutics meets theory-based design. In CHI'89, (pp. 7-15).

Carroll, J. M., & Olson, J. R. (1988). Mental models in human-computer interaction. In M. Helander (Eds.), Handbook of human-computer interaction Elsevier.

Carroll, J. M.,Smith-Kerker, P. L.,Ford, J. R., & Mazur-Rimetz, S. A. (1987). The minimal manual. In Human-Computer Interaction, 3 (pp. 123-153). Lawrence Erlbaum Association Inc.

Carroll, J., & Carrithers, C. (1984a). Blocking learner error states in a training wheels system. Human Factors 1984, 26(4), 377-389.

Carroll, J., & Carrithers, C. (1984b). Training Wheels in a user interface. Communications of the ACM, 27(8), 800-806.

Carroll, J., & Carrithers, C. (1986). Paradox of the Active User (HCI Research Report No. IBM Watson Research Centre.

Carroll, J., & Rosson, M. (1986). Paradox of the active user. HCI '86.

Caylor, J.,Sticht, T.,Fox, R., & Ford, J. (1973). Methodologies for determining reading requirements of military occupational specialities. No. HUMRO.

Chapanis, A. (1965). Words, words, words. , 7, 1-17.

Codd, E. (1974). Seven steps to RENDEZVOUS with the casual user. In J. Klimbie & K. Koffeman (Eds.), Database management (pp. 179-199). North-Holland.

Cohill, A., & Williges, R. (1985). Retrieval of HELP information for novice users of interactive comuter systems. Human Factors, 27, 335-345.

Coles, P., & Foster, J. (1975). Typographic cuing as an aid to learning from typewritten text. Programmed Learning and Educational Technology, 12, 102-108.

Conway, M., & Kahney, H. (1987). Transfer of learning in inference problems: learning to program recursive functions. In J. Hallam & C. Mellish (Ed.), Advances in Artificial Intelligence: Proceedings of the AISM Conference, . University of Edinburgh: John Wiley and Sons.

Coombs, M.,Gibson, R., & Alty, J. (1981). Acquiring a first computer language: a study of individual differences. In M. J. Coombs & J. Alty (Eds.), Computing skills and the user interface London: Academic Press.

Craik, F., & Lockhart, R. (1972). Levels of processing: a framework for memory research. Journal of Verbal Learning and Verbal Behaviour, 12, 598-607.

Craik, F., & Lockhart, R. (1972). Levels of processing: a framework for memory research. Journal of Verbal Learning and Verbal Behaviour, 12, 598-607.

Craik, F., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. Journal of Experimental Psychology: General, 104(3), 268-294.

Cuff, R. (1980). On casual users. International Journal of Man-Machine Studies, 12, 163-188.

Czaja, S.,Hammond, K.,Balsovich, J., & H., S. (1986). Learning to use a word processing system as a function of training strategy. Behaviour and IT, 5, 203.

Damodaran, L. (1974). Trends in trainng naive computer users. (Memo 96). Husat.

Dawson, A.,Coombs, M., & Alty, J. (1982). How to improve computer advisory services. Software practice and experience, 12, 857-877.

Diehl, W., & Mikalecky, L. (1981). Making written information fit worker's purposes. IEEE Transactions on Professional Communication, 24(1), 5-9.

Dobson, D. (1991). Buying skills. Which Computer ?, 14(5), 71-79.

Dreyfus, H., & Dreyfus, S. (1986). The Power of Human Intuition and Expertise in the Era of the Computer. Oxford: Basil Blackwell.

Duffy, T., & Waller, R. (1985). Designing useable texts. Academic Press.

Eames, R., & Starr, J. (1965). Technical publications and the user. Human Factors, 7, 363-369.

Eason, K. (1976). Understanding the naive computer user. Computer Journal, 19, 3-7.

Eason, K. (1979). Man-computer communication in public and private computing. No. 173). HUSAT.

Eason, K. (1981) Manager-Computer Interaction: A Study of a Task-Tool Relationship. Ph.D., Loughborough University of Technology.

Eason, K. (1982). Human Factors in IT. Physics Technology, 13, 196-201.

Eason, K. (1984). <u>The continuing support needs of end users.</u> No. HUSAT.

Eason, K. D., & Damodaran, L. (1974). The needs of the commercial user. In <u>Computing skills and the user interface</u> (pp. 115-143).

Eco, U. (1976). <u>Theory of Semiotics</u>. Indiana Univeristy Press.

Farkas, D., & Williams, T. (1990). John Carroll's *The Nurnberg Funnel* and Minimalist Documentation. <u>IEEE Transactions on Professional Communication</u>, <u>33</u>(4), 182-187.

Faw, H., & Waller, T. (1976). Mathemagenic behaviours and efficiency in learning from prose material: review, critique and recommendations. <u>Review of Educational Research</u>, <u>46</u>(4), 691-720.

Finlay, J., & Harrison, M. (1990). Pattern recognition and interaction models. In D. Dialper,D. Gilmore,G. Cockton, & B. Shackel (Ed.), <u>Human-Computer Interaction - INTERACT '90</u>, . North-Holland.

Fisher, R., & Craik, F. (1980a). The effects of elaboration on recognition memory. <u>Memory and Cognition</u>, <u>8</u>(5), 400-404.

Forrester, M., & Reason, D. (1991). HCI 'Intraface Model' for Systems Design'. <u>Interacting with Computers</u>, <u>2</u>(3).

Frase, L., & Schwartz, B. (1979). Typographical cues that facilitate comprehension. <u>Journal of Educational Psychology</u>, <u>71</u>(2), 197-206.

Garnham, A. (1981). Mental models as representation of text. <u>Memory and Cogniation</u>, <u>9</u>(6), 560-565.

Gick, M., & Holyoak, K. (1980). Analogical problem solving. <u>Cognitive Psychology</u>, <u>12</u>, 306-355.

Gilb, T., & Weinberg, G. (1977). <u>Humanized Input</u>. Cambridge, Massachusetts: Winthrop.

Gilmore, D. (1991). Visibility: a dimensionsal analysis. In D. Diaper & N. Hammond (Ed.), British Computer Society, Human Computer Interaction Specialist Interest Group, VI (pp. 317-330). Edinburgh: Cambridge University Press.

Graesser, A. C., & Nakamura, G.V.. (1982). The impact of schema on comprehension and memory. In G. H. Bower (Eds.), The Psychology of Learning and Motivation (pp. 59-109).

Guillemette, R. (1987). Application Software Documentation. Journal of Systems Management, 38(5), 36-39.

Hartley, J. (1974). Reading hygeine : a different viewpoint. The Journal of Educational Research, 196.

Hartley, J. (1981). Eighty ways of improving instructional text. IEEE Transactions on Professional Communications, 24(1), 17-27.

Johnson, P.,Diaper, D., & Long, J. (1984). Tasks, skills and knowledge: task analysis for knowledge based descriptions. In B. Shackel (Ed.), Interact, . Elsevier-Science.

Kalt, N. C., & Barrett, k. M. (1973). Facilitation of learning from a technical manual: an exploratory investigation. Journal of Applied Psychology, 58(3), 357-361.

Karat, J.,Boyes, L.,Weisberger, S., & Schafer, C. (1985). Transfer between word processing systems. In M. Mantei & P. Orbeton (Eds.), Human Factors in Computing Systems III North Holland.

Kato, T. (1968). What 'question asking protocols' can say about the interface. International Journal of Man-Machine Studies, 25, 659-673.

Kennedy, T. (1975). Some behavioural factors affecting the training of naive users of an interactive computer system. International Journal of Man-Machine Studies, 7, 817-834.

Kintsch, W., & van Dijk, T.A., v. D. (1978). Toward a model of text comprehension and production. Psychological review, 85(5), 363-394.

Lang, T.,Lang, K., & Auld, R. (1981b). A longitudinal study of computer-user behaviour in a batch environment. International Journal of Man-Machine Studies, 14, 251-268.

Lewis, C., & Mack, R. (1982). Learning to use a text processing system: evidence from 'thinking aloud' protocols. In Proceedings of Human Factors in Computing Systems Conference, .

Macdonald-Ross, M., & Waller, R. (1975). Criticism, alternatives and tests: a conceptual framewok for improving typography. Programmed Learning and Educational Technology, 12(2), 75.

Mack, R. L.,Clayton, H. L., & Carroll, J. M. (1983). Learning to use word processors: problems and prospects. ACT Transactions on Office Automation Systems, 1, 254-271.

Mahony, K., & Gower, A. (1991). The Development of a Visual Style for a BT X Window System Toolkit. In D. Diaper & N. Hammond (Ed.), British Computer Society, Human Computer Interaction Specialist Interest Group, VI (pp. 265-280). Edinburgh: Cambridge University Press.

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

Mayer, R. (1967). Computer bases subsystems for training users of computer systems. IEEE Trans. on Human Factors in electronics, 8(2), 70-75.

Mayer, R. E. (1979). Can adavance organisers influence meaningful learning ? Review of Educational Research, 49, 371-383.

Mayer, R. E. (1980). Elaboration techniques that increase the meaningfulness of technical text: an experimental test of the learning strategy hypothesis. Journal of Educational Psychology, 72(6), 770-784.

Mayer, R., & Bromage, B. K. (1979). Different recall protocols for technical text due to sequencing of advanced organizers. Journal of Educational Psychology, 72(2), 209-225.

Muter, P.,Latremouille, S.,Treurniet, W., & Beam, F. (1982). Extended reading of continuous text on television screens. Human Factors, 24(5), 501-508.

Myatt, B. T.,K.H., S.,Kamourni, A. L., & Tykodi, T. A. (1986). Which way to computer literacy; programming or applications experience ? International Journal of Man Machine Studies, 25(5).

Neilsen, J.,Mack, R.,Bergendorff, K., & Grischkowsky, N. (1986). Integrated software useage in teh professional work environment: Evidence from questionaires and interviews. In CHI'86 Human Factors in Computing Systems, . ACM SIGCHI.

Nickerson, R. (1969). Man-computer interaction : a challenge fro human factor research. IEEE Transactions on Man-Machine Studies, 10, 164.

Norman, D. A. (1986). Cognitive Engineering. In D. A. Norman & S. W. Draper (Eds.), User centered system design New Jersey: Lawrence Erlbaum.

Nowaczyk, R. H. (1984). The relationship of problem-solving ability and course performance among novice programmers. International Journal of Man-Machine Studies, 21, 149-160.

O'Malley, C. (1986). Helping users to help themselves. In D. Norman (Eds.), User centered systems design (pp. 377-398). Lawrence Erlbaum Associates.

Pepper, B. (1981). Following student's suggestions for rewriting computer programming books. American Education Research Journal, 18, 259-269.

Pirsig, R. M. (1974). Zen and the Art of Motorcycle Maintenance.

Polson, P. G.,Bovair, S., & Kieras, D. (1987). Transfer between text editors. In CHI + GI 1987, .

Pugh, A. K. (1977). Styles and strategies in silent reading. In Processing of visible language (pp. 431-441).

Rasmussen, J. (1989). The reflective expert and prenovice: notes on skill-, rule-, and knowledge-based performance in the setting of instruction and training. In L. Bainbridge & A. Ruiz-Quintanilla (Eds.), Developing skills with information technology John Wiley and Sons.

Reder, L. M., & Anderson, J. R. (1982). Effects of spacing and embellishment on memory for the main points of a text. Memory and Cognition, 10(2), 97-102.

Reder, L. M.,Charney, D. H., & Morgan, K. I. (1986). The role of elaborations in learning a skill from an instructional text. Memory and Cognition, 14(1), 64-78.

Ross, B. (1984). Remindings and their effects in learning a cognitive skill. Cognitive Psychology, 16, 371-416.

Rosson, M. (1983). Patterns of experience in text editing. In CHI'83 Human Factors in Computing Systems, .

Rosson, M. B. (1984). Effects of experience on learning, using and evaluating a text editor. Human Factors, 26(4), 436-475.

Rothkopf, E. Z. (1971). Incidental memory for location of information in text. Journal of verbal Learning and Verbal Behaviour, 10, 608-613.

Sackman, H. (1970). Time sharing and self tutoring: an exploratory case history. Human Factors, 12(2), 203-214.

Sasse, A.,Johnson, G., & Briggs, P. Introducing word processing to novice users: a study of 'procedural' and 'conceptual' approaches. .

Scharer, L. L. (1983). User training: less is more. Datamation, 175-182.

Schuck, G. (1985). Intelligent technology, intelligetn workers: a new pedagogy for the high-tech work place. Organisational dynamics, 14(2), 66-79.

Schumacher, G., & Waller, R. (1985). Testing design alternatives: a comparison of procedures. Academic Press.

Shannon, G. (1985) Developing an instructional help facility for casual users of a word processor - a feasability study. MSc Dissertation, Loughborough University of Technology.

Shepherd, A. (1976). An improved tabular format for task analysis. Journal of Occupational Psychology, 49, 93-104.

Shepherd, A. (1985). Heirarchical task analysis and training decisions. Programmed Learning and Educational Technology, 22(2), 162-176.

Shepherd, A. (1987). Training decisions and IT tasks. In Workshop: What and how to teach IT, . Bad Hamburg:

Shepherd, A. (1989). Analysis and training of IT tasks. In D. Diaper (Eds.), Task analysis for Human-Computer Interaction Chichester: Ellis Horwood.

Shepherd, A. (1990). An Operating Model of Skill to Support the Design of Training. in press.

Shepherd, A., & Duncan, K. (1980). Analysing a complex planning task. In K. Duncan,M. Gruneberg, & D. Wallis (Eds.), Changes in working life Chichester: John Wiley.

Sinclair, A. (1987) An improved user manual and training solution for a distribution industry software package: tjhe development of a formal methodology. MSc Dissertation, Loughborough University of Technology.

Singley, M., & Anderson, J. (1985). The transfer of text-editing skill. International Journal of Man-Machine Studies, 22, 403-423.

Stammers, R. B. (1971). Training and the acquisition of knowledge and skill. In P. B. Warr (Eds.), Psychology at Work Reading: Penguin.

Stein, B. S.,Littlefield, J.,Bransford, J. D., & Persampieri (1984). Elaboration and knowledge acquisition. Memory and Cognition, 12(5), 522-529.

Stevenson, R. J. (1981). Depth of comprehension, effective elaboration, and memory for sentences. Memory and Cognition, 9(2), 169-176.

Sticht, T.,Caylor, J.,Kern, R., & Fox, L. (1972). Project REALISTIC: Determination of adult functional literacy levels. Reading Research Quarterly, Spring, 424-465.

Sullivan, M., & Chapanis, A. (1983). Human factoring a text editor manual. Behaviour and Information Technology, 2, 113-125.

Szlichcinski, K. (1979). Telling people how things work,. Applied Ergonomics, 10(1), 2-8.

Taylor, W. (1953). Cloze procedure: a new tool for measuring readability. Journalism Quarterly, 30, 415-433.

Tetzlaff, L. (1987). Transfer of learning: beyond common elements. In CHI + GI, (pp. 205-210).

Uhler, H. (1981). Training and support - shifting the responsibility. The Seybold Report on Word Processing, 4(1).

Uhler, H. (1981). Training and support - shifting the responsibility. The Seybold Report on Word Processing, 4(1).

Uhler, H. (1984). Training - managers, professionals and executives. The Seybold Report on Word Processing, 7(3).

Waern, Y. (1985). Learning computerised tasks as related to prior task knowledge. International Journal of Man-Machine Interaction, 22, 441-445.

Walker, R. (1979). Text as diagram: using typography to improve access and understanding. In D. Jonassen (Eds.), The Technology of Text , N.J. Educational Technology Publications.

Waller, R. (1979). Functional information design: research and practice. Information Design Journal, 1, 43-50.

Waller, R. (1979). Typographic access structures for educational texts. In P. Kolers,M. Wrolstad, & H. Bouma (Ed.), Processing of Visible Language, 1 . Plenum Press.

Weber, B. (1985). Technical writing skills: a question of aptitude or interest. Journal of Technical Writing and Communication, 15, 63-68.

Wright, P. (1979). Quality control aspects of document design. Information Design Journal, 1, 33-42.

Wright, P. (1981). "The instructions clearly state ....' can't people read ? Applied Ergonomics, 12(3), 131-141.

Wright, P. (1983). Manual dexterity: a user-oriented approach to creating computer documentation. In CHI'83, (pp. 11-18).

Wright, P. (1984). Designing the documenteation that explains how IT works. , 5(2), 73-78.

Wright, P., & Monk, A. (1989). Evaluation for design. In People and Computers IV Cambridge University Press.

Wright, P., & Reid, F. (1973). Written information: some alternatives to prose expressing the outcomes of complex contingencies. Journal of Applied Psychology, 57(2), 160-166.

Young, S., & Wogalter, M. (1990). Comprehension and memory of instruction manual warnings: conspicuous print and pictorial icons. Human Factors, 32(6), 637-649.

Ziegler, J.,Happe, H., & Fahnrich, K. (1987). Learning and transfer for text and graphics editing with a direct manipulation device. In M. Martei & P. Orbeton (Ed.), Human Factors in Computing Systems III,

Zipf, G. (1965). Human behaviour and the principle of least effort: an introduction to human ecology. New York: Haffner.

APPENDIX 1

The instructions for the first part of the study of prActical problems of measuring skill are shown below.

Question asking experiment

Thank you for volunteering to take part in this experiment. There are no hidden purposes to the experiment and it is not a test.

The main aim of the experiment is to find out what sorts of questions people ask when they are learning to use a word processor. Instead of using a manual, I will be available as your own personal tutor - please ask me anything you need to know. You are not expected to have any prior knowledge or direct experience of using a word processor.

You will be asked to make a copy of a letter using the word processor and to go through the motions of printing the letter, though a printer is not attached to this computer.

Learning to use any word processor can be a difficult and frustrating activity. Please don't be shy about asking questions - even if you have to repeat them many times. If there is anything that is hard to understand it is either my fault through giving a poor answer or just a difficult thing to learn. Please try and make your questions as specific as possible rather than vague ones such as "What do I do next ?"

APPENDIX 2

The empirical results of the pilot studies are shown on the following pages.

Light-Far Group	Tasks												
Subjects	1	2	3	4	5	10	12	6	8	13	7	9	11
1	247	379	183	69	350	218	75	700	648	290	414	212	115
2	832	258	220	147	211	265	175	572	202	352	385	240	192
3	947	294	268	235	297	177	153	430	402	200	286	345	189
4	379	441	308	267	354	219	176	724	408	280	430	266	246
5	727	429	420	266	266	193	179	999	403	180	379	233	224
6	132	90	43	43	46	81	53	310	100	120	532	71	58
7	1141	248	118	126	289	80	86	1146	185	115	505	234	158
MEAN	629	306	223	165	259	176	128	697	335	220	419	229	169
Core Group	Tasks												
Subjects	1	2	3	4	5	10	12	6	8	13	7	9	11
1	196	195	149	68	193	89	76	152	178	166	130	140	148
2	401	184	81	81	114	63	59	124	109	88	129	104	144
3	429	101	83	51	62	52	46	312	102	161	118	125	163
4	209	160	247	83	277	62	62	276	141	131	154	72	57
5	139	76	58	49	66	45	55	119	136	94	181	102	86
6	129	113	52	49	59	49	58	117	113	65	120	108	94
7	173	144	93	72	76	68	55	155	113	80	117	92	103
MEAN	239	139	109	65	121	61	59	179	127	112	136	106	114
Light-2 Group	Tasks												
Subjects	1	2	3	4	5	10	12	6	8	13	7	9	11
1	555	1298	87	70	267	87	87	931	754	78	463	730	53
2	1383	1188	316	233	256	124	145	604	335	174	489	257	273
3	612	702	298	409	314	55	70	472	247	152	426	279	180
4	586	44	85	66	77	53	50	473	101	247	298	95	62
5	604	50	124	53	334	58	42	441	183	107	564	93	63
6	590	82	85	71	146	79	115	845	150	240	436	131	114
7	785	181	85	85	76	76	66	382	109	88	421	93	61
MEAN	731	506	154	141	210	76	82	593	268	155	442	240	115
No-Light Group	Tasks												
Subjects	1	2	3	4	5	10	12	6	8	13	7	9	11
1	790	354	98	72	290	68	54	404	155	171	256	172	65
2	649	317	273	192	244	93	62	799	595	229	318	199	163
3	356	327	70	117	274	60	50	200	251	118	218	157	109
4	399	201	134	120	194	96	71	549	263	156	267	163	139
5	601	577	79	70	165	59	50	466	138	174	303	155	100
6	472	230	163	135	220	165	116	315	182	152	554	167	134
7	823	178	71	54	110	52	58	483	317	137	495	168	143
MEAN	584	312	127	109	214	85	66	459	272	162	344	169	122
Cover-Close_HTA Grou	Tasks												
Subjects	1	2	3	4	5	10	12	6	8	13	7	9	11
1	375	71	60	53	67	53	56	61	117	66	295	162	129
2	491	274	641	167	317	83	60	413	139	231	362	189	172
3	445	322	43	43	115	361	44	361	72	89	333	249	57
4	596	59	64	58	80	53	114	277	153	125	614	167	50
5	428	57	60	76	57	79	48	413	81	98	485	373	77
6	622	106	60	59	119	61	54	306	121	114	263	214	149
7	406	221	116	71	250	55	44	213	128	94	329	194	248
MEAN	480	159	149	75	144	106	60	292	116	117	383	221	126

Light-Far Group	Tasks												
Subjects	1	2	3	4	5	10	12	6	8	13	7	9	11
1	36	27	9	0	30	17	0	12	38	21	26	22	3
2	31	21	16	6	23	19	17	37	51	27	28	24	22
3	42	26	26	26	26	14	12	32	28	15	30	31	20
4	38	29	21	16	33	17	11	30	29	17	26	27	23
5	45	41	38	29	31	28	28	60	31	41	38	33	34
6	24	0	0	0		0	0	20		0	34	4	0
7	30	3	0	0	10	0	0	58	7	0	36	21	15
MEAN	35	21	16	11	26	14	10	36	31	17	31	23	17
Core Group	Tasks												
Subjects	1	2	3	4	5	10	12	6	8	13	7	9	11
1													
2													
3													
4													
5													
6													
7													
MEAN													
Light-2 Group	Tasks												
Subjects	1	2	3	4	5	10	12	6	8	13	7	9	11
1	28	60	0	0	16	0	0	34	0	0	32	34	0
2	60	72	28	21	20	6	8	33	21	12	36	22	24
3	32	42	24	34	22	2	2	14	13	11	30	23	20
4	34	2	0	0	0	0	0	19	0	19	30	0	0
5	34		4	0	12	0	0	13	0	0	36	0	0
6	35	1	0	0	4	0	0	28	0	7	32	0	0
7	42	8	0	0	0	0	0	19	0	0	27	0	0
MEAN	38	31	8	8	11	1	1	23	5	7	32	11	6
No-Light Group	Tasks												
Subjects	1	2	3	4	5	10	12	6	8	13	7	9	11
1	46	13	0	0	25	0	0	19	8	9	27	18	0
2	43	28	22	18	22	2	0	46	46	15	32	26	22
3	28	40	0	0	31	0,	0	9	13	8	24	21	16
4	30	28	8	0	14	0	0	45	19	9	32	25	14
5	40	56	0	0	12	0	0	19	7	9	30	22	4
6	28	28	15	10	24	11	9	27	17	12	43	24	21
7	36	1	0	0	3	0	0	26	21	6	40	26	18
MEAN	36	28	6	4	19	2	1	27	19	10	33	23	14
Cover-Close_HTA Grou	Tasks												
Subjects	1	2	3	4	5	10	12	6	8	13	7	9	11
1	24	10	0	0	7	0	0	5	5	0	24	18	10
2	34	30	53	10	24	1	0	34	10	17	32	21	18
3	33	20	0	0	0	0	0	22	0	0	26	18	4
4	31	0	0	0	5	0	1	12	6	3	31	10	0
5	26	0	0	0	0	0	0	17	0	1	28	22	4
6	35	4	0	0	7	0	0	16	7	9	25	22	18
7	24	19	5	5	22	1	0	9	8	5	35	22	21
MEAN	30	12	8	2	9	0	0	16	5	5	29	19	11

Light-Far Group	Tasks												
Subjects	1	2	3	4	5	10	12	6	8	13	7	9	11
1	15	9	2	0	8	3	0	4	13	5	7	6	1
2	9	6	4	1	7	6	4	7	14	7	7	7	6
3	17	10	11	11	11	6	4	12	11	5	9	7	5
4	9	11	9	6	12	4	5	14	13	4	8	7	6
5	14	14	13	9	10	10	11	13	7	14	11	11	11
6	11	0	0	0	0	0	0	6	1	0	10	1	0
7	9	0	0	0	6	0	0	20	2	0	12	5	3
MEAN	12	7	6	4	8	4	3	11	9	5	9	6	5
Core Group	Tasks												
Subjects	1	2	3	4	5	10	12	6	8	13	7	9	11
1	10	6	1	0	3	0	0	6	3	5	9	6	5
2	6	1	0	0	0	0	0	2	1	0	5	4	5
3	8	0	0	0	0	0	0	6	2	3	6	5	5
4	3	5	5	0	4	0	0	5	3	5	6	3	1
5	8	0	0	0	0	0	0	3	5	1	10	5	1
6	9	2	0	0	0	1	0	7	1	0	9	6	4
7	10	6	0	0	4	0	0	7	2	0	9	6	4
MEAN	8	3	1	0	2	0	0	5	2	2	8	5	4
Light-2 Group	Tasks												
Subjects	1	2	3	4	5	10	12	6	8	13	7	9	11
1	11	21	0	0	8	0	0	12	0	0	11	9	0
2	25	28	12	8	7	1	2	12	7	3	12	6	7
3	12	16	10	9	8	0	0	4	4	3	10	6	5
4	11	0	0	0	0	0	0	5	0	6	9	0	0
5	14	9	2	0	3	0	0	3	0	0	10	0	0
6	13	0	0	0	1	0	0	5	0	0	9	0	0
7	16	5	0	0	0	0	0	4	0	0	7	0	0
MEAN	15	11	3	2	4	0	0	6	2	2	10	3	2
No-Light Group	Tasks												
Subjects	1	2	3	4	5	10	12	6	8	13	7	9	11
1	16	2	0	0	8	0	0	4	2	2	7	5	0
2	16	12	7	6	8	0	0	9	7	2	11	8	6
3	10	10	0	0	13	0	0	3	4	2	7	5	4
4	12	10	1	0	6	0	0	11	6	2	11	7	4
5	12	20	0	0	4	0	0	3	2	3	10	6	1
6	11	9	7	6	10	3	2	7	6	3	13	6	5
7	16	0	0	0	0	0	0	7	6	1	11	6	4
MEAN	13	9	2	2	7	0	0	6	5	2	10	6	3
Cover-Close_HTA Grou	Tasks												
Subjects	1	2	3	4	5	10	12	6	8	13	7	9	11
1	9	3	0	0	1	0	0	1	1	0	7	5	2
2	11	11	16	4	9	0	0	5	3	6	9	5	4
3	8	4	0	0	0	0	0	5	0	0	8	5	1
4	10	0	0	0	1	0	0	1	1	1	6	2	0
5	11	0	0	0	0	0	0	3	0	0	8	6	1
6	13	1	0	0	4	0	0	4	1	2	6	6	5
7	9	6	1	1	9	0	0	3	1	1	11	6	5
MEAN	10	4	2	1	3	0	0	3	1	1	8	5	3

Summary of Time per task for the pilot experiments

	INITIAL TEST				RECALL			TRANSFER			FAR TRANSFER		
	1	2	3	4	5	10	12	6	8	13	7	9	11
Light , Far Task													
MEAN	629.29	305.57	222.86	164.71	259.00	176.14	128.14	697.29	335.43	219.57	418.71	228.71	168.86
SD	380.68	123.25	124.56	92.60	105.97	70.79	54.68	297.27	186.11	90.32	82.42	81.89	64.86
				330.61			187.76			417.43			272.10
				272.71			94.00			287.78			131.24
Light-2, Close Task													
MEAN	730.71	506.43	154.29	141.00	210.00	76.00	82.14	592.57	268.43	155.14	442.43	239.71	115.14
SD	297.26	553.41	105.40	133.38	109.01	24.93	36.76	213.91	229.29	69.26	80.62	230.23	83.13
				383.11			122.71			338.71			265.76
				397.69			90.35			258.88			197.55
No-Light, close Task													
MEAN	584.29	312.00	126.86	108.57	213.86	84.71	65.86	459.43	271.57	162.43	344.43	168.71	121.86
SD	183.86	134.93	73.25	47.72	63.15	39.29	23.31	189.37	156.32	35.14	128.33	14.66	32.80
				282.93			121.48			297.81			211.67
				226.31			79.78			185.09			122.33
Core, Close Task													
MEAN	239.43	139.00	109.00	64.71	121.00	61.14	58.71	179.29	127.43	112.14	135.57	106.14	113.57
SD	123.52	44.19	68.57	14.97	83.61	14.83	9.12	80.49	26.51	40.44	23.71	21.96	38.75
				138.04			80.29			139.62			118.43
				95.90			55.31			59.26			30.46
Cover + Light + 2 - HTA-Close													
MEAN	480.43	158.57	149.14	75.29	143.57	106.43	60.00	292.00	115.86	116.71	383.00	221.14	126.00
SD	95.02	111.50	218.09	41.90	100.19	112.93	24.58	125.21	29.52	53.76	123.67	73.13	71.27
				215.86			103.33			174.86			243.38
				202.49			90.81			114.18			139.77

Summary of Cards per task for the pilot experiments

	INITIAL TEST				RECALL			TRANSFER			FAR TRANSFER		
	1	2	3	4	5	10	12	6	8	13	7	9	11
Light , Far Task													
MEAN	35.14	21.00	15.71	11.00	25.50	13.57	9.71	35.57	30.67	17.29	31.14	23.14	16.71
SD	7.31	14.66	13.96	12.66	8.41	10.24	10.63	18.00	14.40	14.57	4.88	9.55	11.88
				20.71			15.80			27.70			23.67
				14.95			11.51			17.01			10.65
Light-2, Close Task													
MEAN	37.86	30.83	8.00	7.86	10.57	1.14	1.43	22.86	4.86	7.00	31.86	11.29	6.29
SD	10.62	31.35	12.44	13.93	9.29	2.27	2.99	8.75	8.61	7.44	3.29	14.59	10.80
				20.78			4.38			11.57			16.48
				22.05			7.09			11.38			15.19
No-Light, close Task													
MEAN	35.86	27.71	6.43	4.00	18.71	1.86	1.29	27.29	18.71	9.71	32.57	23.14	13.57
SD	7.40	17.71	8.98	7.21	9.52	4.10	3.40	13.77	13.15	2.93	6.78	2.97	8.44
				18.50			7.29			18.57			23.10
				17.43			10.21			12.86			10.05
Core, Close Task													
MEAN													
SD													
Cover + Light + 2 - HTA-Close													
MEAN	29.57	11.86	8.29	2.14	9.29	0.29	0.14	16.43	5.14	5.00	28.71	19.00	10.71
SD	4.79	11.50	19.81	3.93	9.83	0.49	0.38	9.54	3.85	6.19	4.07	4.36	8.34
				12.96			3.24			8.86			19.48
				15.26			6.95			8.56			9.40

Summary of 'Type' cards per task for the pilot experiments

	INITIAL TEST				RECALL			TRANSFER			FAR TRANSFER		
	1	2	3	4	5	10	12	6	8	13	7	9	11
Light , Far Task													
MEAN	12.00	7.14	5.57	3.86	7.71	4.14	3.43	10.86	8.71	5.00	9.14	6.29	4.57
SD	3.32	5.43	5.38	4.74	4.03	3.58	3.99	5.55	5.44	4.76	1.95	2.98	3.69
				7.14			5.10			8.19			6.67
				5.48			4.15			5.57			3.41
Light-2, Close Task													
MEAN	14.57	11.29	3.43	2.43	3.86	0.14	0.29	6.43	1.57	1.71	9.71	3.00	1.71
SD	4.93	10.77	5.26	4.16	3.72	0.38	0.76	3.87	2.82	2.36	1.60	3.87	2.98
				7.93			1.43			3.24			4.81
				8.28			2.73			3.73			4.57
No-Light, close Task													
MEAN	13.29	9.00	2.14	1.71	7.00	0.43	0.29	6.29	4.71	2.14	10.00	6.14	3.43
SD	2.63	6.61	3.34	2.93	4.20	1.13	0.76	3.09	2.06	0.69	2.24	1.07	2.15
				6.54			2.57			4.38			6.52
				6.33			4.02			2.71			3.30
Core, Close Task													
MEAN	7.71	2.86	0.86	0.00	1.57	0.14	0.00	5.14	2.43	2.00	7.71	5.00	3.57
SD	2.50	2.73	1.86	0.00	1.99	0.38	0.00	1.95	1.40	2.31	1.98	1.15	1.81
				2.86			0.57			3.19			5.43
				3.62			1.33			2.32			2.38
Cover + Light + 2 - HTA-Close													
MEAN	10.14	3.57	2.43	0.71	3.43	0.00	0.00	3.14	1.00	1.43	7.86	5.00	2.57
SD	1.68	3.95	6.00	1.50	4.04	0.00	0.00	1.68	1.00	2.15	1.77	1.41	2.07
				4.21			1.14			1.86			5.14
				5.08			2.76			1.85			2.78