Towards the development of a model of User Engagement with packaged software

Cecilia Finnerty

A thesis submitted in partial fulfilment of the requirements of London Guildhall University for the degree of Doctor of Philosophy

February 2001

Abstract

This thesis draws on the body of knowledge in the field of HCI to develop a model of User Engagement with software with the aim of assisting organisations, which do not develop software in-house, in determining the level of user engagement with a software product prior to its introduction. It then goes further to develop the basis of an instrument for measuring that engagement before the commitment to purchase a software package. If an individual is happy with a piece of software, he is more likely to use it productively. If an organisation can assess the level of engagement with software prior to its introduction, it may take corrective action to ensure that its introduction is successful. The focus is on multi-user software of which usage is mandatory.

The User Engagement Model and the instrument to measure the overall User Engagement construct are tested in three stages:

1) By assessing the current level of satisfaction with software in use, by means of an industry tested questionnaire and comparing the results of this and of interviews and discussions with the User Engagement Model.

2) By testing the instrument to measure User Engagement (the ITU) by reference to external organisations, use of statistical analysis and correlation with the industry tested measure of satisfaction.

3) By use of the ITU, interview and discussion in two separate case studies where existing software is replaced by a new package and comparing the results with the User Engagement Model.

Both the User Engagement Model and the tool for measuring the User Engagement construct are necessarily at an early stage of development and more work is required in refining and testing them. However, it is believed that the creation of such a model and measurement tool would be of benefit to those organisations which rely on commercial off-the-shelf software (COTS) in their business. Not just because they would indicate the level of user engagement with a software package before its introduction but because the exercise would lead to questions being asked about why people are giving a negative response to an apparently suitable package. The answers to those questions could lead to a more detailed evaluation about what the organisation wants to achieve with the software and a strategy for addressing the human aspects as well as the technical aspects of introducing new software.

The need to address both these issues has long been recognised. This research re-iterates the need for more attention to be paid to the human issues, particularly when the software is not being designed in-house and the human aspects are likely to be neglected, the onus being placed on the software to match the needs of the organisation without the organisation having considered the need to have the right climate to introduce the software.

Table of Contents

List of Figures and Tables

Acknowledgements

Sincere thanks are due to several people who have been of assistance during this research: To all those, who must remain anonymous, who gave of their time completing questionnaires and taking part in interviews; to the organisations, including that of the author, who allowed people the time to take part in the research and to the people within them, especially Mr. John Seaman, who co-ordinated the handing out and return of questionnaires and provided background information; to Dr. Guy Winter for his comments on the developing questionnaire.

Thank-you also to the staff of the London Guildhall University - to my supervisors, Mr. Robert Scane and Mr. Jeff Naylor who have given of their precious time to offer comments and suggestions; to the library staff, particularly Mr. Phil Jones; and an especial thanks to Dr. Gwyn Jones for giving up his time to help me to get to grips with statistics.

Thanks to Geoff for all his help.

Chapter 1

Introduction

1.1 Background

In 1976 it was estimated that only about 20% of computer systems achieved their intended benefits (Eason, 1988). In 1995 a report by the Standish Group in the USA showed that little had changed, estimating that as many as 84% of computer projects failed to deliver.[1] Very recently, there have been press reports of the failures of high profile systems. Yet, when they are interviewed, the suppliers of these systems say that there is little wrong with their products.

Generally, it is the high profile computer projects which hit the headlines and these are normally systems which are designed for a specific task in a specific organisation. Many small to medium sized organisations (SME) do not have the resources to design in-house systems and rely on commercial off-the-shelf software (COTS) not just for single-user systems such as wordprocessing but for multi-user systems such as purchasing or sales management. This may represent quite a large financial outlay. A large percentage of the systems reported to have "failed" must belong to this category.

Success may be defined in several ways. From the point of view of the organisation as a whole it may be reflected in increased productivity, greater efficiency and higher quality. From the point of view of the individual in the organisation it may just be that he feels happy or satisfied with the software. Although the organisation may place the satisfaction of the individual lower in priority than improved efficiency, very often the two go hand in hand.

1.2 Problem Statement

Research in the field of HCI has been on how to improve the interaction between people and computers. This has necessarily focused on the design and development stage in the case of

[1]Reported by A.Dooley in Project Manager Today, 1998

computer software and various theories have been developed regarding the need to involve the users in this process. There has also been some investigation into the way the culture of an organisation may help or hinder such co-operative development. However, many small to medium enterprises (SME) are not able to develop software. They have to purchase packages which have been designed to fit the requirements of a multitude of companies. These are not just single-user packages such as wordprocessing but packages designed for people working as a group to perform functions such as purchasing and billing. There is little in the field of HCI which has considered the problems faced by such companies when they wish to purchase software that will both meet the functional requirements of the organisation and gain the satisfaction of the users.

This thesis is concerned with the engagement of the individual with the software he has to use in the workplace. It is generally accepted that people who are happy with the tools they are using are more productive and that problems encountered are more readily dealt with (Eason & Damodaran, 1981). How problems are dealt with (i.e. treating them as challenges or as obstacles) is generally accepted as being one of the factors which distinguishes success from failure. The question being addressed in this thesis is whether it is possible for an organisation to determine, in advance of the introduction of a software package, whether the people who will be going to use it have a favourable disposition towards the software. This may then be used as an indicator of how they will respond to any problems which arise and thence to the overall success of the system.

The software is specifically that which is purchased ready to use, designed for a multi-user environment where people work collectively and the software must be used in order to fulfil the job specified by the organisation. A distinction should be made here between a software application which is shared by several people in order to fulfil similar tasks using the same data and a software application which enables people to communicate and share ideas, often referred to as Computer Supported Co-operative Working (CSCW). It is the former type of software application which is of concern here.

1.3 Contribution to be made

Most of the literature on Human Computer Interaction (HCI) assumes that the software is being designed in-house. This is not the case when the organisational policy is to purchase COTS software. When software is being developed in-house there is an opportunity for an organisation to ensure that people are aware of what is happening and to have some part in it before the design of the software becomes fixed. The length of the process means both that problems may be brought to light during that time and that those problems may be addressed. This is not the case with a COTS package. A COTS package is designed to fit a range of organisational sizes and types. Although there may be some flexibility of screen presentation this is usually fairly limited and the core system must remain fixed in order for the supplier to be able to support and maintain it. This means that there is little or no up-front design and that people's first real experience of the software is when it is installed and made live. People are, therefore, adjusting to something at the same time as having to use it productively. A positive disposition towards the software at this stage will mean that problems are dealt with more readily and adjustments to the way of working are accepted.

The aim of this thesis is to develop a predictive model of the factors which contribute to that positive disposition (i.e. engagement), which parallels later satisfaction. If the presence or absence of these factors can be measured to indicate the level of user engagement with a software package prior to its introduction, it will enable organisations to consider the cause and to take corrective action if necessary.

The research is limited by the fact that it is confined to a single user organisation and the user groups are small. However, it has the benefit of having taken place in a "real-world" setting and over a period of time. It was possible to talk to the people using the software with reference to their responses to the questionnaires used and to correlate the responses of the same people with the proposed model. Whilst other research has the benefit of large user samples, the people are generally anonymous and confirmation of what they meant when completing a questionnaire or of obtaining further information, is not possible.

The organisation itself is structured along similar lines to many others and uses a standard method of establishing a requirements specification. It is, therefore, likely that the results of the research are not unique to this organisation, although no organisation is exactly like any other.

1.4 Research Approach

The research took place over a period of five years and is based largely in one organisation which is run along the lines of the public sector (although it is completely self-funding). It has a hierarchical structure which is not unlike that of the larger organisations in the private sector. The organisation has fewer than one thousand employees but is spread across six different locations. The locations are linked via a wide area network. The policy is not to develop software in house but to purchase ready to use packages. Some data was also gathered from other organisations, three of which are in the private sector.

The research question is addressed in four stages:

a) Proposal of the factors that affect user satisfaction with software

b) Development of a simple, predictive model of the key factors

c) Development of an instrument to measure those factors

d) Test of the model in a real-world environment

a) Proposal of Factors

The research into satisfaction with software reported in the literature was undertaken with regard to several different types of system. The thesis draws on the findings of this research to propose those factors for which there is some evidence in the literature and to lay a foundation for a model of user engagement with packaged software.

b) Development of a model of User Engagement

If a model is to be useful in these time pressured days, it must be simple whilst still capturing the key elements. The factors presented under a) are, therefore, re-examined to try to determine which factors are key. Several models predicting people's use of software have been developed (Hartwick & Barki, 1994; Goodhue, 1995; Martinko, Henry & Zmud,1996; Davis, 1989; Mathieson, 1991, Taylor & Todd, 1995). The thesis builds on the work of these models and the findings from the literature to develop a predictive model of the factors which influence an individual's disposition towards a software package.

Five hypotheses are proposed. Four of these relate to the model. The first hypothesis proposes that the characteristics of a software product alone are not sufficient to ensure user satisfaction and results from the findings in a) above. The purchaser of packaged software for an organisation will have a specification of the requirements of the software. It is likely that more than one package will meet these functional requirements. However, if the characteristics of the software alone were responsible for user satisfaction or lack of it, we should surely have seen some improvement on the failure rate over the past 25 years. Even so, it is necessary to test the hypothesis before proceeding with the research.

The four hypotheses which directly relate to the development of the User Engagement Model are:

•People will be positively disposed towards a software package if they perceive it to be easy to use.

•People will be positively disposed towards a software package if they perceive it to be useful in performing their job.

•People who feel they have been involved in the selection process of packaged software, will be positively disposed towards the software.

•The perceived attitude of management and of co-workers towards using a software package, will affect the disposition of the individual towards it.

The first two hypotheses draw heavily on the Technology Acceptance Model (TAM) (Davis et al., 1989). This is one of the simplest and most widely tested of the models predicting software use/acceptance. Perceived ease of use and usefulness are demonstrated to be two key factors which influence usage in this model. It was developed as a means of predicting whether, after short exposure to a system, people would be likely to use it in the future. Usage is often used as a surrogate for satisfaction. However, TAM itself was not suitable as a predictor for the organisational context described above: If there is no alternative tool, or if usage is a requirement of the job, then usage of itself may not indicate satisfaction. However, both the perception of ease of use and of usefulness contribute to the Attitude construct of TAM and this has been shown both in TAM and in related models to predict an individual's intention to use software. This would suggest that it also shows a favourable disposition towards the software. Hence the proposals that people will be more likely to feel satisfied with a software package if they perceive it to be both useful and easy to use.

The third hypothesis proposes that a sense of involvement in the process of selecting a software package is required in order for people to form a favourable disposition towards the software. This hypothesis is based on the wealth of literature on the need for people to be involved in the design and development of software and on one of the models predicting software use described in chapter 3. When software is purchased as a package such involvement is not possible. However, it is possible for people to be involved in the selection of which package to use.

The fourth hypothesis is derived largely from the Theory of Reasoned Action (TRA) (Fishbein & Ajzen, 1975) and is based on the influence that referent others have on individuals. However, there is also support in the HCI literature for the effect of perceived management and co-worker attitudes on software users (e.g. Damodaran, 1991; Martinko et al., 1996).

People's satisfaction with three different software packages already in use in the organisation under study are assessed by use of interviews and an existing instrument for measuring satisfaction. The results are compared both with the propositions on which the model has been built and the factors originally proposed under a) above. The findings lend support to the proposed model.

The comparison of the level of satisfaction with the three packages with an independent evaluation of the software, indicates that there is indeed more at play than the characteristics of the software package suggesting support for the first hypothesis.

c) Development of an instrument to measure User Engagement

Satisfaction with something can necessarily only be measured after experience of it. In the case of packaged software, this would mean after the investment in the software and training has already been made. If it were possible to measure people's disposition towards a software package in advance of the software purchase, this would enable a more detailed look at the factors affecting the outcome and the possibility of taking corrective action. The model developed proposes that people's disposition towards the software, termed User Engagement, is formed as a result of the combination of four factors. It is proposed that User Engagement may be measured after brief exposure to the software, prior to purchase and that it will correlate with measures of satisfaction once the software is in use.

In order to measure the correlation a formal metric of the factors in the form of a questionnaire is designed. Questionnaire design and validation is not something to be undertaken lightly. As Kirakowski (1995a) says: "All attitude questionnaires require several iterations of the cycle of item analysis, data gathering, and statistical analysis before they can be used with confidence" (p.1). The questionnaire developed borrows from other questionnaires which have themselves all been validated to some degree. Nonetheless, the resulting questionnaire is itself new and, therefore, in need of validation. In order to do this external organisations are contacted for assistance. Standard statistical methods are used to validate both the internal consistency of the questionnaire and the overall construct it is measuring. The results from this exercise are such that no changes are made to the questionnaire before using it in the two case studies.

d) Testing of the User Engagement Model

The model is tested by means of two case studies. The replacement of two software packages deemed to be unsatisfactory by their users is described. The new questionnaire is used to measure the presence of perceived ease of use, usefulness, involvement and management and co-worker support before the replacement software is introduced. The results are then correlated with a later measure of satisfaction and confirmed by means of interview and discussion. The process is then repeated in the second case study although, as reported, organisational circumstances meant that the final correlation with the questionnaire to measure satisfaction was not possible.

The results of each case study are discussed and compared with the proposed model. The evidence of the two case studies supports the hypotheses proposed for the User Engagement Model.

1.5 Outline of chapters

Chapter two gives a very brief overview of the field of HCI. It has long been recognised that there is a mismatch between the capability of the technology and its presentation, and the human worker. The different means of addressing this are presented. However it is pointed out that nearly all make the assumption that there is a designer who can take some action. In the case of packaged software, the person who designed it is probably working on a different product by the time the software comes to market.

Chapter three reviews the literature for the factors which affect user satisfaction with software and lays a foundation for a model of those factors.

Chapter four then reconsiders the factors presented in chapter three with a view to simplifying them to those which are key. Five hypotheses are proposed and a model is developed. The research method used and the reason for the approach adopted are then detailed.

Chapter five seeks to test the model inductively. The background to the purchase of three existing packages is detailed. Each of those packages is then tested for current levels of user satisfaction using an industry standard instrument. The results are linked with the results of interviews and discussions and compared with the results of an independent evaluation and related back to the hypotheses proposed in chapter four.

Chapter six describes the development of an instrument in the form of a questionnaire, to measure user engagement with software prior to its purchase, using the factors presented in the model. The attempt to validate the questionnaire by reference to three external organisations and the difficulties encountered are described.

Chapter seven details the use of the questionnaire within the organisation under study when one of the software packages found to be unsatisfactory is replaced. The results are correlated with a later measurement of user satisfaction using the instrument described in chapter five. The results of discussions and interviews and of the questionnaires pre and post software purchase are discussed in relation to the proposed model. The process is then repeated with the replacement of the second package which had been found to be unsatisfactory. The results are compared with those obtained from the first package and with the proposed model.

Finally, **chapter eight** concludes the thesis. The work performed is summarised and assessed and the direction for future work is proposed.

Chapter 2

Interaction between people and systems

The problem addressed by this research is the difficulty faced by those organisations which do not develop systems in-house when they try to purchase packaged software which not only meets the functional requirements of the organisation but also meets the needs of users. Although use of the package may be mandatory, thus ensuring that the financial outlay is not totally wasted, most organisations would prefer that people are happy to use it. If people are satisfied with the software they have to use they are more likely to try to overcome problems rather than to treat them as obstacles to work. Although it is not possible to point to any definitive evidence that this will lead to increased productivity, it is generally accepted that such will be the case.

Over the past few decades research has been undertaken into how to design software that users are happy with. Much of the research has referred to cognitive theory in order to explain how a person interacts with a computer. This describes the one to one relationship of the interaction that was prevalent when personal computers (PCs) were first introduced. The individual and the computer were considered without reference to the context within which they were operating. This study at the detail level was necessary in order to improve knowledge about the physical interaction of a person with a computer and how to design interfaces people felt comfortable with. However, it was soon recognised that human-computer interaction (HCI) took place within a social context and that software design meant more than designing the look of the software on a computer screen, it also meant designing software to fit the task required of it within a particular context.

The purpose of this chapter is to give a brief overview of the different ways in which the problem of interaction between people and computers has been approached. It summarises the development from the first steps of modelling the process of interaction between a person and a computer (e.g. Taylor, 1988) to investigations into how that interaction might be improved upon, to finally considering not just the person - computer interaction but the effect on person to person interaction made by the introduction of a computerised system i.e the socio-technical perspective (e.g. Nurminen, 1988). Thus the focus may be seen to shift from the computer system to the person to the context within which both operate.

This process has not taken place in the linear fashion described but has evolved in a series of parallel and inter-linking lines. As the emphasis has moved from providing purely functional systems to designing systems that users feel comfortable with, the field of study has necessarily expanded until it may now be said to borrow from several diverse disciplines including anthropology, sociology and engineering (Preece et al., 1994, Barnard, 1991).

The word "system" is used throughout the thesis in the sense of a business system and encompasses the hardware and software used in a business application with the main emphasis on the software itself. The chapter concludes with the research question of how organisations, which do not develop software in-house, may choose software that users will feel satisfied with as well as providing the functionality required by the business.

2.1 Definition of Interaction

The dictionary definition of the verb "interact" is to "act reciprocally". This implies a more equal form of working than would be the case if the computer system merely acted upon an instruction. This suggests that there is a communication between the person and the computer. When two people communicate they do so by taking it in turn to listen and respond. The same process is required between a person and a computer.

However, communication may take more than one form depending on the reason for it. In the case of social interaction between two people it may be simply to enquire after each other's health, or it may be to exchange opinions on various subjects. In the case of business interaction, the communication may take the form of negotiation, information exchange or, from employer to employee, simple instruction. The manner of communication, the interaction between one party and another, will depend on the situation, the reason for the communication and the social relationship between the people. In the early days of their development, computers were viewed quite simply as numeric problem solvers. No more was required of them than that they accept instructions and compute the answer. The people giving the instructions were specialists who used machine specific language and they did not expect any assistance from the machine in how to give their commands. As computer usage spread into business, high level programming languages were developed and programs started to be written which meant that the computer user was no longer interacting directly with the machine. In some cases this has been developed to the point where commands may be given verbally to the machine. However, the computer can still not initiate the interaction (although, with current developments on neural networks and artificial intelligence, this may not remain true (Tan, Quah & Teh, 1996)). At present, the communication is always instigated by the person and the prime medium of communication is language - be it written or spoken. Even if the user seems to be mainly responding to the computer by selecting from a choice list of options, it will still be the user who initiates the session with an objective in mind and uses language as the medium. Despite this, the perception of the user is usually that of having a dialogue with the computer and the computer system is often spoken of as having human attributes such as friendly and helpful or awkward and demanding. The field of Human Computer Interaction (HCI) seeks to make the dialogue with the computer as natural as possible, so that use of it does not require specialist skills.

2.1.1 Human Computer Interaction

The term HCI first started to be used in the mid 1980s (Alexander, 1987; Preece et al., 1994). Whilst its roots are in the design of the interface in order to improve the interaction between person and computer, it now draws upon the expertise found in the fields of psychology, ergonomics and linguistics as well as that of computer science and artificial intelligence (Preece et al., 1994; Barnard, 1991). This reflects both the advancement in computer technology itself and the recognition that issues such as environment and training have an impact on the users of computer systems and, therefore, their reaction to those systems.

HCI does not have a single definition. Some definitions seem to place the emphasis on the human issues e.g. "[a] set of processes, dialogues, and actions through which a human user employs and interacts with a computer" (Baecker & Buxton, 1987, p.40). Others tend towards a definition of an equal partnership e.g. "..fundamentally an issue of communication between a person and a computer".(Maddix, 1990, p.30). Others again seem to view the problem as lying mainly with the technology e.g. "..the creation and evaluation of technology that affects the interactional qualities of computers" (Selker, 1996, p.60). Nor does it have a single theoretical grounding. Whilst much of HCI draws on cognitive psychology to model and explain the interaction of person and computer, it has been suggested that this is not sufficient and other frameworks have been proposed. (Long & Dowell, 1989; Kaptelinin, 1996; Kuutti,1996;Mackay & Fayard, 1997).

These different definitions and the continuing search for a theoretical framework are a result of the broad area which HCI now encompasses and the different perspectives from which the problem may be viewed as the technology develops. The subject has developed as the personal computer has become the interpersonal computer and the "interaction" in HCI now means rather more than just the interaction between one human and one computer and also seeks to address the problem of groups of people working with each other by means of computers.

2.2 Modelling the interaction

The early studies into how people interacted with computers focussed on trying to understand what that interaction comprised of. Different models and tools have been created in an effort to describe what is happening. These take various forms and consider the interaction from different perspectives:

a) **The systems perspective** (Nielsen,1986;Maddix,1990;Taylor,1988;Christie,1985; Cohen,1986; Clarke, 1986). In this both the person and the computer are seen as components of the system. For example, Taylor (1988) suggests a layered protocol model to describe the communication between a person and a computer. By this he means a step by step description of what takes place during the interaction. The communication between person and computer may be compared with that between person and person. In order for any two parties to communicate, a protocol is needed. This defines the form by which communication is effected e.g. when there should be a pause to indicate it is the turn of the other participant. It does not describe the content of the communication. Taylor suggests the communication between person and computer should be considered at four levels, that of the physical, the character, the word and the command. The use of layered protocols means that whilst communication may appear to take place at one level, it is actually effected at a lower level. Similar models are used to describe the communication between computer and computer.

b) **The perspective of the task or goal** (Payne & Greene,1989; Moran, 1981; Diaper, 1989; Card, Moran & Newell, 1983). In order for the computer to give the correct responses to the actions taken by the person, the steps needed to fulfill a task need to be taken account of. The Keystroke Level Model (Card et al.,1983) was developed to model the task in order to see how long it takes to perform an action with the system. It assumes that the person performing the task is an expert. Assessing the time taken to perform the task, including thinking time, helps to assess the overall efficiency of a system in use. Task Action Grammar (TAG) (Payne & Greene, 1989) was designed to help to formalise the conceptualisation of the task to be performed to the action needed to perform it. A task is defined at the simple level such as the action "to delete". Similar tasks should require similar actions to enable ease of learning. So, for example, if a certain key is pressed to exit a screen in one part of the system, the same key should be pressed to exit a screen in another part of a system. Representing the actions to be performed in order to achieve a goal in the form of a grammar enables the designer to see if there are inconsistencies in the way in which the system operates.

c) **The perspective of the user** (Christie, 1985; Shneiderman, 1982, 1988, 1991, 1992; Manketlow & Jones, 1987; Staggers & Norcio, 1993; Wallace & Anderson, 1993). Shneiderman (1988) considers the different types of interaction (e.g. menu-driven, direct manipulation) that might be made available and their benefits and drawbacks from the point of view of the person using the system. Manketlow and Jones (1987) consider to what extent an individual will have a mental model of a system and how that might help or hinder their interaction with a computer system.

d) **The perspective of the designer** (Johnson-Laird, 1983; Clarke, 1986; Preece et al., 1994; Dagwell & Weber, 1983). In the same way that the person who will finally use a computer system may have formed a mental model of that operation at the start of the project, the designer will both have his own model of the system and a model of the person who will use it (Dagwell & Weber, 1983). Clarke (1986) suggests a model of the human-computer interface against which a designer could check his interface design.

The aim of each of the above is to assist in the design of a computer system in order that the person using it might feel comfortable with that interaction and no one, on its own, claims to be sufficient to achieve that. The first models were developed when Personal Computers (PC) were in their infancy and the main concentration was on the user as an individual worker with a one to one relationship with his PC. More recently as the power of the PC has increased (often beyond that of old mainframe computers) and they have been networked together, research is being undertaken into modelling Computer Supported Co-Operative Working (CSCW) and how a system mediates between individuals working as a team. (Dourish, Adler, Bellotti & Henderson, 1996; Furner, Watkinson & Clark,1990; Olson, Card, Landauer, Olson, Malone & Leggett,1993; Grønbæk, Kyng & Mogensen,1996; O'Malley, Langton, Anderson, Doherty-Sneddon & Bruce, 1996).

The work carried out to better describe and understand the form of the interaction has assisted in the design of the interface to improve that interaction.

2.3 Design of the interface

It is generally accepted that how information appears on the screen and how intuitive it is to use will influence how the user feels about it. Research into what constitutes good interface design covers a broad area from the detailed research into the use of screen colour and the optimum number of menu options (Kiger, 1984; Murch, 1987; Rubin, 1988; Mosier & Smith, 1986; Parkinson et al., 1988) to the design and use of interface metaphors (Gentner & Nielson, 1996; Carroll et al., 1988; Anderson et al., 1994). In the early days of computers, only character based interfaces were available. Therefore research focussed on optimising use within this constraint. At the same time, work was being undertaken into how to make the PC seem more like a workarea with filing cabinet and waste basket (the desktop metaphor)(Canfield Smith et al.,1982). Improvements in technology have made this possible and the question now is how to represent objects in a meaningful and useful way.

Much work has been carried out into how to design an interface which is intuitive and natural to use (Ogden, 1988; Gentner & Grudin, 1996; Grudin, 1989; Grudin, 1992). However, simply being easy to use is not sufficient. Interaction with the computer must also serve some useful purpose. In order to be useful, the computer must enable the user to carry out a task more efficiently (either in terms of speed or of quality) than would be the case if it were done manually. An understanding on the part of the designer of the task to be performed is, therefore, essential. This entails a shift in focus from how the system should be presented to the task the system must perform.

2.4 Analysing the Task

In order to design an interface a software engineer must decide what actions are to be made possible via the intermediary between person and computer. In order to do this the task to be achieved must be understood. Various approaches have been used in order to gather information about the task and how it is carried out, from the more formal task analysis methods (Stammers & Shepherd, 1995; Whitefield & Hill, 1994) to the less formal methods where the analyst acquires information by actually working alongside the intended users of the system (Shapiro, 1995).

Task analysis maps the links between one step in a task and the next and the various decision points which may occur. However, whilst task analysis may be helpful to the designer in building the program logic, used on its own, without further reference to the user, it will not suffice to create an acceptable interface. Understanding what is required is more than just understanding the functionality of the system, it is understanding how that functionality is to be put into action. It is in this area that the role of system designer has been altered over the years. The classic role of the system designer is to translate the required functionality of the system into a flow which the programmer can take and code into a computer system. The missing step here is the representation of the functionality to the user - the design of the user-interface.

The difficulty in achieving this has been recognised for some time and there are various discussions in the literature on the user's and designer's mental or conceptual model of the system, how they are acquired, what influences them and the problem of marrying the two together without, on the one hand, the designer slavishly following the user's model or, on the other hand, disregarding the user's wishes on the basis that the system would function better another way (Sutcliffe & Springett, 1992; Christie, 1985 ; Kieras & Bovair, 1984.; Downton, 1993; Roast,1994; Manketlow & Jones, 1987). This points to a much closer involvement of the user in the design process than is achieved simply by analysing the task they perform. It was recognised at an early stage that, in order to ensure that this is the case, the user must be involved in the process of design (Zmud, 1979). In this the focus now shifts from the characteristics of the system to the characteristics of the user and how he perceives the system.

2.5 User involvement in the design process

Cognitive style, personality and demographic variables are all reported to have an effect on the success of a computer system (Zmud, 1979; Ramamurthy, King & Premkumar, 1992; Martinko et al., 1996). Intelligence and level of education, whether someone is introvert or extravert, age and gender have also all been shown to have an effect on MIS success, although the reported effects have not always been consistent (Zmud, 1979). Potential effects of the user characteristics on the success of a system are more likely to become apparent if the user is involved in the system design.

Another key argument for user involvement in the design process has been that a main factor in the failure of computer systems is the alienation often felt by the users (Hirschheim & Newman, 1988; Clegg, 1988). However, what constitutes involving the user in the process of design has several interpretations. Asking people about their requirements at the start of the design process could be said to be involvement. Similarly, asking them to evaluate the final product could be claimed as involving the user. Neither of these is sufficient if potential problems with the system are to be flagged before change becomes too costly. The user must be involved at all stages of the design process (Mantei & Teorey, 1988; Shackel, 1990). To try to achieve this, different forms of prototyping have been used. There is the simple technique of using paper cut outs to represent actions and what happens as a result of those actions; the more expensive (in terms of tool, although maybe not in terms of time) use of computer based presentation software to represent screen layouts and the results of actions taken on them; and the prototype system which is then further developed into the final product. Involving users in this way in system design and development has influenced formal methodologies such as SSADM which now includes prototyping and design iteration as part of the life cycle (Weaver, 1993). It has also encouraged the development of the Star Life Cycle which advocates evaluation at all stages (Preece et al., 1994). The involvement of the users generally takes place at specific points in the life-cycle from the initial discussions to evaluation prototypes or separate system modules until the final product is ready for acceptance.

The philosophy of Participative Design (PD), which has its roots in the Scandinavian countries and aims to promote the control of workers over their work, goes beyond simply involving the users at different phases of the development process. Asking people to evaluate a prototype which is then revised by the designers and re-submitted for evaluation, is not sufficient to be classed as involvement. PD advocates active participation in the design and development of a system (Kensing & Munk-Madsen, 1993; Muller et al.,1993). This is often difficult to achieve. If a system is being developed for a large group of people in a company, it is difficult to bring them all together at one time and it could be a lengthy process to reach a design decision. In recognition of this difficulty, research has been undertaken into how to select representative users according to their individual characteristics (Hawk, 1993).

The Joint Application Design (JAD) methodology has PD at its roots but specifically does not involve all users of the proposed system. Representatives of the users are co-opted into the design and evaluation process. JAD imposes a formal structure which helps to ensure adherence to timescale and is largely used in the US (Carmel, Whitaker & George, 1993).

The main emphasis of the work in both PD and JAD has been on the development of a software system within the organisation. The whole project, from feasibility through analysis, design and evaluation is assumed to be undertaken in-house.

The Dynamic Systems Development Methodology (DSDM) is a UK developed methodology which has the philosophy of PD at its roots and a structure which is similar to JAD. Its primary aim is to produce systems on time and to budget without sacrificing user involvement in the design. Ambassador Users are nominated to keep the communication going between the developers and end users. Whilst this method has been used for the implementation of packaged software, all reported uses have been for when the package requires tailoring to the specification of the organisation in which it is to be used.

Although JAD and DSDM may be used by a software supplier rather than an in-house development team, the result is still a bespoke system, rather than an off-the-shelf package which may be used by several organisations.

"Co-development", as described by Anderson and Crocca (Anderson & Crocca, 1993) is similar to each of the above. Software engineers and users are brought together on a specific project with the aim of helping the software engineers to understand the needs of the user. However, users are not always able to explain what it is they want to achieve with a system so even this close working relationship does not necessarily overcome the communication problems between user and designer.

Research has also been carried out into how a software supplier may work with the potential customers in order to create products which are acceptable across different organisations (Eason & Harker, 1988: Keil & Carmel, 1995; Beyer & Holtzblatt, 1995; Hutchings & Knox, 1995;

Poltrock & Grudin, 1994). However, the literature focuses on what is required on the part of the supplier in order to design a package which will be suitable for a range of organisations. This does not offer assistance to the software purchaser for an organisation, who needs to ensure that a package will be satisfactory to the users in that organisation. When a package is to be purchased, installed and used with no tailoring the question must be posed of how the user is to be involved.

2.6 Evaluation

Continual evaluation of a system, whilst it is being developed is advocated by the adherents to the PD philosophy. This is because "It [PD] views the users' perceptions of technology as being at least as important to success as fact, and their feelings about technology as at least as important as what they can do with it" (Schuler & Namioka, 1993, p.ix). Involvement of the users all the way through the process is assured if people are continually reviewing and evaluating what is being offered and this may help them to adjust their expectations and be more likely to lead to a favourable perception of the system (Grønbæk, Grudin, Bødker & Bannon ,1993).

However, evaluation itself is a broad field with research into when evaluation should be carried out (Christie, Scane & Collyer,1995) and the different ways in which it may be performed (Ravden & Johnson, 1989; Reiterer & Oppermann, 1993; Desurvire, Kondziela & Atwood, 1992; O'Donnell, Scobie & Baxter,1991; Böckle et al.,1996).

Although Böckle et al. (1996) describe an evaluation method which concentrates on the performance of the system, this is not how HCI considers evaluation. It is no longer sufficient to evaluate software according to how well it meets functional requirements: The needs of the user must also be taken into account. As has been said ."Although a system may be evaluated favourably on every performance measure, the system may not be used very much because of the user's dissatisfaction with the system and its interface" (Chin, Diehl & Norman, 1988, p213).

As described above, use has been made of prototypes and system modularisation to enable the user to evaluate the design of a system as it is built. If this is not possible, then the final system should be made available to the users or to human factors experts to evaluate (Molich & Nielsen, 1990). The latter method has been found to be successful in detecting problems with the interaction but it loses the benefit of end-user evaluation in that the people who will be using it are presented with a finished product.

The need to have the software available for a period of time in order to evaluate it presents a problem to the organisation which is purchasing packaged software. Whilst it might be possible to buy a single user PC package (e.g. for wordprocessing) and discard it, having given a small number of people training on it and time to use it, this is not possible for packages intended to replace large manual systems, as the initial outlay would be too great. Nor is it always possible to introduce a system on a gradual basis. What is required is a means of predicting the likelihood of a package having the potential to both satisfy the needs of the organisation and of being properly utilised to that end by individual users who feel satisfied with its use.

2.7 Socio-technical aspect

The interaction between a person and a computer does not take place in a vacuum. There will be a social context be it that of the home or the workplace. When the context is the workplace the characteristics of the organisation including factors such as the power politics and managerial style will have an effect on the interaction (Keen, 1981; Markus, 1983; Kling & Iacono, 1984). It is this need to fit the technology to the functional and social needs of the organisation that the socio-technical aspect of HCI addresses (Mumford,1993). It recognises that acceptance of a computer system and satisfaction with it depends on more than just the features of the system and the users' interaction with it (Mumford, 1991; Damodaran, 1996).

The Effective Technical and Human Implementation of Computer-based Systems (ETHICS) technique (Mumford,1991;1993) considers the introduction of a computer system as only a part of the whole process of change. It aims to balance the social and technical aspects of a revised way of working according to need. The introduction of technology should not be considered as something apart from the rest of the working environment as it will have an impact upon it. The emphasis on the human issues should be equal to the emphasis on the design of the computer side.

Thus the socio-technical aspect of HCI extends the human factors aspects beyond the perspective of the interaction of an individual with a computer system to taking account of the social factors which affect an individual's attitude towards that computer system.

PD, arising as it does initially from the desire to give people the means of controlling all aspects of the environment in which they work, from the physical design of the workplace to being involved in the long-term planning and decision making of the organisation and, therefore, being concerned with the social aspects of the introduction of new technology (Kensing & Munk-Madsen, 1993), perhaps more properly belongs under this heading rather than under that of user involvement in the design process. The PD philosophy goes beyond the focus on the interaction between software engineers and end-users and how to improve the amount and quality of communication between them. Its starting point is the operation of the organisation and how workers and management may co-operate to make improvements. Indeed Ehn (1993) makes a distinction between PD and the socio-technical approach because the latter does not give explicit power to the trade union side and actions are instigated by management. It is important to involve all levels in the organisation in the design and implementation of the software. However, whilst reported PD projects recognise this, they do not report explicitly on the organisational structure or politics in place at the time (Clement & Van den Besselaar, 1993). The concentration is on the means of ensuring manager and user participation in a particular project (e.g. Williams & Begg, 1993). This means that the difficulties or benefits presented to PD by different organisational structures can not be compared. Wagner (1993) does report on the problems faced by a software designer when attempting to design a system to support open, co-operative working in an organisation where it is not usual for decision making to be open and which relies on the judgement of specific individuals. However, the paper raises the problems and questions to be addressed rather than offering solutions.

With its concentration on the balance between the organisation and technical issues, the socio-technical approach is applicable to the introduction of packaged software as well as to software being designed in-house. However, where such an approach is reported, it is generally with regard to software which is either developed in-house or has undergone substantial tailoring.

2.8 HCI issues applied to packaged software - the development of standards

In recognition of the fact that software purchasers need some assurance that human factors have been taken into account during the development of a package, software design standards have been advocated.

The ISO 9241 standard covers the requirements for all office tasks which use a visual display terminal (VDT). It is divided into several parts and part 11 deals specifically with usability. Usability is defined as "The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" (ISO/DIS 9241-11 p.2, 1995). The standard gives guidance on how usability should be assessed. It recognises that the degree to which a product is deemed usable may vary depending on its context of use and, in order to allow flexibility, the standard only offers guidelines not rules.

Evadis 11, as described by Reiterer (1992), is a method by which software may be evaluated against ISO 9241. There are 5 stages to the evaluation - examination of the software itself; examination and selection of required tasks for testing; exploration of user characteristics via a questionnaire; evaluation of the software via the test tasks; and interpretation of the results culminating in a written report. The evaluator should preferably be a human factors expert. Whilst Reiterer (1992) says that the method is suitable for purchasers of software as well as developers, the process of evaluation would seem to be rather lengthy and complex and there are few software purchasers in small to medium size companies who are expert in human factors. In addition to this, it is questionable whether an HF expert would find the problems that occur as a result of the context in which the software is being used. Dzida (1995) suggests a two stage evaluation method for the application of the standard. First of all he points out that "..it may occur that a product conforms with the standard but lacks usability" (1995, p.94). In order to try to avoid this, the first step is to assess the applicability of the standards in their context of use i.e. "regarding the user characteristics, the task and organizational requirements as well as the physical environment" (1995, p.95). The second step is to evaluate the software against the applicable standards. Again, this is not a simple process.

Both the methods suggested as a means of evaluating software against ISO 9241 and, indeed, the standard itself would seem to assume that either a product is being custom written for a company and so is available for assessment in all its parts or else that the software is already in use in a company.

An evaluation method proposed for use in evaluation of in-house products, evaluation of purchased products or evaluation by a third party (neither vendor nor purchaser) is the Software Assessment and Certification Program, Europe (SCOPE) (Welzel & Hausen, 1995). The guide to the evaluation method which has been produced assists the evaluator in determining which aspects of the software should be evaluated depending on the requirements of the client (person wanting the evaluation). The method used concentrates on the attributes of the software itself as specified in ISO/IEC 9126 rather than the human factors aspect. Again, the purpose is to obtain/ award certification for software quality. The fact that software achieves this standard does not necessarily mean it will be suitable for all organisations.

The draft standard ISO/IEC DIS 12119 is also intended as a tool to help software purchasers determine which products are likely to meet their needs (Wegner, 1995). The standard defines the documentation which must be available with the product. This documentation must state the intended work task, requirements for hardware etc. The standard also gives guidance on how a product and its documentation should be tested against the quality requirements. If a product passes the test in all areas, it will be awarded a Quality Mark. However, Wegner (1995) points out that "the fine products with the Quality Mark have moderate success, [whereas] software packages with rich functionality, violating many requirements of this standard swamp the offices"(p.117). This may, of course, be due mainly to the marketing of the other software as high sales do not necessarily indicate satisfactory use.

Whilst both the certification and quality mark proposals go some way to assisting a purchaser to determine how well a product performs and how well it meets certain design standards, they are not a means of ensuring that users will be satisfied with the software.

Developments are also underway to establish Common Industry Format tests (CIF)[2] (Bevan, 1999a). The aim of the initiative is to raise the profile of usability both in supplier and purchaser companies by making it part of the acceptance criteria of a product. As described by Bevan, the usability tests may be carried out either by a software supplier and the reports made available to a potential purchaser, or they may be used by the purchasing organisation. These tests are not just to demonstrate ease of use but to show "the extent to which the product can meet the needs of the user as a result of providing appropriate functionality, performance, [and] reliability" (Bevan, 1999b*). This is described as "quality in use". However, in order to test the software against this criterion, various steps are required through which representative users are selected to carry out tasks as they would be performed in a normal working environment. These together form the

[2] CIF is an initiative supported by the US National Institute of Standards and Technology (NIST) in co-operation with ACM SIGCHI

Context of Use i.e. the characteristics of the users in terms of age, experience, knowledge etc.; the type of task (that is the whole task of which computer action forms a part) and the total work environment in terms of lighting, furniture, temperature, any legislative requirements, work practices and organisational structure. Having established the context of use, the usability requirements must be specified i.e. acceptable levels of efficiency, productivity and satisfaction. At this stage it should be possible to set up an environment for evaluation of the software.

Whilst CIF does recognise that the feelings towards a piece of software may be different in one company from another, the investigations that are suggested could be very time consuming and beyond the scope of small to medium size organisations, especially if more than one package is to be considered.

2.9 Summary

There are essentially two areas of knowledge which contribute to the discipline of HCI - that of Human Factors and that of Software Engineering (Long & Dowell, 1989).

Software engineering is primarily concerned with the design of the computer system itself (its appearance and capabilities). This may be seen in the work on modelling the interaction, designing the interface and modelling the task.

Human factors is concerned with the people issues (ergonomics, social context). These are the main focus of the work describing the effects of user characteristics, involvement in design and evaluation and the influence of organisational factors.

As the use of computer systems and their technical capabilities has extended so has the scope of HCI. However, there is always an underlying assumption that the people using those computer systems may, in some way, be involved in their design, development and evaluation or, at least, have some control over how or whether to use them.

The purpose of the various standards is to give some assurance that human factors have been taken into account during the design of software packages written for general distribution. However, these are limited to generalisable factors such as ergonomics and can not take account of the social factors in different organisations. The proposed CIF tests will enable companies to make choices between packages based on reported testing, but these tests will necessarily be away from any organisational context or else specific to one. To carry out its own evaluation, an organisation will need to invest both time and money which will add to the cost of the project.

It would be preferable for small to medium size organisations if there were a means of assessing the acceptability of a package i.e. likely user satisfaction with it pre purchase without the need to spend time evaluating it for anything other than how well it meets the functional requirements specification. It is this problem which this thesis seeks to address in the following chapters.

The next chapter defines what is meant by satisfaction with software and considers the factors derived from the literature which contribute towards it.

Chapter 3

Factors affecting user satisfaction with software

The previous chapter finished by arguing that the focus of most of the work on user acceptance of software has been at the design, development and evaluation stages of the life cycle. The people who will be the end users of packaged software are not present during these stages of the life-cycle. Even evaluation methods are dependent on having the software available for a sufficient period of time for people to gain some familiarity with it. This is rarely possible when an organisation is buying commercially available software as it would effectively mean installing the software and providing training.

The new ISO standards are intended to ensure that software meets basic usability criteria and methods of testing this are being developed, as are methods of ensuring that software continues to evolve in line with business requirements (Hather, Burd & Boldyreff,1996; Bennett,1996; Reiterer,1992; Reiterer & Oppermann, 1993). However, the ISO standards cannot ensure that a piece of software is suitable for a particular organisation unless the testing is performed within the organisation itself. For example, the Software Development and Certification Program Europe (SCOPE) project has detailed a set of evaluation methods by which a software product may be assessed (Welzel & Hausen, 1995). This, of necessity means that the full software product must be available for evaluation. Third party laboratories would be able to undertake the evaluations on behalf of the potential purchaser. The first step would be to agree what is to be evaluated and to what level. What is important to one organisation in terms of functionality, may not be given the same importance by another organisation, so the evaluation is effectively on a per company basis. Under the scheme it would be possible for a software house to submit its own product for evaluation and receive a certificate of quality. However, the software house would specify the criteria against which the software was to be evaluated. Provided the software met those criteria (which would have to also be within the scope of the ISO standards ISO/IEC 9126, ISO 9000 to 9004) and fulfilled what it claimed, it would receive a quality certificate (Welzel

& Hausen, 1995). The possession of such a certificate would serve only to confirm that the software, as a product, met a certain standard. The meeting of such a standard would not guarantee user acceptance in every organisation. Nor is any evidence presented to show that a software package achieving the quality assurance certificate has resulted in user acceptance.

The difficulty in doing anything other than measure on a general basis is confirmed by Dzida (Dzida, 1995), who states that "any specification of requirements concerning task and user would produce a tremendous proliferation of standards, and it may be questionable whether scientific research will ever be able to provide empirical evidence for requirements comprising all combinations of user, task and environment" (p60). However, even if this is accepted, a means of assisting an organisation in assessing whether a new software package is likely to result in user satisfaction, without first investing time and money in the product, is needed.

Of course, software houses do not write products without taking any account of their targeted customers. Before writing a line of code, they will have gathered information through some form of market research be it by use of focus groups, questionnaires or other methods. Holtzblatt and Beyer (1993) advocate working very closely with potential customers and using the Contextual Inquiry technique to test out the software model so far designed, within the context of the workplace. However, it is not generally commercially possible for a software developer to spend a great deal of time with different organisations, neither has the software purchaser any means of knowing whether any organisations which were involved in such a development have any similarities to the organisation being bought for.

Software, then, is written to fill a perceived market need and, as the adherence to standards starts to become recognised as a stamp of quality, so more software houses are likely to adopt them. But from the point of view of the organisation buying the software, what is also important is to know that any system purchased is going to be accepted in the workplace. In order to do this, account needs to be taken of the various factors which affect user satisfaction with software.

This chapter considers the factors which are proposed in the Human Computer Interaction (HCI) literature as influencing user satisfaction. Section 3.1 first defines the scope of that satisfaction. Section 3.2 then proposes the factors which affect that satisfaction, under the three main headings derived from Chapter 2. Each of the factors is discussed and the reasons for the inclusion of each as a contributor to user satisfaction is given. The chapter concludes by asking the question whether these same factors are key with regard to user satisfaction with packaged software. This question is then addressed in Chapter 4.

3.1 Scope and meaning of satisfaction

Although this study is concerned primarily with satisfaction with computer software, reference has been made to investigations into the acceptance of "new technology". Software has to run on hardware and the speed of the hardware may affect perception of the software, so the software should not be taken in isolation from the whole system. It has been assumed that, in the workplace, the same factors are at play with regard to satisfaction with new technology, satisfaction with information systems and satisfaction with software applications, and the terms have been used interchangeably. Different methods of measuring satisfaction exist and these are discussed later.

Satisfaction with something may be defined as having one's desires or expectations met. User satisfaction, as used here, means that people are content with the use of the software. In the literature, user acceptance of software is generally taken to be synonymous with user satisfaction in the sense that people are happy to use it. Whilst acceptance does not necessarily equate with satisfaction - it is possible for a situation to be unsatisfactory but accepted - it is generally used in the literature in the sense of something being acceptable and, thereby, satisfactory. MIS[3] success is also a frequently used term in the literature. By this is meant the benefit that the introduction of information systems has brought to the organisation. As such success is difficult to measure, it tends to be related to the amount of use that is made of a system and that, in turn, has been taken to indicate satisfaction. For these reasons, the investigation into the literature on the factors affecting user satisfaction with software has also referenced studies into the acceptance of technology and MIS success as well as those which explicitly investigate satisfaction. Satisfaction with software is not the same as software usability although it may form a part of it. Usability has been given various definitions but all require that there be an element of efficiency and productivity as well as that people are satisfied with use of the technology.

[3] Management Information System. MIS or IT dept. is also the department in an organisation responsible for providing Information Services

3.2 Contributors to user satisfaction

In addressing the problem of interaction between person and computer in the workplace, there are three global areas of concern - the computer system, the individual and the organisational context. In the previous chapter these were described as:

- System Characteristics (2.2 - 2.4)
- Individual Characteristics and Perceptions (2.5 - 2.6)
- Organisational Characteristics (2.7)

Each of these contains various factors within it. These are now discussed with regard to their effect on user satisfaction as defined in the broad terms above. The studies described tend to focus on factors belonging under just one of these headings but include references to factors which have been placed under a different heading. For ease of discussion, Individual Characteristics and Individual Perceptions are treated as separate headings.

3.2.1 System Characteristics

The System Characteristics are those elements which distinguish one system from another. Three factors are proposed as belonging under this heading. These are:

- Functionality
- Match to the Task
- Usability

3.2.1.1 Functionality

Specifying the functional requirements of a system forms part of all standard methodologies (e.g. SSADM). The functional requirements specification generally entails detailing what the system must be capable of. How many records may be held, time taken to process, which data fields must be held, printing capabilities, are all part of the functional requirements. If a system is to be designed in-house, a requirements specification of the functionality is agreed at the outset and used as part of the system evaluation during and post development. If a commercial off-the-shelf system (COTS) is to be purchased, the requirements specification may be used to evaluate and compare potential systems. The functionality of a system means all the functions it is capable of performing and these may extend beyond those required for the task itself. Even games software may be said to be functional in that it provides a leisure activity. Alternatively, a system which has many features may still be rejected because it is not capable of performing the key functions (tasks) required of it.

Functionality, then, is the capability of the software to enable the required task to be performed. It is an essential part of the system which would have no meaning without it.

Functionality may then be said to be a key characteristic of a system, without which people will not be satisfied with the software simply because it will not enable the required task to be performed.

3.2.1.2 Match to the Task

The software should not be designed in isolation from the situation in which it is to be used. Various factors external to the computer system may affect how data may be prepared and input to that system. This should be taken account of so that the system helps the workflow.

Although it is necessary to establish the overall requirements of a system at the outset in order to be able to measure the final system or to assess progress during development (Christie, Scane & Collyer, 1995), the requirements specification will not generally detail the actual method of system operation. When a system is to be developed in-house this lack of detail in the requirements specification may not be important as it may be incorporated during the design stage. However, for a COTS system it is necessary to know the required method of working as it is possible for the same end result to be achieved by different approaches to the task. As Gentner and Grudin (1996) say "Even when tasks are nominally the same, differences among people and work situations mean that one user-task model will not fit everyone"(p.30)

In support of the argument of the need for Match to the Task, Goodhue (1995) argues that the better the task-technology fit, the better the job will be performed. According to the Task Technology Fit (TTF) model (Goodhue, 1995), user evaluations of a technology will reflect not just how they view the system itself but also how well they perceive that system to meet their task needs and their own abilities.

Clegg (1988) describes this overall fit to the task as "appropriate" technology i.e. the technology supports and enhances the role of the operators as well as improving the efficiency of the organisation. The case for appropriate technology focuses on factory automation and considers a broad view of technology but the need to match the tool to the task, for the appropriate tool to be used, remains the same even at the more specific level of software.

Kyng (1991), Greif (1991) and Rasmussen (1992) all variously make the point that it is necessary to fit a system to support the way in which people work and, in order to do this, the work practice must first be understood. Rasmussen (1992) further makes the point that, as technology itself changes the way people work, "design cannot be based on transfer of an existing work practice to the new medium" (p.4). Not only may the same task be achieved by different approaches, but the computerisation of a task may itself alter the way in which the task is performed, perhaps imposing a standard way of working, perhaps requiring a new skill to be learnt (Medcof, 1989). Neither can it be assumed that because a system may fit a task in one organisation that it will necessarily meet with the same success in another. Whilst it may appear that the job to be performed is the same, the needs and personality of the person performing the job may affect how the system is used and viewed (Eason & Damodaran, 1981).

An example of the way in which computerisation of a task may affect the performance of that task is given by Webby and O'Connor (1994). Their study showed the effect of Decision Support Software (DSS) on how some people manage the task, perhaps taking longer to reach a decision because more alternatives are offered. This may or may not be a good thing depending on the task needs of the organisation. The type of task and the way in which it is computerised may also affect the job satisfaction of the people performing it. For example, a task at the simple but structured level of data entry operators, where the task is to key in data in a fixed format, needs to be computerised such that people are neither confused nor bored by the speed of response of the system. In such a case, the system response time may make the difference between an operator feeling satisfaction or dissatisfaction with the job (Barber & Lucas, 1983). For the system to properly match the task, it must enable the user to work at a pace which is suitable to him. This demands flexibility of a system which is designed to be used by different people.

However, there are times when the flexibility of a system may cause conflict (Kjær & Madsen,1995; Turoff, 1982). For example, if a system is a single user system, it may be possible to alter the way in which the functionality is presented in terms of screen layout and shortcut options from one user to the next without impacting on the use of the core system. For group working this may not be so simple. Such a system needs co-operation between users which may need to be obtained through uniformity of, say, form presentation. This may cause disagreement between people who used to have differing approaches to a task under a manual system and who believe that their way of working is better. Such conflicts are not caused by the system but are brought to light as a result of its introduction

Match to the Task is more than designing (or purchasing) a system that fulfils a specific function. Consideration must be given to how that function is currently fulfilled. However, to build a system based on existing work practices would be to ignore the fact that the introduction of the system will itself have an impact on working practices. It may be possible to increase the scope of an otherwise boring task by the introduction of a system or to inadvertently make a tolerable task intolerable by virtue of the response time. This poses an interesting problem for the purchaser of a COTS package.

Functionality is essential and the task to be performed determines what that functionality needs to be. However, it is important to match the operation of that functionality to the method of performing the task, which itself is situated within the social framework of an organisation. Task Match is, therefore, placed as an element of System Characteristics which impacts on user satisfaction.

3.2.1.3 Usability

In order to meet the requirements specification a system must have the correct functionality. One system is distinguished from another in how that functionality is presented and whether it matches the required method of working. However, it is generally accepted that people also want software that is user-friendly or "usable". Whereas functionality may be broken down into the components that a system must be capable of and task match may be checked to some extent by performing a walkthrough of system operation with the users, usability is not so readily defined.

The term "usability" is often used in place of "ease of use": Software which is easy to use, it is argued, must be usable. It is this aspect of software usability on which research into interface design often focuses. Much work has been undertaken into which interaction styles are easier to use (Dix, Finlay, Abowd & Beale, 1993; Molich & Nielsen, 1990) and which are more efficient in terms of speed of achieving a goal (Lim, Benbasat & Todd, 1996). However, what is easy to use for one person may be cumbersome and slow to another; what is efficient to one may be complex to another (James, 1981).

The difficulty of designing something to suit a range of people is not peculiar to software.The field of ergonomics is evidence of this. However, as the results of studies in ergonomics have shown, it is possible to design for flexibility. In this respect, Usability may be defined as something that may be adjusted to suit an individual whilst retaining the core functionality. Various means of doing this have been suggested, one of which is to separate the interface from the system to the extent that it acts as an intermediary between the user and the software. This may take the form of "protective ware" which interprets the intention of the user and passes instructions on to the core system (James, 1981) or of an "intelligent" agent which assists the user's navigation of the system (Preece et al., 1994). Kilgour (1987) argues that the interface should be seen as an entity completely separate from the system.This view of the interface suggests that it is possible to have the same core system but with several different interfaces each of which will present the system differently according to who is using it. Enabling the individual to determine how the system should be displayed should result in the interface being more meaningful to the user and, thereby, easier to use. However, software usability is more than just ease of use.

Fitzpatrick and Higgins (1998) argue that Functionality and Match to the Task (i.e. Suitability) are each an attribute of a software product. Usability itself is not an attribute of the software but, they argue, is an all embracing description of software. In place of Usability as an attribute they put ease-of-use and learnability. Match to the Task corresponds to the "Suitability" attribute i.e. a product is suitable for the task for which it is designed. Thus, under their definition, a "usable" product will be, amongst other things, functional, easy to use and to learn and suitable for the task.

This definition of Usability, encompassing a number of different attributes of the software is perhaps more precise than to show it as a single factor. However, it is questionable whether functionality is one of those elements.

The definition of functionality needs clarification. Functionality is often used to mean all the features of a piece of software both what it can do and how it does it e.g. whether it is possible to undo actions (i.e. restore the situation to as it was); whether options are accessed via menu options or icons on the toolbar; whether there is any help in how to achieve a goal etc. as well as enabling the user to perform the function itself for which the software was written. This meaning of Functionality goes beyond the simple description of the utility of the software although it still describes the software itself. It also goes beyond the strict remit of the requirements check list as it is generally applied.

To assess the Functionality of a software package by also checking the way in which it requires or enables tasks to be performed is to cross the boundary into checking whether it Matches the Task and whether it is usable in terms of being easy to use. In order to perform these checks, it is necessary to know not only what the software is capable of but the means by which this is achieved, whereas an assessment of the functionality of the software is an assessment of the functions that may be performed with it.

Functionality in this thesis means the utility of the software, which it should be possible to measure against the requirements checklist. It is not, therefore, an element of Usability.

The other attributes of Usability, as defined by Fitzpatrick and Higgins (1998) are shown in table 3.1.

Attributes of usable software		
Installability	Ease of Use	Safety
Adaptability	Learnability	Security
Interoperability	Efficiency	Correctness
Reliability		

(shown as a single column by Fitzpatrick and Higgins, 1998, p.18) Table 3.1

Of these, those attributes which appear in the first and third columns may be measured against a checklist to show that the software meets the required standard. Indeed, it is difficult to separate them from the Functionality of the software. For example, a system which is not reliable will hardly serve a useful purpose. On the other hand not all organisations will require software to be interoperable - that would depend on the circumstances of use. Such an item would appear on the requirements checklist.

In the case of the attributes in the middle column, it is more difficult to abstract the measure from the perception of the user. For example, what one person finds easy to use, another will find difficult. However, it is these attributes which are more commonly associated with the term usability.

It would seem that there are two aspects which combine to make up the usability of a software product:

a) The measurable attributes

b) The perceived attributes

These are not different attributes but different perspectives of the same attributes. In other words, whether or not a software package is *thought* to be usable may vary from one context of use to another, the attributes which combine to make up that usability will not vary.

Ease of use and, thereby, Learnability (it is generally accepted that if something is easy to use it will also be easy to learn, as what is being learnt is how to use it) and Efficiency may be objectively measured by measuring user performance. This is generally done by measuring the time taken to complete a number of tasks and comparing that with the number of errors made in the process (Nielsen & Levy, 1994).

Shackel (1991) defines usability as: "..the capability, in human functional terms to be used easily and effectively by the specified range of users, given specified training and user support, to fulfil the specified range of tasks, within the specified range of environmental scenarios" (p.24). This definition places the usability of software as something which may not be separated from its context of use. Within a particular context it must be easy to use the software and that use must be effective.

This definition of Usability is very similar to that of ISO 9241/11 (1995). However, the ISO 9241 definition also adds that people must feel satisfied with their use of software. According to this definition, if people are not satisfied with the software, then it may not be classed as usable even if it is easy to use and effective in use.

Yet there is evidence to suggest that if people find software easy to use they are more likely to be satisfied with it (Nielsen & Levy, 1994). Their meta analysis of the literature on user preference for different interfaces and input devices and their performance with them, found a correlation between ease of use and efficiency of performance. Generally, they found that people reported a preference for those systems on which they performed better (although the correlation was lower for experienced users than for novice users). They also found that individuals tended to perform better with interfaces they said they preferred (performance thus being a function of preference). However, a note of caution is added here and examples are cited of people having consistently preferred to use one method of performing a task when they performed more efficiently with another.

There is a subtle difference here between what people perceive to be easy to use and what is measurably easier to use i.e. how long it takes to learn and to be put to use. In the case of both the definition given by Shackel (1991) and the definition given in ISO 9241, the ease of use and effectiveness elements of usability are treated as measurable. If Usability is to be considered as a characteristic of a system, not an individual's perception of that system, it should be possible to objectively measure it in some way. It is this objective view of software which would be taken in assigning the Quality Mark described by Wegner (1995). It is also used here, under System Characteristics, to mean the measurable attributes of the software.

Usability, used in the sense of measurable ease of use and effectiveness, is placed as the third and final element of the System Characteristics as there is evidence to suggest that people will not be satisfied with a system which is difficult to use and which is not efficient in use.

3.2.2 Individual Characteristics

As has been mentioned, the same task may be performed differently in one organisation from another (Eason & Harker, 1991). Similarly "different systems will be usable to varying extents by different users" according to their characteristics (Dillon & Watson, 1996, p.628). Whilst someone's ability to use a system does not necessarily predict their satisfaction with it, there is some support for this (Nielsen & Levy, 1994) and much of the literature has concerned itself with individual user characteristics and system rejection (Zmud, 1979, Martinko et al.,1996). The key user characteristics which influence satisfaction for which there is support in the literature are:

- Familiarity with the task (domain expertise) (Ramamurthy et al., 1992)
- Prior use of software (system experience) (Ramamurthy et al., 1992)
- Whether the individual has an optimistic or pessimistic outlook (Martinko et al.,1996)

- Prior experience (Martinko et al.,1996)
- Level of education (Rice & Tyler, 1995)

These are each discussed below.

3.2.2.1 Domain Expertise and System Experience

Ramamurthy et al (1992) consider how the different characteristics of people using a Decision Support System (DSS) may affect how effectively they use it. They acknowledge that differences between products, tasks and organisations could also influence the effectiveness of a DSS system, but the purpose of the study is to determine the influence of the user characteristics on how effective the system is. In order to remove the product variable, the same software was used by all user subjects; in order to remove the task variable, exactly the same task was required to be performed by a group of MBA students. Using MBA students as the subjects also meant that any potential variation due to educational level was removed. As the study was undertaken in a single educational environment, the affect of organisational context was also removed to the extent that there were no surrounding politics or conflicting interests.

They define Effectiveness as consisting of a) objective effectiveness (decision performance) and b) perceived effectiveness (user satisfaction with the DSS). Objective Effectiveness is further divided into the speed with which decisions are made and the quality of those decisions.

Having researched the literature concerning user characteristics, they selected those of gender, intelligence, cognitive style, attitude towards computer usage, domain experience, domain expertise, system experience and personality trait (extrovert or introvert) as potentially having an effect on performance and satisfaction.

Cognitive style, personality trait, experience, attitude and some personal/demographic variables were collected by use of a survey instrument. Intelligence and cognitive style were assessed via standard tests. Satisfaction was measured using the instrument developed by Ives et al. (1983).

The Attitude toward DSS usage measured by Ramamurthy et al. (1992) was defined as the pre-conceptions on the part of the user which influence satisfaction with DSS usage. It is the expectation that someone has of the system before having used it. The seven items used to measure Attitude (reliability, quality, creative learning, flexibility, experimentation, facilitation, feedback) do not take account of the possibility of any impact on attitude which may directly result either from prior experience with other systems or from the influence of the attitude of others. As the whole organisational variable has been removed from the study, it would not, indeed, have been possible to consider the influence of others, apart from that of the other students.

Domain expertise (by which is meant the skills and knowledge which result in performance above the norm in a specific domain), system experience and gender were all found to have a significant correlation with satisfaction. However, no significant correlation was found between other individual characteristics and satisfaction. It is worth noting that the questionnaire used to determine satisfaction was devised at a time when a central computer was the norm and satisfaction was measured mainly by checking the users' satisfaction with the results of the system rather than their feelings towards actually using it. However, in terms of satisfaction as an element making up effectiveness, it does measure what was intended.

The correlation between domain expertise and satisfaction was positive. They suggest that this could be because the people who were more familiar with the work domain and, therefore, of what might be required, could see the potential benefits of the system. System Experience, however, was found to be negatively correlated to satisfaction. This, they suggest, could be because users experienced in the use of computer systems tend to have higher expectations of the software which are not always feasible. Whilst females appear to be more satisfied with the DSS than males the correlation is low and it is suggested that this should be viewed in the light of other findings in the literature and should not be viewed as a purely gender related correlation. The hypothesis tested against this element was that male subjects would differ from their female counterparts in the level of satisfaction with a DSS. Neither cognitive style nor personality trait was found to have any significant correlation with satisfaction.

The results of this study by Ramamurthy et al. (1992) suggest that domain expertise and system experience are two individual characteristics which contribute to satisfaction with system use and they have, therefore, been placed under that heading. Gender has been omitted even though a positive correlation was found because of the comments on this element made by Ramamurthy et al.(1992)

3.2.2.2 Attributional Style and Prior Experience

Martinko et al.(1996) are concerned with the individual in an organisational setting and the influences on the individual which affect his attitude towards IT. They argue that it is the users' prior perception of a system which affects how open they are to accepting it. The Attributional Model of Resistance to IT (AMRIT) demonstrates this.(Figure 3.1)

In building the model they concentrate on the factors which cause individuals to reject a system for reasons which have little to do with the functionality of the system. Rejection is defined as lack of use. The individual characteristics they consider are "prior experience" and "attributional style". By the latter is meant the way in which an individual tends to view success or failure - mainly by attributing it to external factors over which there is no control or by attributing it to internal factors such as ability and effort. By prior experience they mean neither domain expertise nor system experience (as considered by Ramamurthy et al.,1992). Prior experience relates to experience of previous system implementations not experience with a particular system or a specific work domain.

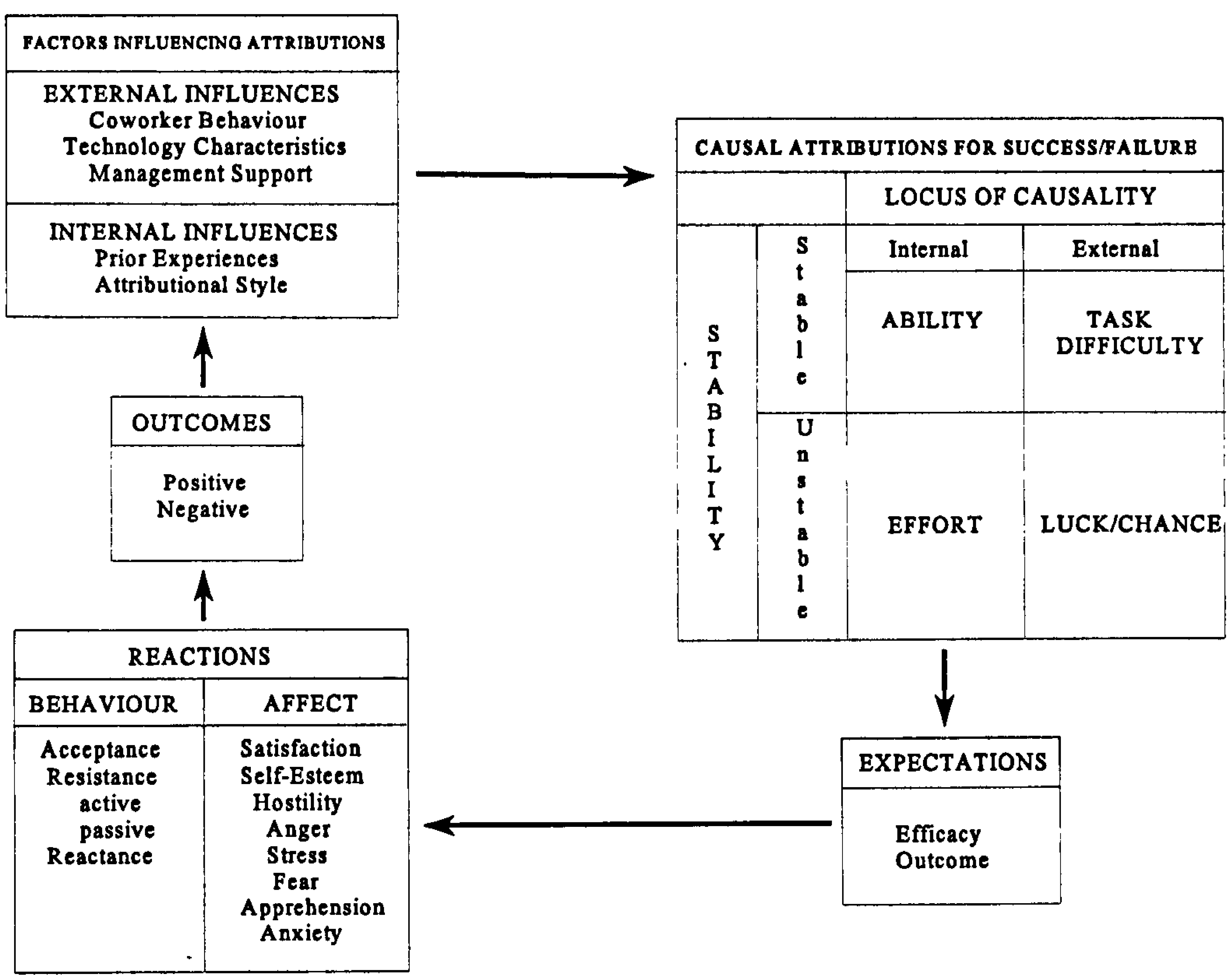


Figure 3.1

The model (which has itself not been empirically tested but draws on the literature as its foundation) is based on the argument that user expectations are influenced by environmental and intrapersonal factors, perceived usefulness of the new system plus previous experience with IT implementations and that these, combined with their own attributional style, causes users to behave in a particular way towards any new system introduced, which behaviour again affects the result and potentially re-enforces their original expectation. This, they say, is in line with Learned Helplessness Theory[4], where experience of failure either on the part of the individual or within an organisation tends to have more influence than experience of success.

AMRIT is very much focused on the individual as the source of resistance or acceptance of software and demonstrates (through the literature) that users may already have formed an opinion about or an attitude to a new system before even seeing it, based on prior experience. The idea that the individual may have formed a positive or negative attitude towards a system before using it, is also used by Ramamurthy et al. (1992). However, the Attitude construct used by the latter is concerned with the individual's pre-conceptions about a particular system, whereas Martinko et al. (1996) suggest that Attitude towards any system is influenced by how successful the implementation of any other system has been. Thus if a previous system implementation was not thought to have been a success, it will be anticipated that the next will also fail.

AMRIT also indicates that the attitude of a new person may be influenced by that of co-workers and that the attitude of the group may affect the overall acceptance of the system. If there is a positive group feeling towards the system then individuals within the group will support that positive attitude. The group feeling itself is heavily influenced by the person or people in charge.

The attributional style of the individual discussed by Martinko et al.(1996), (an optimistic style means that the cause of failure is seen as external and unstable (changing) and a pessimistic style sees the cause of failure to be internal and stable) are similar to the cognitive styles discussed by Ramamurthy et al.(1992) and defined as ".. the interaction of the preferred way of looking at life and the preferred way of judging what is seen" (p.484), but they are not the same. Ramamurthy et al.(1992) are concerned with how an individual processes information and the effect of different cognitive styles on quality and speed of performance. Martinko et al. (1996) are

[4]Learned Helplessness Theory, Martinko & Gardner (1982) Learned helplessness: an alternative explanation for performance deficits, Academy of Management Review, 7, 195-204. Reference from Martinko et al. (1996)

concerned with why an individual may be more inclined to attribute success or failure to factors which are not within his control, not to how the individual processes information. They do not discuss whether the attributional style of an individual has a greater influence than the reported experience of colleagues or than the degree of management support, although they do suggest that lack of management support can result in negative attitudes being formed.

Although it has not been empirically tested, AMRIT demonstrates an argument based on the literature for the effect of both the Attributional Style of the individual and the effect of Prior Experience on the level of satisfaction he will report with regard to a new system. These two factors are, therefore, added to domain expertise and system experience as individual characteristics which affect user satisfaction.

3.2.2.3 Level of Education

Rice and Tyler (1995) propose innovativeness, tenure and education as possible influences on system acceptance with specific regard to the use of Voice Mail (v-mail).

Education is measured by comparing the highest educational qualification gained. Tenure is defined as the length of time in the organisation and may, therefore, be similar to domain experience as used by Ramamurthy et al.(1992). Domain experience relates to familiarity with the environment within which the task is performed. Length of tenure implies that familiarity. However, tenure is measured as length of time with the organisation generally and that does not necessarily mean that experience with the domain in question has actually been gained. Acceptance is determined by the amount of intentional usage and usage extent of the v-mail system.

The findings are that neither individual innovativeness (and, therefore its opposite conservatism) nor length of tenure have any significant influence on system acceptance, but that education does. The hypothesis that those users with higher educational qualifications will be less satisfied with software is found to be supported.[5]

---

[5] Rice and Tyler note the fact that the literature suggests that general education is negatively related to use. However, their model shows education as being positively related to use (p.333). It is assumed that this is an error as their results claim the "predicted" negative relationship (p.336)).

These findings seem to run counter to those of Ramamurthy et al. (1992). In their study, the construct of Domain Expertise is related to education, and Domain Expertise is found to have a positive correlation with satisfaction. Ramamurthy et al. (1992) do say that these findings do not agree with those generally found in the literature. However, Domain Expertise, as measured by Ramamurthy et al. is not the same the Education measured by Rice and Tyler (1995). The latter use the highest educational qualification as a basis of comparison. Ramamurthy et al. use the level of education in a specific field. There is here a distinction between higher educational qualifications per se and higher educational qualifications in the specific field for which a computer system is being used. Zmud (1979) reports that people with higher education tend to utilise MIS less, whereas those with a higher task knowledge utilise it more. Ramamurthy et al.(1992) are measuring higher task knowledge when they measure Domain Expertise. Indeed, they suggest that the reason for their findings is that the people with the greater knowledge in the field for which the software was being used, were also the ones who could see the potential of the tool for being useful. Rice and Tyler (1995) are simply measuring educational qualification. Using this measurement, there would be no distinction between any of the subjects used by Ramamurthy et al. as they were all MBA students. In the case of the Rice and Tyler study, it is also possible that those people who had a higher educational qualification were in posts with a different work emphasis so did not feel that the software supported the tasks they had to perform and that it was this rather than educational background itself which was the influencing factor.

This would suggest that the education construct contains rather more than that which may be measured by considering educational achievement alone. However, there is support for an association between level of education and satisfaction, be that positive or negative hence its inclusion as the fifth element of the user characteristics.

3.2.2.4 Other candidates

The user characteristics so far mentioned are rather different in kind from those suggested by Schneider (1982) who suggests six levels of user characteristic. These characteristics (mode of interaction; object generalisability; operations generalisability; pre-planning; unit and grouping) focus on the characteristics of the user which affect direct interaction with the software not his perception of it. Schneider highlights them as factors to be taken account of at the design stage but suggests that they could also be used as a predictive tool at the requirements gathering stage to match the system interface to the user. For example, "mode of interaction" which means "does the user prefer to be prompted by the system or to control the interaction" could be ascertained and used as part of a system evaluation. These user characteristics describe how an individual behaves with regard to a task, rather than predict what their attitude towards an object is likely to be. Possibly if the design could be matched to the behaviour, this might affect someone's attitude towards the object of their behaviour but that is neither suggested nor discussed. For these reasons they have not been included as elements of the user characteristics which affect software satisfaction.

3.2.3 Individual Perceptions

There is much in the literature to suggest that how users perceive software will influence their acceptance of or satisfaction with it (Wiedenbeck & Davis,1997). These perceptions may arise as a result of prior experience or the influence of co-workers (Martinko et al.,1996) or as a result of how well the system seems to meet their task needs (Medcof, 1989). The perceptions of ease of use, usefulness and of involvement are included under this heading.

3.2.3.1 Perceived Usefulness and Perceived Ease of Use

In the study by Hirschheim and Newman (1988) perceived usefulness or lack of it is one of the factors affecting acceptance of the new system. In their study (which considers the causes of resistance to a computerised system in the Commercial Lines section of an insurance company) they suggest nine possible causes for resistance to change of which only three relate to individual characteristics. All the others relate to external factors which are mainly concerned with the environment within which the individual is working (e.g. management support, involvement in the selection process). The individual characteristics cited are those of innate conservatism (a lack of desire to change), uncertainty (fear of how the system may affect the job) and lack of felt need (similar to conservatism in that there is no perceived benefit to changing). These are completely different from the characteristics tested by Ramamurthy et al.(1992) where the idea that there was no requirement for a Decision Support System was not even considered and the user sample was deliberately taken from a student group so that the organisational variable could be removed, thus meaning that individual attitude to change was not relevant. The characteristics of innate conservatism and uncertainty (which are derived from the literature) are similar to the willingness to change (or lack of it) studied by Rice and Tyler (1995). Lack of felt need as a reason for lack of use is the same as lack of perceived usefulness cited by Martinko et al. (1996,p.317) and it is this that is best supported by the interviews reported. The individual characteristics of innate conservatism and uncertainty are not so well demonstrated by the interviews.

No measure of satisfaction was used in the study by Hirschheim and Newman (1988) so conclusions have to be drawn from the interviews themselves. From the reports of the attitude of the Personal Lines section in the company, it might be concluded that the feeling of satisfaction with the system comes from the control it gives to the users over how they manage their work and that this control enhances their perception of the usefulness of the system. Conversely, the commercial underwriters feel that they have to follow the strictures of the system to the detriment of their work i.e. they have less control as a result of the system and their perception is that it is not useful to them but is there for the benefit of others. This is also a good example of how the introduction of a computer system can alter the task from the user's point of view (Medcof,1989). One of the key points arising from this study by Hirschheim and Newman is that a system must be perceived to be useful in order for it to be accepted.

Perceived Ease of Use is one of the key determinants of satisfaction in the study by Olaniran (1996). This study considers the factors affecting group satisfaction with the results of computer mediated meetings (Olaniran, 1996). The two other determinants are decision confidence and participation. In the test of member satisfaction with performing a task where face to face communication was compared with communication via computer, ease of use was found to be the dominant factor in the computer mediated communication and not in the face to face communication. Olaniran goes on to argue that ease of use must be present in any communication medium in order to ensure its use. The fact that people are so comfortable with face to face communication means that they no longer see ease of use as a problem. It is also pointed out that familiarity with use of the computer mediated system might also change the user's perception over time and remove ease of use as a dominant factor. Even so, the proposal is that the actual need for ease of use will not disappear, nor will the characteristic of the communication medium change, but the user's perception of ease of use and the need for it may change.

In Olaniran's study it is clear that it is ease of use as perceived by the subjects, not ease of use as a property of the medium which is important. It is this perceived attribute that forms the other half of Usability discussed under System Characteristics.

Whilst Hirschheim and Newman (1988) show that perceived usefulness is important to user acceptance and Olaniran (1996) shows that perceived ease of use is important, the Technology Acceptance Model (TAM) (Davis et al.,1989; Davis, 1993) is based on the premise that both perceived ease of use and perceived usefulness will affect the user's attitude towards using a system. The attitude thus formed will influence to what extent the system will be used. As in the studies by Martinko et al.(1996), Hirschheim and Newman (1988) and Rice and Tyler (1995), the amount of use is taken to be indicative of acceptance or rejection. Perceived ease of use is defined as "the degree to which an individual believes that using a particular system would be free of physical and mental effort" (Davis,1993, p.477) and perceived usefulness is defined as "the degree to which an individual believes that using a particular system would enhance his or her job performance" (Davis,1993 p. 477).

Davis (1993) changes his original model slightly to specify a particular factor (system design features) which influences the user's perception of perceived ease of use and perceived usefulness (figure 3.2). He further proposes that a system which is perceived as easy to use is more likely to be perceived as useful.

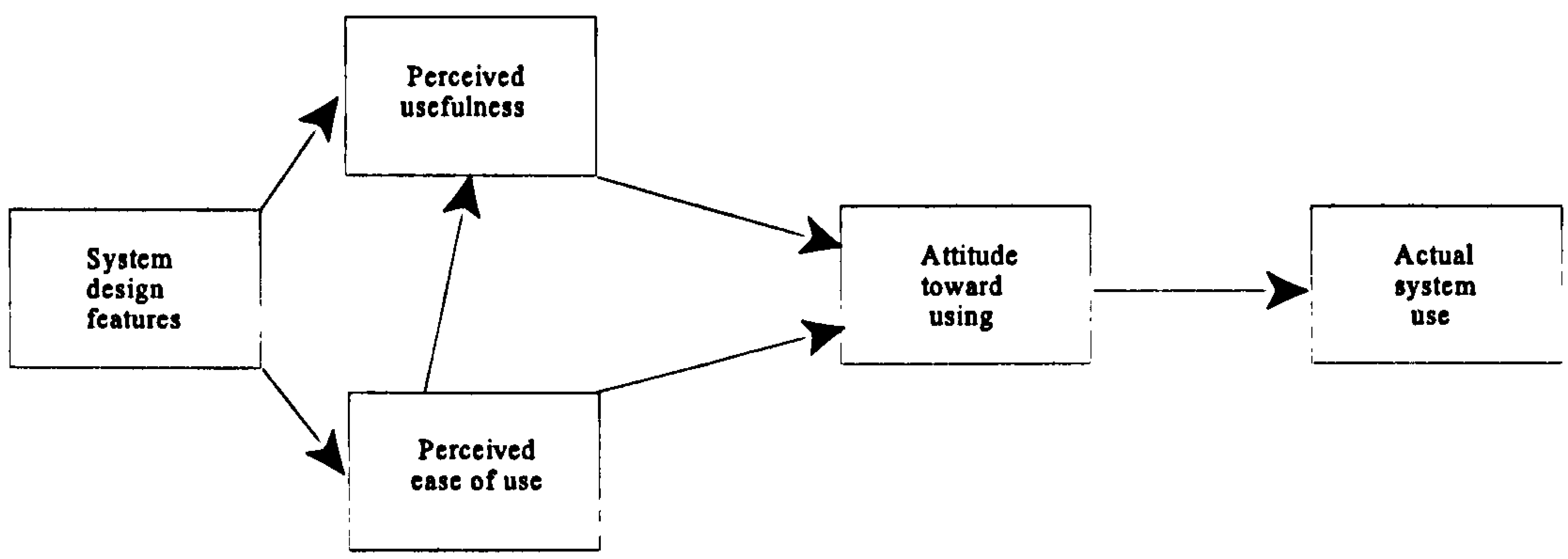


Figure 3.2

Users of an electronic mail system and of a text system in a large company were asked to complete a questionnaire on their perceived ease of use and perceived usefulness, attitude towards using and self reported usage of the systems. Use of both systems was voluntary. The results of the study suggest that perceived ease of use and perceived usefulness both influence attitude toward using and that the system characteristics influence both perceptions of ease of use and usefulness. However, whilst ease of use was found to influence perceived usefulness, it is worth noting that Davis found perceived usefulness to be more important than perceived ease of use in affecting how much a system is used, stating that "Users may be willing to tolerate a difficult interface in order to access functionality that helps them on their job, while no amount of ease of use can compensate for a system that doesn't do a useful task" (1993, p.484). Perceived usefulness was found to have a direct effect on usage as well as an effect via Attitude. Davis (1993) says that although it is unusual for a belief to directly affect a behaviour there is some support in the literature for this finding. System Characteristics were also found to have a small but significant influence on attitude. This, Davis suggests, could be because people enjoy using the system for its own sake without reference to ease of use or usefulness. Whilst this may be the case when people are experimenting with software, it is unlikely to be true in the work situation.

The above reported studies provide support for the inclusion of perceived ease of use and perceived usefulness amongst the perceptions which affect user satisfaction with software.

3.2.3.2 Perceived Involvement

The idea of a "user-centred" approach to system design and development, by which is often meant that the user should take part in the process, is a central theme of the HCI literature. In many cases such participation is reported to be beneficial in ensuring success with the system (e.g. Mumford, 1991; Baroudi, Olson & Ives, 1986). Although there have been suggestions that user dissatisfaction could arise as involvement raises their expectations of the system and they place too many demands on its development (Heinbokel et al.,1996) such studies are in the minority.

There is a need to bring together those who design the functionality of the technology and those who use that functionality (Grudin, 1992). This need has long been recognised, as has the difficulty it presents. Research into the mental model of the system held by the designer and user has attempted to bring the two together by trying to understand the two different perspectives (ref. Staggers & Norcio, 1993 for an overview). The model the user has of the system and his expectations of it may be influenced by his prior experience (Martinko et al.,1996), in which case it may be difficult for the user to modify that preconception. Similarly, the designer may have preconceptions from other work carried out and make assumptions about what is required. (Some nice examples of this are given by Gentner & Grudin (1996) where they give an example of a tractor designed to operate in a mode similar to that of cart and horse, the tractor being the horse and being directed by use of reins). The purpose behind involving the user is to try to ensure that the user and the designer gain a good understanding of each other's work domain and thereby are able to communicate on the same level, using the same language and various methods and techniques have been developed to this end:

Contextual Inquiry: CI is described as one of the tools in the PD toolbox (Muller, Wildman & White, 1993). It is based on the premise that "..users cannot describe what they really do because they are not conscious of it and do not reflect on it." (p.93). According to CI, it is the role of the designer to understand the work situation of the user and to place the needs of the user at the centre of the design process. It is, therefore, necessary for the designer to work with the user by taking an apprenticeship role and learning what is required by seeing it happen and asking questions when necessary to understand why things are done (Beyer and Holtzblatt, 1995).

Keil and Carmel (1995) investigated the success of custom built and package software in various organisations. Their findings were that designers very rarely get close to the real person who is actually performing the task for which they are designing. In most instances it is some form of surrogate, be that an analyst from the MIS department or a consultant, who takes on the role of user for communication with the developer. Their results show that projects often suffer from a lack of real knowledge on the part of the designers, of the task to be performed. Keil and Carmel express surprise at the high use of surrogates in communicating with the designers, but do not explore why they are used in preference to direct links. One reason could be that expressed by Holtzblatt and Beyer (1993) that the users themselves have difficulty in articulating what is needed and that it is, therefore, perceived to be more efficient to use an intermediary. It is this problem of direct user-designer contact and communication that Contextual Inquiry (CI) is designed to overcome.

However, whilst CI may help the software designer to gain a better insight into the needs of the people for whom he is designing, this does not necessarily help the user, whose part in the process is to describe the task by performing it. It is the designers who visualise the work as it currently is and then consider how the introduction of a system may impact on that. Whilst the user may be asked to evaluate the results of the designer's deliberations, he does not take an active part in them. Although the user plays a central role in CI, his involvement is at the bottom end of the involvement-participation continuum (i.e. the extent of involvement is at the level of that of information passing) as the role is a passive rather than an active one.

Participative Design (PD) has the active participation of users at its heart. However, the philosophy has grown out of the trade union movement to ensure that workers have some say in the environment in which they work. This means that the focus of PD is on more than the computer system and making that "user-friendly" or acceptable to the users. It is about co-determination between workers and managers on the development of the workplace and using technology as a tool to aid improvement. This means that both the organisation and the technology must be flexible and allow things to be done differently (Clement & Van den Besselaar, 1993). There is no single PD method. The approach to PD itself must be flexible to suit the needs of the moment and use may be made of prototyping, storyboarding, CI etc. but none of these alone will suffice (Muller et al., 1993).

PD aims for the full and active user participation, which gives people some control over all aspects of their work environment. It may be said to be at the top end of the involvement-participation continuum.

A more recent development in PD is the MUST[6] Method for Participatory Design (Kensing, Simonsen & Bødker, 1998). This focuses on the early stages of development and it stresses the need for management as well as user participation. It is stressed that the aims of the new system must be transparent, even if it may lead to job cuts. In the MUST method, it is essential that the vision of the system is understood and accepted by all levels of the organisation so that, if change to the work pattern is needed, that change can start to take effect whilst the system design is still in process, thus ensuring that all levels participate in the change. The method recognises the need for the IT professional to understand the needs of the user and the need of the user to understand the technological aspects. However, the formalisation required of a method means that the IT professionals are the key drivers to system design. There is also a steering committee who will deal with any political issues that may arise. This steering committee comprises management, and one or two user representatives. Although the user representatives should be elected, not selected, there is more of a controlling factor from the top level in the process than has been evident in other research into PD. Although Kensing et al. (1998) claim that MUST follows the

[6] MUST is a Danish acronym for theories of and methods for design activities

PD philosophy, it is closer to JAD and DSDM in its approach in that the user is involved in decisions regarding the system design, not whether there should be a computer system. This is perhaps in part due to the recognition by Kensing et al.(1998) that the recent move to downsizing has also reduced trade union power and participation in the high level decisions, embraced in the PD philosophy, often no longer takes place.

Joint Application Development (JAD) "involves" users by having user representatives attend workshop sessions from the outset of a project. These sessions are well-structured and have a defined plan of action. A facilitator ensures that knowledge and ideas are brought out for group discussion and that the workshop agenda is adhered to. Other roles, such as Scribe for noting down everything that is discussed, are also pre-defined and allocated so that everyone present is aware of what their contribution to the workshop is intended to be. The people who act as user representatives must themselves be users and should be the ones who are most knowledgeable about the area of work (Carmel, Whitaker & George,1993). The purpose of JAD is to speed up the design and implementation of systems without detriment to the end result. It is rarely possible to involve all potential users of a system in its design and adhere to a timescale, hence the need for representative users. These are not the same as the surrogates mentioned by Keil and Carmel (1995) but people who perform the tasks. However, although they are representative in that they are there in the place of the general user group, they are not representative in the sense of being average. The users selected are chosen because of their good knowledge of the work domain and of what is required. Even the fact that they are selected suggests a restriction on choice and decision making from the user community.

The **Dynamic Systems Development Method** (DSDM) is very similar to JAD but is Europe based whereas JAD is North American. Regular workshops are held with facilitators to help elicit and exchange knowledge; people have pre-defined roles and the project has pre-defined check points (time boxes) to ensure its timely completion. Ambassador users are chosen to champion the system and to ensure that the rest of the user community is kept informed of progress and encouraged to take an interest in developments. Advisor users are brought in for prototyping or test sessions (these users must be from the group that will be the key hands-on users of the new system). The Visionary user is there to ensure that the all-round business need of the system is not lost in the drive to introduce new technology.

Both JAD and DSDM stress the need for user involvement in the design process. The involvement proscribed for each is more active than that of conventional methods such as SSADM where user requirements are determined via interview. The workshop sessions enable users to suggest alternative approaches to the task rather than leave this to the designers as is the case with CI. However, the decision to introduce new technology and how it will be used is made as a business decision before referring to the users. The involvement of the users is at the point when it has been decided that new technology will be introduced. People are able to influence its design but within certain limitations. This involvement gives the user a greater role in decision making than does CI but does not extend that role into being able to influence the surrounding context as proscribed by the PD philosophy.

JAD and DSDM have an explicitly economic purpose to the method i.e. to provide a system on time and to budget. PD is more concerned with good worker relations. However the goal of all three in involving the user at the outset of a project, is to try to overcome the problem of a system being designed which meets a functional specification but which does not function because it does not meet the requirements of the user. In other words, it contains all the technical capabilities required of it but not in a way which the user finds accessible. Rather than the designer try to determine what the user needs, the user directly influences the design by being involved in the process of design. At the same time, it is clear that the "involvement" of CI is not the same as that of JAD and DSDM and that the "participation" of the PD philosophy is different again.

The terms "participation" and "involvement" have both been used in the literature to describe the inclusion of the user in the process of design and development. For example: "User participation in information system development is considered to be an important factor influencing implementation success or failure" (Wong & Tate, 1994, p.51). "..user involvement is a necessary condition for successful development of computer based information systems" (Ives & Olson, 1984, p.586). Given the differences just described in the process of involvement and participation it is worth considering whether there is any difference in meaning between user involvement and user participation or whether both mean essentially the same thing. This is discussed in the next subsection.

3.2.3.2.1 Involvement versus Participation

Baroudi, Olson and Ives (1986) found that user involvement has been used in several ways, from providing feedback to assigning users to the design team or setting up formal communication channels with users. However, in their view user involvement means the degree of influence that the user exercises over design decisions. If there is no influence, then any "involvement" is no more than symbolic. Mumford (1991) defines participation in much the same way saying that "Participation in systems design is 'the involvement of the end users, and others who will be affected by the system, in its design'". (p.271). The fact that this is not symbolic involvement is evidenced by the fact that she goes on to say that such participation means that design is shared between technical specialists and users, that it helps give a sense of ownership and helps effective use of the new system because the users have become familiar with it during its development. O'Neill (1998) also describes participation as meaning more than just being present when decisions are made, rather it means actively taking part and contributing to decisions. This view of participation is also taken up by Hornby and Clegg (1992), who describe three elements to participation:

- the type and level of interaction;
- the flow of information;
- the "nature and extent of the influence of one party on another over the decisions being taken" (p.294).

This usage of involvement and participation to mean essentially the same thing would seem to be disputed by Kuhn and Muller (1993) who describe participation as something more than "mere" involvement (p.28). Referring to Kuhn and Muller (1993), Damodaran (1996) too, seems to consider that participation is distinguished from involvement. Participation is described as being the active taking part in decisions which affect more than the design of a system but the surrounding work context. However, as Damodaran describes how to facilitate this participation, it becomes clear that the distinction being made is between approaches to user involvement and how far that involvement extends beyond the design and introduction of a new system, rather than a proposal that user involvement implies something distinct from user participation. The different forms of involvement, she says, will lie somewhere between it being merely informative, to actively influencing decisions relating to the whole system, not just design. In this sense the involvement of the user extends to any decisions regarding change of task or workplace structure which may arise as a result of the computer system. The user is an active participant in change not just an object of it.

For all the above authors it would seem that involvement and participation may be regarded as meaning the same thing. Where a distinction is made, it is to describe different points on the same continuum and involvement tends towards the low end of user reference, and participation tends towards the high end of input and decision making. However, it is not always clear where on that continuum they are for different authors.

Another view is taken by other authors who define participation and involvement as two distinct constructs (King & Lee, 1991; Kappelman & McLean, 1991; Barki & Hartwick,1994; Hartwick & Barki, 1994;). Barki & Hartwick (1994) define User Involvement as "a subjective psychological state reflecting the importance and personal relevance that a user attaches to a given system" (p.75) whereas User Participation means that the user takes part in an activity and has responsibility for it. Participation, they say, is behavioural. Involvement is psychological. The model they develop shows the influence of the two constructs of User Participation and User Involvement on system usage. (Figure 3.3 from Hartwick & Barki, 1994).

In the model, User Participation has three dimensions:
•Overall Responsibility
•User-IS Relationship
•Hands-On Activity
These three together encompass the behaviours and activities of users during the development process.

User Involvement is a single psychological construct and is the importance and personal relevance of the system to the user.

The model was tested by issuing a questionnaire both pre and post development of new applications to users in commercial organisations. Although the initial sample size was very large, just 105 responses were from the same individuals both pre and post implementation.

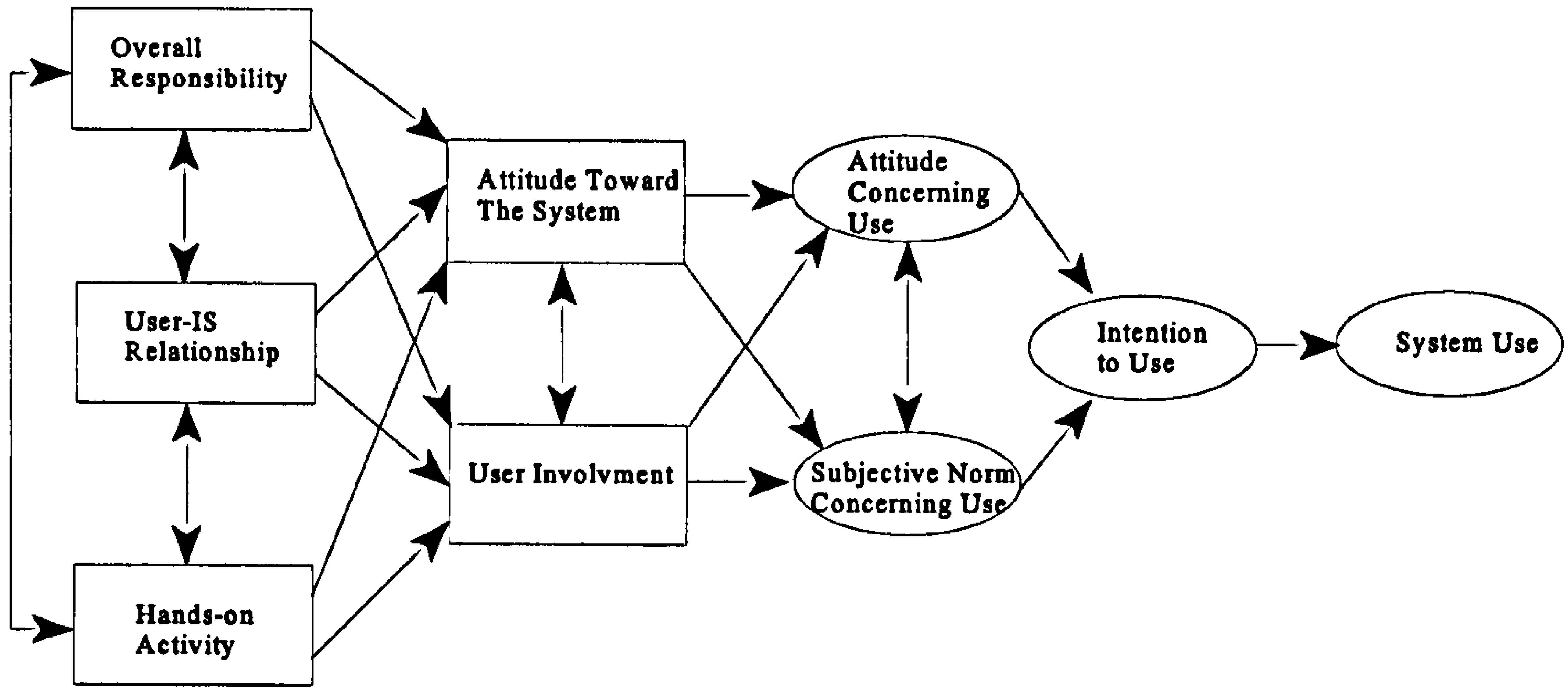


Figure 3.3

Overall Responsibility was found to be the dimension of User Participation which most influences post-implementation Attitude Toward the System and post-implementation User Involvement. Hartwick and Barki (1994) state that post-implementation User Involvement is more important than pre-implementation User Participation in explaining actual use. However, User Participation, particularly Overall Responsibility, will influence the belief of the user regarding the importance and personal relevance of the system (i.e. Involvement).

Participation as defined by Hartwick and Barki is neither the low level involvement of physical presence, nor the high level participation of active decision making. It goes further than participation as used by PD practitioners and allocates overall responsibility for the system to individuals. Whilst the elements of User-IS Relationship and of Hands-On Activity could be seen as the same as the user taking part in joint discussions and decision making with the IT professionals, Overall Responsibility implies that one person is in charge of the process. Indeed Overall Responsibility is described as reflecting overall responsibility for the success of the project. Hartwick and Barki argue that it is this sense of personal autonomy and control which leads to the increased sense of involvement. This is similar to the definition of participation by Wong and Tate (1994) who say it is a combination of the end-users of the proposed system being both involved in the design meetings and also having the need to ensure that the process worked successfully. Whilst Hartwick and Barki (1994) suggest that this sense of responsibility may be spread amongst individuals by assigning key project tasks to them, this would still seem to limit the possibility of a large number of people falling into the "participating" category.

Participation, then, as defined by Hartwick and Barki and shown in the results of their model, depends more on an individual having a stake in the success of a project than simply being given the opportunity to influence its direction.

Although not shown within the model, the same study also considered whether there was any difference between people for whom system use was mandatory and those for whom it was voluntary. It was found that the constructs of both User Participation and User Involvement were irrelevant for mandatory users of the system but are important in explaining usage for voluntary users. This is of assistance in considering the difference between participation and involvement as defined by Hartwick and Barki(1994) and in relation to their use by other authors: Whilst it is quite possible for people to be elected as user representatives, to be included in design meetings, to be asked their opinions and, therefore, to have all the outward signs of being participants in an activity, unless they feel that they have a responsibility for the system and a real influence over decisions, they are unlikely to feel it is important to them.

In the Hartwick and Barki model, the User Involvement construct is more important than Participation in explaining system use. That is, the feeling that the system is important and personally relevant will influence the individual's attitude towards the system. The results of their study show that, if someone has had responsibility for the system pre-implementation, they are more likely to feel that it is important to them when implemented.

The definition of User Involvement by Hartwick and Barki is quite different from the User Involvement construct used by Baroudi et al. (1986) as may be seen from the questions used to measure it. Baroudi et al. (1986) ask questions such as whether someone had developed a cost justification for a new system (more akin to the questions asked by Hartwick and Barki to measure Participation), whereas Hartwick and Barki use statements such as "I consider the new system to be of no concern to me" (1994, p.463). The definition of User Involvement as having a sense of importance and personal relevance is very similar to Davis' (1989) definition of perceived usefulness.

In the Technology Acceptance Model (TAM, Davis et al., 1989), Perceived Usefulness is defined as the extent to which the system is perceived as beneficial in the performance of a job. This may be that the individual perceives the system as a help in the performance of a particular task or that the use of the system will lead to promotion. That being the case, the system will be important to the individual.

Whilst the definition of Participation in the Hartwick and Barki (1994) model is not dissimilar to how the term has been used in other parts of the literature, it is more restrictive in its application in that it encompasses overall responsibility for the whole or part of the project. Participation and Involvement as used by Franz and Robey (1986), Mumford (1991), O'Neill (1998) and others means the inclusion of all who will be affected by the system in decisions regarding the design of that system. The definition of Involvement by Hartwick and Barki (1994) as the importance and personal relevance of the system is closer to the Perceived Usefulness construct of Davis (Davis 1989; Davis et al.,1989).

Doll and Torkzadeh (1989) consider User Involvement from a different angle by focusing on the differences between individuals and their desire to be involved. They are concerned specifically with user involvement in End User Computing (EUC) by which they mean the situation where PC based applications are developed either wholly or in part by the PC user. In this environment, they argue, Involvement should be defined as "the extent to which the user engages in systems analysis activities such as project definition and logical design decisions" (1989, p.1154). This idea of involvement means that it is the user who is responsible for deciding what is required and how that need should be met. It is not a case of actively taking part in discussions with others and "having a say" in what happens, but of being in control, possibly of the whole process. This is similar to the definition of User Participation by Hartwick and Barki (1994) as the user has both overall responsibility and hands-on activity.

From the above it would seem that the following definitions are possible:

- Both Involvement and Participation may be taken to refer to anything from physical presence at design meetings to active decision-making in all aspects of work related change.
- Participation means overall responsibility for the system design and implementation as well as hands-on activity.
- Involvement means overall responsibility for the system design as well as hands-on activity.
- Involvement means personal relevance and importance.

The conclusion that must be drawn is that there is no implicit meaning of either Involvement or

Participation that distinguishes one from the other and that they may be used interchangeably. Neither is there a single accepted definition of either Involvement or Participation.

In their investigation of the literature on the relationship between Involvement and Satisfaction, Ives and Olson (1984) pointed out the lack of consistency between authors in measuring either Involvement or Satisfaction and hence the often contradictory results that were reported. It would seem that little has changed. Whilst there is agreement that people must be involved at all stages of the process, the nature of that involvement remains imprecise.

3.2.3.3 Involvement and Perceived Involvement

Perhaps, however, the "objective" amount of involvement i.e. that which may be measured by the number of involvement situations (e.g. brain-storming sessions, discussion forums etc.) and the amount of input by the user in those situations, is not as important as the degree to which the user feels involved. Doll and Torkzadeh (1989) touch upon this when they suggest that the difference between the individual's desired level of involvement and his perceived level of involvement will affect his level of satisfaction. It is, after all, the user who is important and his perception of having been involved is what matters.

Despite the variations in the definition of involvement, the evidence in the literature suggests that the level of involvement i.e. the amount of actual physical involvement, and the resulting perception of involvement i.e. how content the individual feels with his involvement, in the design and development process will influence satisfaction with the resulting system. Perceived Involvement, defined as the degree to which the user feels involved in the process, is therefore included as one of the factors affecting user satisfaction.

3.2.4 Organisational Characteristics

The context in which software is used can be as important to its success as the software itself (Bjørn-Andersen, 1988; Eason, 1989; Damodaran, 1996). Organisational issues include how the system may affect the way in which people interact with each other; how it may affect the user roles within the organisation; and how it may affect their relationship with management. The need to be aware of the surrounding organisational environment is a central issue of both PD (Damodaran, 1996) and of the socio-technical approach of Mumford (1993) which is to obtain a balance between the social and the technical aspects of any change in the work environment.

Clement (1994) describes how the effect of a strict hierarchical structure within an organisation and the introduction of computer systems which both forced a way of working and monitored workers' performance, led to a walk out by staff who believed they were being physically harmed by the equipment they were using. In fact, the computer equipment itself was quite safe, but the psychological effect of people feeling imprisoned in their work environment was that people suffered physical distress. Eason (1989) describes how systems often fail because the way in which they force tasks to be performed does not match the distribution of roles and responsibilities in an organisation. Sometimes it may be preferable to re-evaluate the current way of working and to improve it with the introduction of a computerised system. However, this of itself may involve a change within the organisation and the importance to the users of this change in work pattern should not be underestimated as it may be as important to them as the change of technology (Eason & Harker,1991). Hornby and Clegg (1992) in their study into the effect of user participation in the introduction of new computer systems in a UK bank, state that "organizational design can have a powerful influence on the way in which key actors perceive and interpret their own and other's roles" (p.294). Finally, Tomlin (1991) describes how the effective delivery of information technology and its contribution to the success of business is dependent on the corporate culture in place.

All of the studies so far mentioned under System and Individual Characteristics also make some reference to the influence of organisational factors on the degree of acceptance of a system. Ramamurthy et al. (1992) deliberately use students in order to remove the organisational context and to concentrate on the characteristics of the individual. Martinko et al. (1996) demonstrate that the attitude of the individual may be influenced by that of co-workers and of management. Hirschheim and Newman (1988) highlight how the attitude of a manager towards a system can send a negative or positive signal to the staff. Markus (1983) argues that the power shift brought about by a new system had a greater impact on user acceptance of the system than factors concerning the system itself.

An organisation is made up of individuals who may work on their own or as part of a group. Groups may consist of individuals in a single location or may be a group of individuals remote from each other but working co-operatively. In the sense that everyone in an organisation is working to the same goal, the organisation may be regarded as a single group of co-workers. However, within this there will be different levels of authority and responsibility. Access to and control of information is often seen as a way of gaining or retaining power. The use of a computer system to make information more accessible or, alternatively, to control the way it is keyed in, may vary according to the type of organisation and the attitude of management toward the system. It may then be viewed as an enabling or controlling technology by the people who use it.

The many aspects which combine to form the characteristics of an organisation can loosely be grouped as

- Organisational Type
- Organisational Politics
- Management and Co-Worker Attitude

These are each discussed below.

3.2.4.1 Organisational Type

By "type" is meant the nature of the organisation - whether it is service-centred or profit oriented; whether it is conservative (i.e. bureaucratic) in outlook or innovative; whether management control is from the centre or devolved. Government departments can be loosely described as service-centred since their existence does not depend on the ability to make a profit and Private Sector organisations can be described as profit oriented as their existence depends on making a profit. There is also a general perception that service centred organisations are more likely to be conservative in outlook than are profit oriented organisations which are continuously looking for new ways to improve profits. Central control or devolved responsibility are found in both types of organisation.

Rice and Tyler (1995) hypothesised that the introduction of a new system, in the form of voice mail (v-mail), would be less successful in a more conservative organisation than in an innovative one (innovative is defined as being open to new ideas and methods). Their findings were, in fact, that organisational conservativeness had a significant positive relationship with the use of v-mail. They conclude from this that voice mail is not perceived as innovative and that it was the non-innovative i.e. cost cutting potential of voice mail that influenced its use. However, they do also mention that the result lends weight to the idea that more bureaucratic organisations are more successful in ensuring the use of new systems. This latter suggestion implies that use of v-mail may not have been entirely voluntary in which case its "success" (measured in terms of usage) may be success in the view of the organisation as a whole and not of the individual.

The findings of Rice and Tyler (1995) contrast with the results of the study by Caldwell and Uang (1995) into the satisfaction with the use of voice mail across several State Government agencies. The hypothesis relating to the use of voice mail by different organisations is that "Differences between organizations in voice mail system and appropriateness ratings exist because of differences in task demands and organisational contexts" (p.309). Information was gathered by use of a five section survey collecting data on the individual; context of voice mail use; general comments; system feature priorities; and the overall satisfaction with the system. General comments were requested as well as answers to specific questions. Overall satisfaction was measured by use of the Questionnaire for User Interaction Satisfaction (QUIS). The results are different from each organisation. All the organisations are government based and should, therefore, fall into the category of service-centred and conservative. Following the argument that this type of organisation is generally successful in ensuring the usage of a new system, it could be expected that usage should be similar in each agency. However, this was not the case.

Usability (measured by QUIS and relating to questions such as ease of use) was also rated differently by each organisation. The "Labor" organisation gave a lower satisfaction rating overall than all the others and was also lower on the scales which measured appropriateness of v-mail. The "Resource" organisation gave a higher rating for v-mail appropriateness which Caldwell and Uang attribute to the fact that the employees tend to be more dispersed than in the other organisations.

It is possible that the assumed structured discipline of the organisations taking part in the study did not influence the outcome because use of voice mail was at the pilot stage. This suggests the possibility that formal procedures for its use had not yet been laid down. Once they have been, usage may increase. It is also possible that the type of work performed by the organisation is more influential over people's view of the tool provided than the rules and regulations of the organisation itself. However, perhaps most importantly, the way in which acceptance is measured is also different from that used by Rice and Tyler (1995). Caldwell and Uang (1995) issued a "satisfaction" questionnaire to individual users, whereas Rice and Tyler based acceptance on extent of use within the organisation. Whilst Caldwell and Uang do measure extent of usage across gender, it is not used as a measure of acceptance.

Caldwell and Uang also indicate that usage is variable across the organisations. Ackerman and Starr (1996) found that the use of groupware software would increase or decrease depending on how much people felt it was being used. It is possible that the use of v-mail was subject to the same influences and that people would respond to messages using the same mode in which they had received them, regardless of how they felt about using the technology themselves. This would mean that usage could be affected by factors external to the organisation itself.

Insufficient is known about each of the five organisations studied by Caldwell and Uang (1995) for anything but speculation to be proffered about why they vary so much in their perception of the v-mail system. However, it is interesting to note that the same v-mail software is being used in each at much the same stage of implementation. Despite this, the view of how usable it is and the overall satisfaction rating is different between the organisations. The ratings are different between organisations but, apparently, consistent within them. This very much suggests that factors other than the software itself are influencing people's perception of it.

Hornby and Clegg (1992) consider the effect of the organisation on the participation of people in the introduction of new systems. They distinguish between two types of organisational structure:

- The hierarchical structure where everyone has a clear view of their role and what it involves and work is governed by routines and procedures;
- The flat structure, where people are not in fixed roles but are expected to make decisions within their area of capability and act upon them to the benefit of all.

The first type of organisation they refer to as "deterministic" and the second type as "organic". It is their proposition that the type of organisation will influence the form that user participation takes and that the success or otherwise of this might then affect how users view the benefit of participation in the future. The study takes place in a large UK bank which is introducing a computerised system. The organisation is modelled along the lines of the Civil Service and so is described as deterministic. The new system will cause a change of working practices in some areas and so represents new ground for the organisation. In order to introduce the system, various groups and sub groups with different responsibilities are created. The intention is to involve users in the process.

However, the results of the questionnaires and interviews show that most of the users felt either that they had not been asked for their views or that their views had not been included in the system design. Neither did the majority feel that they had had any influence over the process. Overall, there was dissatisfaction on the part of the users with the process. Hornby and Clegg suggest that this dissatisfaction is, at least in part, due to the fact that change, with unknown consequences, is occurring in an organisation which is used to being able to follow rules and procedures. The result of this is that people do not know how to pass on information or what the decision chain is. Suggestions are made for how changes in the organisation could have been made prior to introducing the system in order to ease the communication problems.

There is no report of how well the system was accepted some months after it had been introduced and the concentration is on the perception of people of the amount of involvement they had and how happy they were with it. Nevertheless, this study offers support for the argument that the type of organisation may have an influence on the perceptions of the people who work within it and how they react to change.

From these three studies it may be concluded that Organisational Type does have an influence on the people within the organisation and how they view the software tools they are using, although there is no clear evidence of whether that influence is positive or negative.

3.2.4.2 Organisational Politics

"...Information systems development is political as well as, sometimes far more so than, technical in nature" (Keen, 1981, p.31). How the decision-making process is managed within an organisation will affect the power that different groups have to influence their own work and that of other people. Whilst it is similar to organisational type in that the type of organisation may influence the decision making structure, organisational politics is a reflection of how the different groups within the structure interact with each other. If a group perceives a new system as being a way of increasing their power or influence within the organisation they are likely to have a favourable opinion of it. Alternatively, if they see it as a vehicle for use by others and possibly reducing their own autonomy, they are likely to have an unfavourable attitude towards it (Shneiderman,1982). This effect of group perceptions on systems development is demonstrated in a role playing exercise described by Klein and Hirschheim (1983) in which a workshop group assumed the roles of different sections of an organisation. The conflicting interests and priorities of the different groups quickly become apparent.

Markus (1983) demonstrates the difficulties in introducing a centralised system into a decentralised organisation. The aim behind the introduction of the software was to enable global reporting from the offices in different locations. Prior to the introduction of the new system, the accountants at the divisional offices were fairly autonomous and determined their own way of reporting. The purpose of the new system was to locate all data centrally rather than for it to be sent by the divisional offices and collated. This had the effect of making information immediately available to the corporate accountants rather than relying on the divisional accountants to provide it. It also meant that the divisional offices had to conform to a single standard format of data recording. There was resistance to the new system by the divisional accountants who complained that the system did not help them and was not an improvement on what they had used to do. The corporate accountants, on the other hand, felt that the system was successful and had several benefits for them. Whilst the way of working was different between the corporate group who were mainly financial accountants and the divisional groups who were management accountants, Markus concludes that the effect of the introduction of the system was a power shift from the divisional accountants to the corporate accountants and that it was this, rather than problems with technical aspects of the system which caused the divisional accountants to reject it. Even after technical improvements had been made, the system was still disliked at divisional level. Even if it was not a deliberate attempt on the part of the corporate accountants to gain power, the divisional accountants perceived a loss of power as a result of the system introduction.

Even after a system has been introduced, its further development might still be used by one group to gain power or authority in the organisation (Kling & Iacono,1984). Their study demonstrates how the manufacturing division of an organisation manages to maintain and increase control over the Data Processing function by being the key user of that area. The Marketing group, who did not feel the system to be of benefit, were also reported to not properly understand its purpose and not to belong to the "social world" of the core users. Despite continuing problems with the system it was generally reported to be beneficial by those who supported the power of the manufacturing division rather than that of the marketing division. It should perhaps also be noted that there was strong management support within the manufacturing division itself for the use of the system and that the users were generally happy with it.

As well as demonstrating the political aspect of organisational culture, this study perhaps also indicates the power of management support for a system from the top to influence its acceptance by staff.

Franz and Robey (1986) consider organisational factors and system success by considering how the organisational factors (age, size and decentralisation) affect the MIS department responsible for introducing new systems. The structure of the MIS department arising from that of the organisation, may influence the involvement of users and thence their perceived usefulness of the system. A larger organisation is perhaps more likely to have a better resourced MIS department and, therefore, be better able to involve people in the process of introducing new systems. However, their findings are that there is no significant direct relationship between organisational factors and user involvement. What they do find is that the higher the MIS manager is in the organisation, the more likely it is that users will have a sense of involvement and perceive the system to be useful. Whilst Franz and Robey suggest that this relationship could be due to the fact that the MIS manager was likely to be at a higher level in larger and older organisations, which therefore had more resources and the ability to involve users, they do also point to contradictory findings which suggest that larger MIS departments had a lower success rate because they are more bureaucratic. However, if the findings are considered in the light of the research by Saunders and Scamell (1986), it is the power and influence, not the size, of the MIS department within the organisation that matters. It is the politics within an organisation which determine which departments hold most power. It is, therefore, the very fact that the MIS manager is in a powerful position, perhaps at board level, that raises the profile of new systems introduction and engenders a more positive response towards them.

These examples lead to the inclusion of Organisational Politics as a factor affecting user satisfaction.

3.2.4.3 Management and Co-worker Attitude

Participative Design (PD) projects are generally held to be worker rather than management driven (Carmel et al.,1993), yet the need for management support in order to gain acceptance is still recognised (Clement & Van den Besselaar,1993). The MUST Method (Kensing et al., 1998) has the stated purpose of ensuring involvement at management level. Damodaran (1996) stresses the importance of leadership by top and middle management if IT projects are to be successful.

The AMRIT model (Martinko et al.,1996) shows management support as an external factor influencing the attributions of the individual, proposing that such support will positively affect the acceptance of an IT system. (Caution is, however, given about the form this "support" takes and an autocratic approach, dictating the use of a system is more likely to produce a negative than a positive effect).

Further evidence of the need of support not just at line management but at senior management level is provided by Damodaran (1991). The initial impetus given to the work in the DHSS by a group of Human Factors (HF) consultants by the appointment of a director who recognised the need for the work, was lost as the people who were originally co-owners of the project moved to other jobs. Whilst Damodaran (1991) is arguing the case for an HF strategy when she says that people at Board and Senior Management level must both understand and endorse the need, the argument endorses the need for such support for any implementation if it is to succeed.

This argument is echoed by Poltrock and Grudin (1994) and Axtell, Clegg and Waterson (1996). The former describe the development problems for software suppliers. The main concern of the study is how Suppliers can properly incorporate design standards in their products to the benefit of the user. However, the two case studies described are good examples of how a determined, knowledgeable and positive leadership, backed by top level management can lead to product success, as can good cross-functional communication. Axtell et al. (1996) are also concerned with software development but this time within an organisation-user not supplier. In their study, the aim is to investigate a method for ensuring user participation in the development of a system. One of the problems found initially was the recruitment of users to participate. It is suggested that this was due to a lack of senior management support for the method. If the task is not perceived to be of importance at senior management level, people will not be encouraged to volunteer to take part nor allowed the time to participate.

The effects of different management attitudes is also discussed by Hirschheim and Newman (1988) in their case study in an Underwriting company. Lack of management support is one of the nine possible causes of resistance they cite. They give an example of a manager who claimed to have adopted a laissez-faire approach allowing the underwriters to use the system or not as they wished on the basis that it would gradually gain acceptance. At another branch the manager had a coercive approach and ensured that the system was used by removing the paper documentation meaning that reference had to be made to the system. It is evident from the interview with the underwriter at the first branch that the result of the laissez-faire approach was to give the impression that the manager had no interest in the system and was giving no positive leadership towards its use. Whilst there is no note of an interview with an underwriter at the second branch, the system was not accepted there either and the conclusion may be drawn that forcing the system to be used did not help improve the users' view of it. Positive leadership in use of the system was lacking at both sites.

Cheney (1984) found that the attitude of management in the way that tasks and procedures are specified and the support and trust given to staff, affected the amount of job satisfaction reported. Although this study was concerned specifically with the job satisfaction of computer programmers, it is intuitively likely that the same effects would be found with other workers.

However, it is not just the management support or lack of it for a system that affects how people perceive it. The attitude of others at the same level may also have an effect. Martinko et al. (1996) posit that if respected co-workers blame the computer system for their inability to use it, so the individual is more likely to blame any problems he encounters on the system rather than on his own skill. Ackermann and Starr (1996) base their argument for social activity indicators for groupware systems on the argument from social psychology that people will react to a situation largely in accordance with how other people react. In other words, if people perceive that a system is being used, they will be encouraged to use it themselves. Thus if managers are perceived not to have a favourable attitude towards a system, staff are less likely to view it favourably and this view will be re-enforced by seeing that others are reacting in a similar unfavourable manner.

The attitude of management and co-workers may well be influenced by the type of organisation in which they work. However, there is little in the literature to suggest that a particular organisational type will always result in a particular managerial or organisational attitude. The combination of management and co-worker attitude is, therefore placed as a separate factor of organisational characteristics.

3.3 Summary and Discussion

The key influences on user satisfaction as they have just been discussed are shown in table 3.2.

System Characteristics	**Individual Characteristics**	**Individual Perceptions**	**Organisational Characteristics**
Functionality	Domain Expertise	Perceived Usefulness	Organisational Type
Match to the Task	System Experience	Perceived Ease of Use	Organisational Politics
Measurable Usability	Attributional Style	Perceived Involvement	Management & Co-worker Attitude
	Level of Education		
	Prior Experience		

Table 3.2

Although each of the four key areas of influence and their sub elements have been described individually, it is evident that it is the combination of factors that affects user satisfaction. It is also clear that some factors are affected by others and that there is an overlap between the four areas described.

There are several problems with the data as it has currently been presented, from the point of view of someone who has the task of purchasing a COTS system for use within a small to medium sized organisation.

- There are too many factors for information regarding each to be gathered and analysed in a reasonable time-scale.

- It is not clear whether any one factor is more important than another in affecting overall satisfaction.

- It is not clear whether a particular combination of factors has an important influence.

The current design proposals of the Common Industry Format (CIF) as described by Bevan (1999a) seek to address this, at least in part, by making it the responsibility of the software supplier to provide a report on the usability of the software as defined under ISO 9241/11. The report should detail:

a) the description of the product

b) the objectives of the test(s) performed

c) the format of the test and the results including measures with regard to time taken to complete the task, amount of assistance required etc..

d) the number and key characteristics of the participants

e) the context in which it was tested

Details of the items from a) to c) would assist the prospective purchaser in determining whether the System Characteristics met the requirements of the organisation.

With regard to d), it would be a complex and delicate process to compare the key characteristics of the people who took part in the test with those of the potential users of the system. Listed under key user characteristics are such things as age, level of experience and education. If the result of a comparison were that the potential user group was not very like the test group, what impact would this be likely to have?

Finally, the context of use of the software during testing is necessarily restricted to describing the physical environment. Whilst this may well be an important factor - if someone is asked to concentrate in a noisy environment it may affect their performance - it would not be a reflection on the characteristics of the software so much as on the ability of an individual to concentrate under difficult conditions. The context of use, in the sense of type of organisation or management and co-worker attitude, is not something which can be easily measured in the real world or replicated in a test environment but which may influence the overall satisfaction with a software package.

In summary, whilst CIF reports may assist in raising the profile of usability in the industry and their availability may assist purchasers in comparing one software product against another, there is still the problem of assessing user engagement with a product in advance of its introduction.

The next chapter addresses this problem by reconsidering each of the factors and proposing a model of those which are key to user engagement.

Chapter 4

Developing a Model of User Engagement

The previous chapter discussed the factors in the literature which have been shown to influence user satisfaction with software. User satisfaction is defined as a favourable attitude towards the software as might be demonstrated by user acceptance and usage of it when usage is voluntary. Four areas of influence were derived from the literature:

- System Characteristics
- Individual Characteristics
- Individual Perceptions
- Organisational Characteristics.

However, much of the research into user satisfaction with software has concerned itself mainly with software which is designed and developed in-house. Most small to medium sized organisations do not develop in-house and purchase COTS software to meet their needs. Of its nature, such software must be of a general fit design. Account must also be taken of the fact that such organisations do not have the resources to investigate and analyse all the potential factors which may influence user satisfaction. Whilst it is evident from the literature that there is a complex web of factors which lead to satisfaction, the aim of this research is to develop a simple model of those factors which influence user engagement with software prior to its introduction. Positive user engagement is posited to increase the likelihood of satisfaction with actual use. In order to be useful to a small organisation the model must be simple but still capture the key elements.

This chapter examines each of the areas above and their constituent elements more closely with a view to simplifying the number of factors whilst retaining those which are key to user engagement with the software. Hypotheses are proposed and a model of User Engagement is developed. Section 4.6 then describes the research method adopted in order to test the proposed model.

4.1 System Characteristics

The previous chapter suggested that there are three elements to System Characteristics:

- Functionality
- Match to the Task
- Usability.

4.1.1 Functionality

Although "functionality" is sometimes used to describe all the characteristics of a system, it is used here in the very limited sense of meaning the tasks it enables to be completed. For example, people who wish to purchase a package in order to carry out an accounting function will generally all list the same functions that the software must be able to perform. Depending on the size and spread of the business, some may also require multi-currency conversion. The presence or absence of that function would help them to narrow down suitable packages. The functionality of a system is essential if it is to be a useful tool. Functionality is, therefore, retained as a key element of system characteristics.

4.1.2 Match to the Task

Although different organisations may require essentially the same functionality from their software, there are as many different ways of performing a task as there are organisations. Not only must a software package be capable of performing certain functions but it must be possible for those functions to be performed in a way that fits the organisational processes and procedures. A mismatch between the operation of the software and the operation of the organisation will lead to dissatisfaction (Eason, 1989).

Determining whether a software package is suitable for a task is not a simple matter. Although it may at first appear that the mode of performing a task may be described and the actions required by the software be checked against this, performing a task is more than the simple operational procedure. Different parts of a task may be undertaken by different people depending on the structure of the organisation and the software would need to allow for this. In addition, as previously described, the introduction of a computerised system may itself effect a change to working practice (Rasmussen, 1992). This raises the question of whether it is indeed possible to assess the fit of a software package to the purpose for which it is being used in a particular organisation.

Goodhue (1995) suggests that it is possible to measure how well a system matches the task and that this may be achieved via user evaluations of the system. It is his contention that people will give higher evaluations of a system depending on how well it meets their individual capabilities and on how well it matches the task(s) they need to perform. The characteristics of the technology, of the task and of the individual and the interaction between the three, will all influence the individual's evaluation of the technology. The Task Technology Fit (TTF) model (figure 4.1) describes how user evaluations of a system may be used as an assessment of suitability.

Each of the factors in the model is described and discussed below.

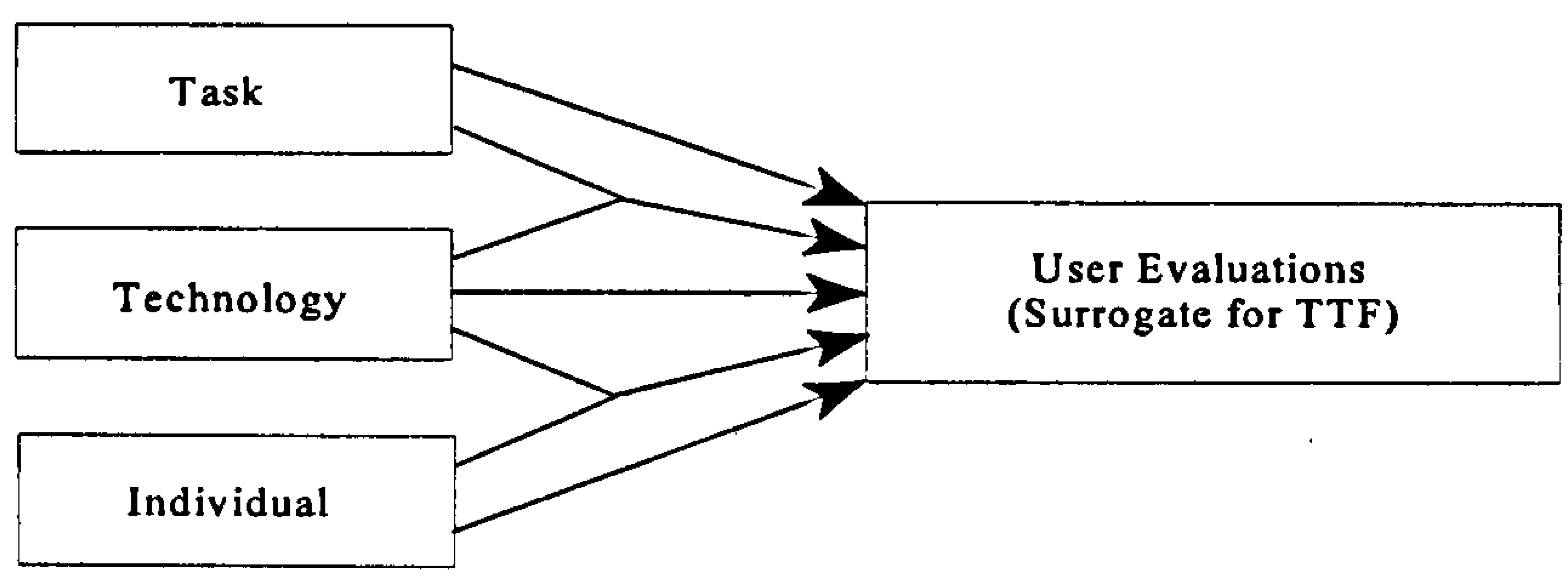


Figure 4.1

• Technology characteristics

Goodhue uses four dimensions to assess the technology. The assessment is not carried out by the people using the systems but by personnel from the IS department. The four dimensions are:

Integrated systems

Number of workstations available

Number of people available to give assistance

Location of those assistants (i.e. near to the users or centralised).

However, these areas give an overview of the structure of the technology in an organisation and how that structure is supported. They do not reflect the actual technology, in terms of software and hardware, in use for a specific task. Nor does the task specified reflect the user's interaction with the computer system. The task is the use of quantitive information in managerial posts. In other words the task is to obtain information by use of the computer system. However, of the four items listed, only the first - whether or not different systems have been integrated - would affect the user interaction with the technology in order to obtain information. The availability or lack of workstations would affect the possibility of interaction but not the interaction itself. The number and locality of assistants might have an effect on being able to obtain the information if the user has difficulty with the interaction but would not explain why someone found the system difficult to use to begin with. Whilst the task of using quantitive data is specific, the technology itself is only specific in that it fits a type i.e. that of consisting of integrated systems rather than being composed of several separate systems. The characteristics of the different software that combine to create the integrated system are not mentioned. It is the structure of the technology that is important in the model not the individual components.

•Task characteristics

The characteristics of the task to be performed are divided into 3 areas. These are:

Routine or non-routine

Interdependence

Hands-on

The latter does not so much refer to the task itself as to how it is performed. Its inclusion as a dimension is in order to distinguish between people who use reports from the system and those who extract data directly. Those people who have "hands-on" use of the system are the ones who use third and fourth generation languages to access the data and so have direct interaction with the technology. Their relationship with the technology is, therefore, different from those who receive system outputs. Goodhue (1995) suggests that this direct interaction will lead to higher demands of the technology and, thereby, greater likelihood of low evaluations. This does not necessarily follow. It is not the direct interaction per se which is likely to lead to a higher expectation and lower evaluation. It is the need for greater flexibility in obtaining data that leads to hands-on interaction and the evaluation of that interaction will be determined by the perception of how easy it is to obtain the required data.

Whether the tasks are routine or not, or are interdependent will affect the nature of the data required and how it is obtained. According to Goodhue, people are more likely to be able to work around difficulties with the system if they are performing routine tasks. They are also less likely to experience difficulties if the data is always obtained from the same source and they do not need to identify links between different sources. Both of these again relate more to how the structure of the technology i.e. whether different systems are integrated, matches the structure of the tasks i.e. routine or ad hoc, rather than to how specific software products facilitate specific tasks. For example, some financial reporting may require data from several different systems (stores, purchasing and sales) each of which may be operated independently of the other.

•Individual characteristics

The skills and abilities of the individual are defined as how computer literate those individuals are. The argument is that the more familiar people are with the system, the better trained or more competent, the less likely they are to experience difficulty and the more likely they are to give higher evaluations.

Whilst this might intuitively seem to be the case and Goodhue quotes support from the literature, the contrary has also been found. People who are more computer literate may also have higher expectations of a system and be disappointed, thus giving lower evaluations (Rice & Tyler,1995). The diverse findings in the literature with regard to these user characteristics would suggest that, whilst the characteristics do influence the perceptions of the user with regard to software evaluations, the influence is not necessarily always the same way, suggesting that other factors will also play a part. Goodhue defines computer literacy as being comfortable with the use of spreadsheets, creating reports or writing simple applications. It would be possible for an individual to fulfil these criteria and yet not have expertise in the particular area for which they need to obtain management information i.e. they would be lacking the domain expertise which might positively influence their evaluation of the system (Ramamurthy et al.,1992). The results reported by Goodhue reflect the difficulty in ascertaining what the effect is of computer literacy on user evaluations.

•Interaction between the characteristics

Goodhue argues that how a user evaluates a system will depend on how important a characteristic of the system is according to the task the user has to perform and how capable the user is of performing the task with that system. The example he gives is of how someone needing to access data from a variety of systems should find that simpler to achieve if the systems are integrated rather than if different data is held by different individual systems.

This example describes a type of task i.e. that of extracting data produced by a variety of systems and a type of technology i.e. an integrated structure. It does not describe the task of extracting data itself nor the steps by which the system enables this to be achieved. This compares with the need for "appropriate" technology as described by Clegg (1988). Whilst it is necessary to have an appropriate tool for the job, what is being described by Goodhue is the appropriate structure for use of a tool rather than the appropriateness of the tool itself. However, the means by which the appropriateness or fit of the technology to the task is measured by Goodhue is worthy of consideration. The type of technology (integrated or not) may be relatively easily determined since it is a question of fact rather than opinion. This is much the same as the inherent functionality of a system. Similarly, the type of task is a matter of description not of opinion. But the fit between the task and the technology relies more on the perception of the person performing the task as reflected in the type of statement used by Goodhue e.g. "It is easy to learn how to use the computer systems that give me access to data" (p.1841).

In describing the TTF model, Goodhue is describing a type of task and a type of technology. Both these may be objectively stated. However, the fit between the two is a subjective evaluation. That evaluation may be influenced by the characteristics of the user e.g. prior experience and computer literacy but it is the *perception* of task technology fit that matters. Even at the detailed level of how a single software package matches the task for which it is to be used, the people who use it will have a *perception* of how well it fits their task needs. This leads to the conclusion that Match to the Task, is not something that a software package inherently possesses. It is perceived to be present or absent depending on its context of use. As such, for the purpose of the model being developed it should be placed under the Individual Perception group of factors.

However, the question then arises as to whether perceived match to the task is a factor distinct from the perception of usefulness. Although it may be argued that a system may match a task as it is currently performed but to be useful it must help to improve the performance of that task, such an argument assumes that "useful" contains within it the idea of improvement. Whilst this may be the case in some instances, it is not so in all cases. Whereas, if a system is not perceived as suitable for performing a task, it will not be perceived to be useful. Match to the Task is, therefore, part of usefulness. Although Davis (1993) does suggest some evidence for people believing computer systems to be useful simply because they enjoy using them, it is hard to believe that this would be possible in the work context. Match to the Task will, therefore, be incorporated into the Perceived Usefulness construct in the Individual Perception group.

4.1.3 Usability

This is defined as the objectively measurable aspects of ease of use, learnability and efficiency. An organisation may wish to asses a system based on objective usability only, in order to ensure greater efficiency (Davis, 1989), and perceptions of usability may not be considered of prime importance to an organisation (Morris & Dillon, 1996). Such would seem to be the case in the study reported by Gray, John and Atwood (1993).

Their investigation into the benefits of a replacement system in a telephone company, found that the proposed system, which they thought "..a good example of a graphical user interface whose designers paid attention to human-computer interaction issues" (p.243) would result in a high cost to the telephone company as transactions would take longer. There is no report on how satisfactory the users felt either the existing or the proposed system to be, but it is evident that the economic needs of the organisation were of first importance. The fact that the existing system was measurably easy and efficient to use is what counted. Morris and Dillon (1996) also suggest that subjective usability (in this case meaning simply the user's liking for a system) may not be of prime importance to an organisation.

Although Nielsen and Levy (1994) find a correlation between objective usability and subjective usability in that people are more likely to be satisfied with software which helps them to perform better, generally this satisfaction will depend on whether people actually perceive that it is helping them to perform better (Davis, 1989) i.e. on subjective usability. It has also been found that people may prefer systems which are not more efficient from the point of view of the organisation (Davis, 1989; Nielsen & Levy, 1994).

Whilst it may be said that a system must possess the objectively measurable characteristics of functionality and of usability in order to lead to user satisfaction, the presence of these characteristics will not necessarily lead to that satisfaction. The context in which the system is to be used will affect the user perception of its characteristics. In fact, as Bevan and Macleod

(1994) say "The characteristics of the context...may be as important in determining usability as the characteristics of the product itself" (p.135). (Usability is used here to mean the subjective measurements of ease of use, learnability and efficiency).

This leads to the amendment of table 4.1 below to show Functionality and the objectively measurable aspects of Usability under System Characteristics. It has been argued that Match to the Task is not objectively measurable and forms part of the perception of usefulness. The characteristics of a system are, then, proposed to consist of measurable items. This might suggest that it would be possible to state a score for these items which, if achieved, would lead to satisfaction with the system. If this were the case, the research effort should focus solely on the characteristics of the system. However, it is argued that, even if a system scores well on objective measures, that will not necessarily lead to user satisfaction. This leads to the proposal of the hypothesis (H1) that "*the characteristics of a software product alone are not sufficient to ensure user satisfaction.*"

System Characteristics	**Individual Characteristics**	**Individual Perceptions**	**Organisational Characteristics**
Functionality	Domain Expertise	Perceived Usefulness	Organisational Type
Measurable Usability	System Experience	Perceived Ease of Use	Organisational Politics
	Attributional Style	Perceived Involvement	Management Attitude
	Level of Education		Co-worker Attitude
	Prior Experience		

Table 4.1

4.2 Individual Characteristics

Five factors are listed under this heading for which there is evidence in the literature of influence on user satisfaction with software. Although Goodhue (1995) restricts the user characteristics in the TTF model to just two i.e. their skills and abilities (by which he means their hands-on experience of third and fourth generation languages), he acknowledges that other factors may play a part in user satisfaction. Bevan (1999a) lists three areas of user characteristics - Personal details, Skills and Knowledge and Personal Attributes. These, he says, form part of the context in which software will be used and should, therefore, be taken account of. The individual elements listed under each of these loosely correspond to the five factors of Attributional Style, Prior Experience, System Experience, Level of Education, and Domain Expertise discussed in the previous chapter.

4.2.1 Attributional Style and Prior Experience

Although Individual Characteristics form part of the context of use, they are also influenced by it. Whether or not someone attributes any difficulties he has to himself or to the software, will depend not only on his own Attributional Style (pessimistic or optimistic) but also on the attitude of those around. Whether his expectations of the software are met will, to some extent, depend on the prior experience which has led to the formation of these expectations (Martinko et al., 1996). In both the case of Attributional Style and of Prior Experience, it is the *perception* of the user that is formed as a result, that will influence his satisfaction with the system. That perception could be that the system is easy or difficult to use, or that it serves a useful purpose. Attributional Style and Prior Experience, then, may help to explain why an individual perceives a system in a certain way but it is the perception that leads to satisfaction. They will not, therefore be included in the model.

4.2.2 System Experience and Level of Education

These are not affected by the organisational context. Educational qualifications are something which a person possesses regardless of the job he is performing. Whilst the benefit of an education may affect the individual's approach to the job, the job and surroundings in which the individual works will not affect the level of education. As has been discussed, there is conflicting evidence as to whether level of education is likely to improve or detract from satisfaction with the software. It is possible that this is because it is the type of job being performed as a result of educational qualifications rather than the qualifications themselves, that is influential.

The evidence in the literature is mixed as to whether educational qualifications affect user satisfaction positively or negatively and, indeed, to what extent they have an effect away from the context of the job. That being the case, the inclusion of this factor as an indicator of likely satisfaction would serve no useful purpose. It will, therefore, be dropped from the model being developed.

System Experience, as defined by Ramamurthy et al. (1992), means the amount of experience with the software being used. In the case of a package about to be purchased in order to help perform a task, system experience will not be applicable as, by definition, people will not have experience of using the software. Whilst it may be helpful later to explain why there are different levels of satisfaction amongst a user population, it can have little to tell us before that familiarity has been gained. Therefore, this factor will also be dropped from the factors affecting user satisfaction with packaged software.

4.2.3 Domain Expertise

Domain Expertise means that the user has a specific competency in the task to be performed. Whilst this was relevant to the study by Ramamurthy et al. (1992) where a DSS to support financial decisions was being considered, it is doubtful that this would be the case in relation to all software packages. The particular skill required in the study by Ramamurthy et al. was a good aptitude for decision making. This is a skill for dealing with information that is presented and is helpful in certain job categories. It is not a skill acquired through familiarity with the work domain (this is Domain Experience which was found to have no significant influence on satisfaction). Ramamurthy et al. (1992) agree that there may be some overlap between Domain Expertise and Domain Experience but say that it is useful to make the distinction for the particular work domain under consideration i.e. that of a DSS. This distinction would not apply in most areas of work. For example, someone may be very experienced in stock management. However, they may only have used the principle of first in first out (FIFO) and standard costing. If they were asked to use a system based on last in first out (LIFO) and average costing, it is their Domain Experience that would be important not their Domain Expertise.

Satisfaction with the DSS was measured according to satisfaction with the result of using the DSS i.e. satisfaction with its effectiveness in aiding decision making, not with use of the DSS itself. In addition to this, it was the *perception* of the people with greater Domain Expertise that the output information was relevant and useful. It is possible, therefore to hypothesise that it was the user perception that led to overall satisfaction. Their Domain Expertise *strengthened* the perception but it was the overall perception that was important. Therefore, this factor will not be included in the model.

Thus it is argued that Attributional Style, Prior experience and Domain Expertise will all result in an individual forming a perception. When trying to assess how well a software package will be accepted before it is introduced, it is the current user perceptions of it which are more likely to give an accurate indication. Other factors may help to explain why these perceptions have been formed but it is the perceptions themselves that will indicate future satisfaction. This means that none of the factors under Individual Characteristics will be included in the model.

4.3 Individual Perceptions

The Theory of Reasoned Action (TRA) (Fishbein & Ajzen,1975; Ajzen & Fishbein,1980) supports the hypothesis of the importance of user perception. TRA (figure 4.2) is a social psychology model which predicts actual behaviour (B) from behavioural intention (BI). Grounded in social psychology it has a general application and predictive models of technology usage have been based on it. Behavioural Intention is formed by a combination of influences which are both internal and external to the individual.

Internal influences derive from how an individual feels about a behaviour i.e. Attitude (A) towards the behaviour. This feeling (attitude) will be formed as a result of beliefs (perceptions) relating to the consequences of behaving in a particular way and the perceived benefits, or otherwise, of those consequences. External influences are a combination of how an individual perceives how others wish him to behave and a desire to comply with those wishes. This is termed the Subjective Norm (SN).

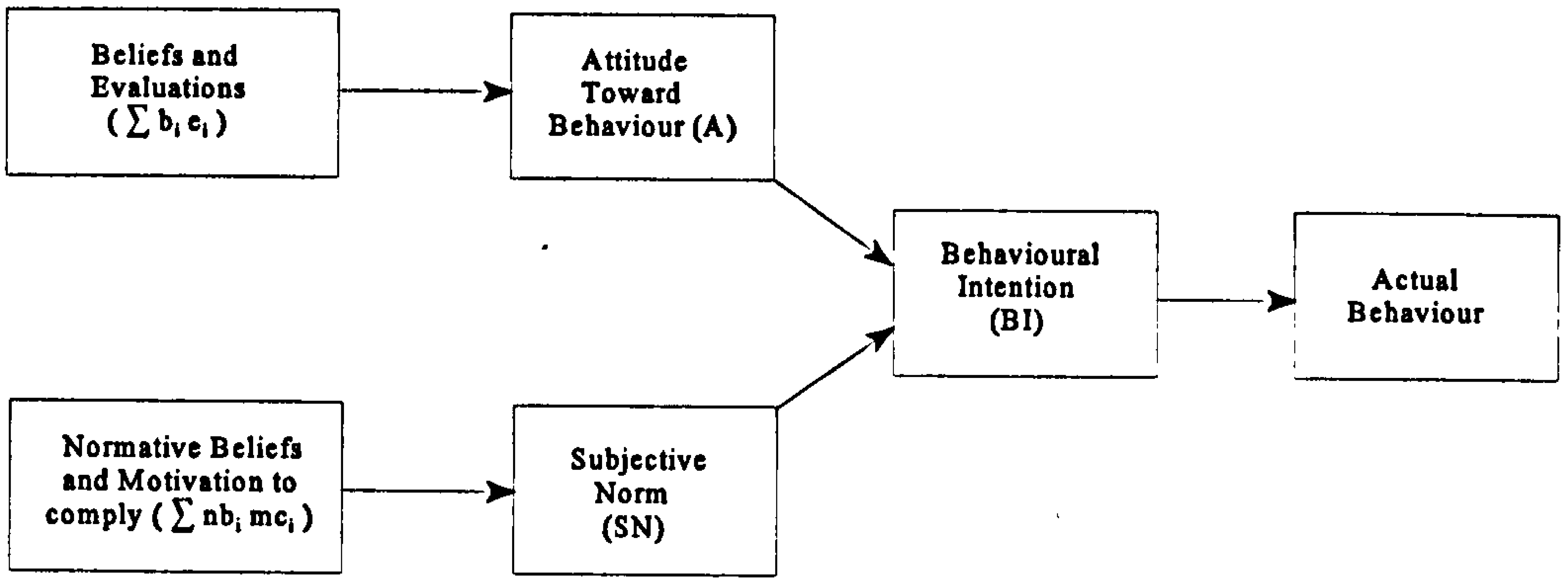


Figure 4.2

Both Attitude and SN, therefore, affect the BI of the individual to perform a behaviour. If these can be determined, BI and thence B, may be predicted, as factors which affect BI do so via one or other of these routes. However, it is the perceptions of the individual which determine their Attitude or Subjective Norm. TRA does not specify the specific beliefs or attitudes that may apply in any particular circumstance. It is up to the researcher to determine the belief set applicable to the behaviour in question.

Fishbein and Ajzen (1975) accept that it is possible to argue that an individual's beliefs about what others think, may affect his Attitude towards a behaviour, but they argue that it should be seen as a distinct construct because "a given factor may have different effects on the attitudinal and the normative components of the theory" (p.304). It is, therefore, preferable to retain two constructs (A and SN) in order to reflect the effect of different factors on each.

The Attitude construct consists of beliefs and evaluations of those beliefs. The majority of these beliefs will be formed as a result of prior information. This is similar to the suggestion by Martinko et al. (1996) that people's view of a computer system will be influenced by their prior experience with other systems. They are more likely to view a new system unfavourably if they have prior experience of a system that was difficult to use.

4.3.1 Perceived Ease of Use and Perceived Usefulness

Whilst TRA supports the hypothesis that user perception is important in influencing attitude, the Technology Acceptance Model (TAM) (Davis et al.,1989) supports the hypothesis that Perceived Ease of Use and Perceived Usefulness will influence satisfaction as defined by usage. TAM builds on TRA by first focusing on the particular area of information technology. Davis et al. (1989) contend that there are two principal perceptions which affect people's attitude towards using a technology (for which we may read software, as reported tests with TAM always use software as the object). These perceptions are: the perception of how free of effort using the software is i.e. Perceived Ease of Use (PEOU) and the perception of to what extent the software will help them to perform their job better i.e. Perceived Usefulness (PU). The antecedents to these beliefs are important only if it is necessary to establish what has caused them. It is the beliefs themselves that result in the formation of an attitude toward using the software. As in TRA, the attitude toward using will affect the behavioural intention to use and thence the actual behaviour. Whereas TRA is a generalised model, TAM is specific and the behaviour in question is that of using the software. The model is shown in figure 4.3. The model also shows Davis et al. (1989) findings that perceived usefulness has a direct influence on behavioural intention.

TAM uses self reported usage as an indicator of actual use. The amount of use which someone makes of software on a voluntary basis is taken to be an indicator of their acceptance of or satisfaction with it. However, this measure can not be used when use of a software package is dictated by the task to be performed and when an individual has no option but to use it. Such is often the case with packaged software purchased for use in an organisation. The way in which a task is to be performed and the means by which it is to be performed are often not at the discretion of the individual, more especially for those who are not at management level. TAM

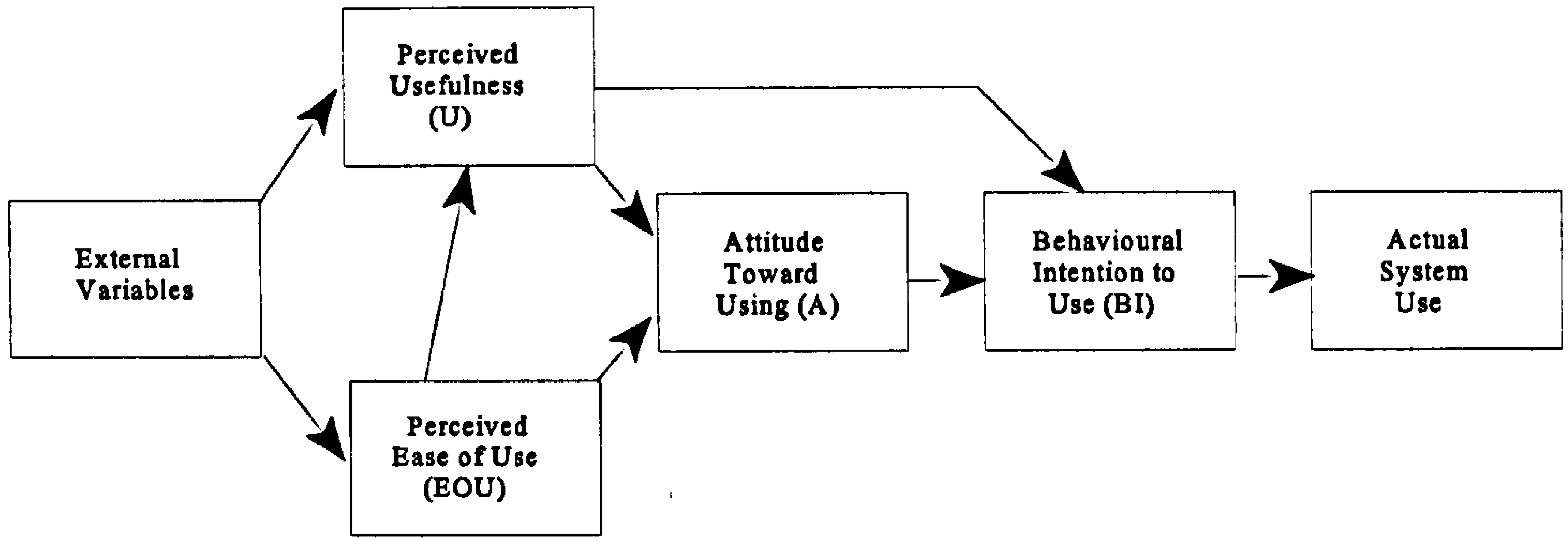


Figure 4.3

is a predictive model of an actual behaviour following from behavioural intention (BI). Satisfaction is not a behaviour. It is a feeling towards an object. To take usage as an indicator of satisfaction in the case of non-voluntary users would, therefore, be misleading.

The evidence of TAM is that PEOU and PU will influence a person's Attitude toward using a piece of software and the attitude thus formed will influence his behavioural intention to use the software. This is not the same as the ex ante Attitude described by Ramamurthy at al. (1992) who argue that Attitude should be treated as a separate construct from satisfaction. In their argument, Attitude is formed as a result of prior beliefs and experience and exists in relation to software in general. Satisfaction occurs as the result of actual experience with a specific software application. The Attitude construct of TAM is the result of perceptions formed in relation to a specific software application rather than the result of prior beliefs or experience. It is, therefore, more akin to the ex post Attitude of Ramamurthy et al. and thus to satisfaction.

Although Davis (1989, 1993) states that the Attitude construct relates to the Attitude towards a Behaviour rather than towards an object, this is not made clear in the TAM questionnaire. In the reported studies using TAM only that by Taylor and Todd (1995) does not specify a software application by name, although they do refer to a specific Computer Resource facility. If TAM were to measure only attitude towards a behaviour, the statements should have used the *type* of software application e.g. "I would find spreadsheet software useful in my job", rather than name an application. Naming a specific software application, confounds the response, which could refer to the action of using or to the object of the action. Indeed, in the case of the Perceived Ease of Use construct especially, it is difficult to see how this could be used in relation to anything but a specific object. This difficulty in separating Attitude towards the Object from Attitude towards the Behaviour with regard to software might help to explain why the Attitude construct has been found to be comparatively weak in explaining Behavioural Intention (Davis, Bagozzi & Warshaw, 1989).

There is much empirical evidence to support the two constructs of PEOU and PU used in TAM as being key factors in the formation of Attitude and BI (Adams, Nelson & Todd, 1992; Hendrickson, Massey & Cronan, 1993; Chin & Todd,1995; Davis & Venkatesh, 1996). Both of these are, therefore, retained.

4.3.2 Perceived Involvement

PD, JAD and DSDM all require that the users be involved in the design and implementation of a system. The purpose of this is to enable people to communicate their ideas to the system designers and, thereby, help to develop the system they want. It is also argued that, if people are fully involved in the development and implementation of a system they are more likely to understand what the system is designed to do and to accept its use (Mumford, 1991; Damodaran, 1996).

The review of the literature by Ives and Olson (1984) into the case for user involvement having a positive influence on system success found that, in fact, there was little evidence to support the theory. However, this review took place in the early 1980s when PD and other design methods and techniques for involving users were in their early stages. The argument of Ives and Olson (1984) was that Participation had not been properly defined in the sense of who should be the participants and in which type of systems design participation was beneficial. Neither had the construct of Satisfaction been measured in the same way across studies. These criticisms were of the lack of consistency in defining Involvement and measuring resulting Satisfaction. They did not suggest that there was no relationship between Involvement and Satisfaction, only that who was to be involved and when, required more research. They themselves propose a descriptive model of Involvement and their measure of Satisfaction includes Perceived Participation as one of the factors (Ives et al., 1983; Baroudi et al.,1986). Evidence of the influence of Involvement on Satisfaction is also provided by Alavi (1984). In a controlled experiment using MBA students she found that involving users by creating prototypes and encouraging feedback, resulted in a higher level of satisfaction, than did a structured design method.

Whilst it continues to be true that there is no consensus reporting in the literature of who should be involved in an organisation for the development of different systems, this must be partly due to the varying nature of the organisations themselves and the fact that it is not possible to control the variables within them as it is in a laboratory experiment. Neither has the nature of what it is to participate been fully defined (O'Neill, 1998) but there is general agreement that it means an active role in decision making on the part of the people who will be using the system. This involvement activity has to not only take place but to be *perceived* to take place by the users. Whilst this is not always explicit in the literature, it is implicit in the assumption that the involvement will lead to greater acceptance of the system.

Measuring the satisfaction that occurs as a result of involvement, in a consistent fashion, also remains a problem (Clement & Van den Besselar, 1993). Indeed Heinbokel et al. (1996) report an increase in difficulties with projects according to the degree of user involvement. However, in their study, success is measured according to the reports of the software designers. It is not known whether the users themselves were happy with the end result. As Heinbokel et al. point out, the results are not an argument against user involvement but evidence that it may not be all good and that the benefits should be weighed against the cost.

Despite these problems with the user involvement construct, methods such as JAD and, more recently, DSDM continue to increase in use as does the principle of design iteration, which requires user input. These may still be largely based on "intuition, experience and unsubstantiated perceptions" (Ives & Olson, 1984,p.584) but the weight of their combined evidence is heavily in favour of the need for the user not just to be involved but to perceive himself to be involved in the process if he is to feel satisfied with the end result.

Methods to involve people in software development include such things as newsletters, having user representatives on steering committees and even in design teams (Eason & Sell, 1981). It is not possible to involve the end user of a COTS package in its development. The software is developed for sale in multiple organisations and is based on a general model which should take account of the measurable attributes of Usability. However, when a COTS package is considered, the purchaser will assess how well it meets the functional requirements of the organisation by measuring it against the requirements checklist. It is possible to involve the people who will use the product both in the initial specification and in the checking of that specification against a package and to keep them informed of what is happening. It is also possible to involve them in the final determination of which package to purchase. This would enable people to express their opinion about what is important in a package and also to influence the decision on which package should be purchased. This compares with the definition of Involvement as people being present at decision-making meetings and having the opportunity to communicate their views and to have an influence on the decisions taken. However, it is necessary not just that this should be the case but that the individual should perceive it to be the case. Perceived Involvement is, therefore, retained.

Table 4.2 shows the revised list of the key factors affecting an individual's disposition towards packaged software.

System Characteristics	**Individual Perceptions**	**Organisational Characteristics**
Functionality Measurable Usability	Perceived Usefulness Perceived Ease of Use Perceived Involvement	Organisational Type Organisational Politics Management Attitude & Co-worker Attitude

Table 4.2

Individual characteristics have been removed as it is argued that it is the effect of those characteristics on the user's perceptions that is of primary importance in the development of a simple model. This is not to say that individual characteristics are not important but that, as with the external variables in TAM (Davis et al.,1989), their importance would lie in the explanation of how a perception has arisen not in predicting the result of that perception. Perceived ease of use, perceived usefulness and perceived involvement are all retained and it is argued that each will contribute to the disposition of the individual towards the software. Perceived ease of use and perceived usefulness are derived from the evidence of TAM (Davis et al., 1989) but it is proposed that they will influence an individual's disposition rather than his behaviour when software use is mandatory. This leads to the hypotheses that:

H2 *"People will be positively disposed towards a software package if they perceive it to be easy to use"*

H3 "people will be positively disposed towards a software package if they perceive it to be useful in performing their job".

The influence of PI is derived from evidence in the literature. However, in the case of packaged software the nature of the involvement is necessarily different from when software is developed in-house. This leads to the hypothesis that:

H4 *"people who feel they have been involved in the selection process of packaged software, will be positively disposed towards the software".*

4.4 Organisational Characteristics

Organisational characteristics are the final factors for consideration. The factors listed under this heading are: Organisational Type, Organisational Politics, Management Attitude and Co-worker Attitude.

4.4.1 Organisational Type

Hornby and Clegg (1992) describe two types of organisation:- a) the Mechanistic, where activities are proceduralised and the structure is hierarchical, and b) Organic, where activities are not formalised and there is no strict delimitation between jobs and responsibilities. According to Hornby and Clegg, the people in each of these organisational types will react differently under certain circumstances. The aim of their study is to assess the effect on participation of different organisational contexts. The physical structure of the organisation in terms of all personnel being on one site or being spread over a geographical area is not discussed. The results of their study in a UK bank, described as mechanistic in structure, support the proposal that the introduction of a system which has far reaching effects on the existing structure, can affect the normal communication channels and result in authoritarian decision making and dissatisfaction with the amount of perceived participation.

It is unclear from the case studies by Rice and Tyler (1995) and Caldwell and Uang (1995) whether the organisations were mechanistic or organic in nature, although the state departments described by the latter are more likely to be mechanistic and the former describe one of the organisations in the study as conservative and proceduralised, which would also suggest it is mechanistic in nature. In the case of Rice and Tyler (1995) the more conservative and, therefore, possibly mechanistic organisation seemed to report greater satisfaction with the system. In the case of Caldwell and Uang (1995) the results are mixed. The main aim of both these studies was to investigate satisfaction with software, whereas Hornby and Clegg (1992) are investigating satisfaction with the perceived level of participation. However, they are worth comparing as they are all concerned with satisfaction and all refer to the type of organisation.

All three studies would seem to support the influence of organisational type on user satisfaction, either with the software or with the degree of participation but they do not present any clear indication of what this influence is or how it occurs. Indeed, Hornby and Clegg (1992) warn against assuming some form of organisational determinism because, they argue, organisations are not fixed structures. They consist of people and people with authority in an organisation have the power to alter the context of projects under their control. Indeed, Hornby and Clegg say that the effect of the new system on the organisation described, was to disrupt normal communication channels so that decision making became authoritarian. This was the reaction of managers within the organisation to a situation. Although the type of organisation may have influenced the form the reaction took, it was the reaction of management (i.e. the authoritarian mode of decision taking) which affected people's perception of their involvement.

This may also explain the difference between the studies by Caldwell and Uang (1995) and Rice and Tyler (1995). Although the organisations they describe appear to be of a similar type, the outcomes of the studies are different. This could be because the reaction of the managers to the situation of the introduction of new software was different in the different organisations. Thus two apparently similar studies achieved different results.

This suggests that, in considering the factors which affect user satisfaction with software, there are more direct influences than the type of organisation in which the user works. Whilst the type of organisation may influence the management approach, it is the management approach itself which directly influences the individual. Organisational type will not, therefore, be included as a separate factor.

4.4.2 Organisational Politics

The effect of the power politics within an organisation has been shown by Keen (1981) and Markus (1983). People may feel allegiance to one group and, therefore, favour software that is seen to increase the authority of that group even though the software itself may be lacking (Kling & Iacono, 1984). The politics within an organisation may be affected by the structure of the organisation. But, like the structure of an organisation, the politics within it are dependent on the people with influence. Where those people are in the hierarchy of the organisation - close to the end users of computer software, as in an organic structure, or more removed as, possibly, in a deterministic structure - may depend on the organisational type but it is the people themselves and their relationship to the end user of the software who are most likely to influence that end user's perception of the software. Thus Organisational Politics as a separate factor will also be dropped. This is not to say that organisational politics do not have an impact on the people within the organisation. People may behave differently according to the situation in which they perceive themselves to be and the politics within an organisation may affect the behaviour of management just as individual characteristics may affect how someone reacts to a situation. However, it is the effect of that behaviour, not its cause , which is of interest in this research because it affects the context within which the individual works.

4.4.3 Management and Co-Worker Attitude

Igbaria and Nachman (1990) find a significant correlation between leadership style and satisfaction. Managers who are considerate towards their staff and who have a clearly defined and understood approach towards a project, have staff who are more comfortable with use of the computer system. Hartwick and Barki (1994) also found that the perceived attitude of managers and co-workers towards the system influenced intention to use and thence usage and that this influence was stronger in the early stages of development before people had gained sufficient experience to form opinions more independently of others.

The influence of others whose views are important to an individual is termed Subjective Norm in the Theory of Reasoned Action (TRA) (Fishbein & Ajzen,1975; Ajzen & Fishbein,1980). An individual's normative beliefs (Subjective Norm) will depend on the perceived attitude of others whose views are important to the individual and on the individual's desire to comply with that attitude. Who these people are will vary depending on the circumstances. The attitude of others may be inferred rather than known by the individual. So, for example, if a referent is known to believe that something will be beneficial, the inference is likely to be that they are, therefore, in favour of it. Whilst in many instances, the referent others are likely to be friends and family, in the workplace they are more likely to be co-workers and managers. In the case of friends, the motivation to comply with their wishes will probably be simply out of a desire to please them. However, there is some evidence that motivation to comply may also be related to the referent's power over the individual (Fishbein & Ajzen, 1975). This suggests that, in the workplace, motivation to comply may be strengthened if the individual perceives others to be in a position of power.

It should be noted that a normative belief is when an individual believes that a referent other wants him to perform a certain behaviour. It is not the same as believing that the action of performing a behaviour would please the referent other (Ajzen & Fishbein, 1980). This distinguishes it from that part of Perceived Usefulness (PU) in the Technology Acceptance Model (TAM) (Davis, 1989; 1993) which indicates a belief on the part of the individual that using a system will please those in authority and potentially result in a pay rise or promotion. In fact, in the TRA model, this part of the PU construct would form one of the elements of the Attitude construct as it could be said to arise as a result of the individual's beliefs about the outcomes of using a computer system and his evaluations of the benefits of those outcomes.

TAM (Davis et al., 1989) suggests that perceived ease of use (PEOU) and PU affect Attitude and that the Attitude thus formed influences Behavioural Intention (BI). It is BI which predicts the actual behaviour of usage. It is also argued that, contrary to the TRA model, the beliefs formed regarding whether use of a system will improve chances of job promotion which are captured as part of PU, will directly influence BI regardless of the Attitude (positive or negative) a person may have formed towards the behaviour itself (Davis et al.,1989). The intention to perform the behaviour in such cases is based on a cognitive appraisal of how the action itself will be of benefit. In this a parallel may be drawn with acceptance of a system even though it is disliked, because there is no alternative. Such acceptance does not equate to satisfaction.

It is useful, here, to draw a distinction between Attitude towards an Object and Attitude towards a Behaviour. TRA is based on the Attitude formed towards a Behaviour. TAM is concerned with the specific behaviour of use. In the model being developed, the interest is in the factors which will lead to satisfaction with a specific object (i.e. packaged software). Satisfaction is a state of mind formed as a result of certain perceptions and experiences. It is the factors which lead to the formation of an attitude which corresponds to later reported satisfaction, which are of importance in this thesis. The behavioural intention arising as a result of that attitude or influenced directly by other factors, which results in a behaviour, is not of concern because the behavioural intention can only predict voluntary behaviour. Thus the PU construct must have a narrower definition than that assumed (although not explicitly stated) by TAM: The definition of PU is that the software is perceived to help the individual in the performance of the job, without the extension that using the software will result in improved chances of job promotion.

TAM (Davis et al., 1989) uses TRA (Fishbein & Ajzen, 1975) as its theoretical basis. However, the SN construct does not appear in TAM. It is argued that it is difficult to show the direct affects of SN on BI separately from its indirect affects via Attitude (A) (Davis et al., 1989). In their comparison of the TRA and TAM models, Davis et al. (1989) found little support for the effect of SN on behavioural intention with regard to use of a computer system. However, as they acknowledge, the study was carried out using MBA students as subjects and the software was a wordprocessing package. The type of software used (a single-user application) means that there was no inter-dependency between the students on usage. The fact that the subjects were students also meant that there were no direct referent others in a position of authority (usage of the software was not required as part of the course) who might influence them. They were also MBA students - i.e. of a managerial level and thus more likely to be autonomous with regard to their work. Under these circumstances, it is perhaps not surprising that the SN construct appeared to have little influence.

Mathieson (1991), comparing the Theory of Planned Behaviour (TPB) and TAM, reports similar results with regard to SN. TPB is an extension of TRA adding Perceived Behavioural Control (PBC) (i.e. the individual's perception of his control over performance of the behaviour) to the constructs of A and SN as influencing BI. This study also uses students and, for the task they are given, it is the result of the task, rather than the tool used to achieve it, that would be important. Thus the statements used such as "Those people who are important to me would (strongly support/strongly oppose) my using [the spreadsheet] rather than my calculator for the assignment" would have little real relevance to the respondents. The responses are anonymous, and which tool was chosen to complete the assignment would not affect the marks given. As in the study by Davis et al.(1989), it is not surprising in the circumstances described, that there should be little support for the effect of the influence of others.

The findings of Davis et al. (1989) and of Mathieson (1991) with regard to the influence of SN contrast with the findings of Taylor and Todd (1995) who compare TRA, TPB and the Decomposed Theory of Planned Behaviour (DTPB) which proposes specific antecedents to Attitude, SN and PBC. Taylor and Todd (1995), find that SN does help to predict usage, especially where people do not have prior experience of the system. They explain the difference between their findings and those of Davis et al.(1989) as being due to the fact that, although they too use students as their subjects, the situation is more naturalistic: The students were using the software on real projects and in a relatively competitive environment. Unlike the study by Mathieson (1991), how the students carried out their projects was likely to be influenced by their perceptions of what their professors would want, due to the potential impact on their grades of their actions. Thus it would seem that, in a context where others are in a position of influence, SN does have an effect on the action of the individual.

The study by Taylor and Todd (1995) provides evidence in support of the SN construct as influencing BI. Davis et al. (1989) omitted this construct from TAM because of the difficulty in ascertaining its effect on BI separately from its influence on the Attitude towards performing the behaviour. There is evidence in the literature to suggest that SN will indeed influence people's attitude towards an object (e.g. Martinko et al.,1996; Hirschheim & Newman,1988; Damodaran,1991). It is the contention of this research that, whilst SN may directly affect BI where usage is voluntary, when usage is a requirement the Attitude formed towards the object of that usage will also be influenced by the perceived attitude of referent others.

The final list of key factors affecting an individual's disposition towards packaged software is shown in table 4.3 below.

System Characteristics	**Individual Perceptions**	**Organisational Characteristics**
Functionality Measurable Usability	Perceived Usefulness Perceived Ease of Use Perceived Involvement	Management Attitude & Co-worker Attitude

Table 4.3

Organisational Type and Organisational Politics have both been removed from the list because, it is argued, these are an indirect influence via the mediation of management and co-workers. Whilst the type of organisation and the politics within it may influence user satisfaction with software, it is the perception of the attitude of referent others i.e of the people who are important to the end user (regardless of how that importance has arisen) which will have a direct influence on his perception of the software. This leads to the hypothesis (H5) that "*the perceived attitude of management and of co-workers towards using a software package, will affect the disposition of the individual towards it*".

4.5 A model of user engagement

In summary, the factors which have a direct influence on the disposition of the individual to feel favourably towards packaged software, prior to its implementation, may be expressed in the model below. It is suggested that a positive disposition towards a software package will correlate with satisfaction post implementation. This positive disposition may be compared with the "user engagement" of Kappelman and McLean (1991). They argue that user engagement arises as a result of the combination of involvement in the design and development process and the personal relevance of the system that occurs as a result and that this will lead to satisfaction.

The model below (figure 4.4) proposes that the perceived ease of use (PEOU) and perceived usefulness (PU) of the software combined with the perception of involvement (PI) in the decision process leading to the choice of package and the combined influence of line managers and co-workers (SN) will lead to a sense of engagement with the software. The strength of this engagement will indicate the strength of satisfaction that is likely to be felt with the software post implementation.

Model of User Engagement

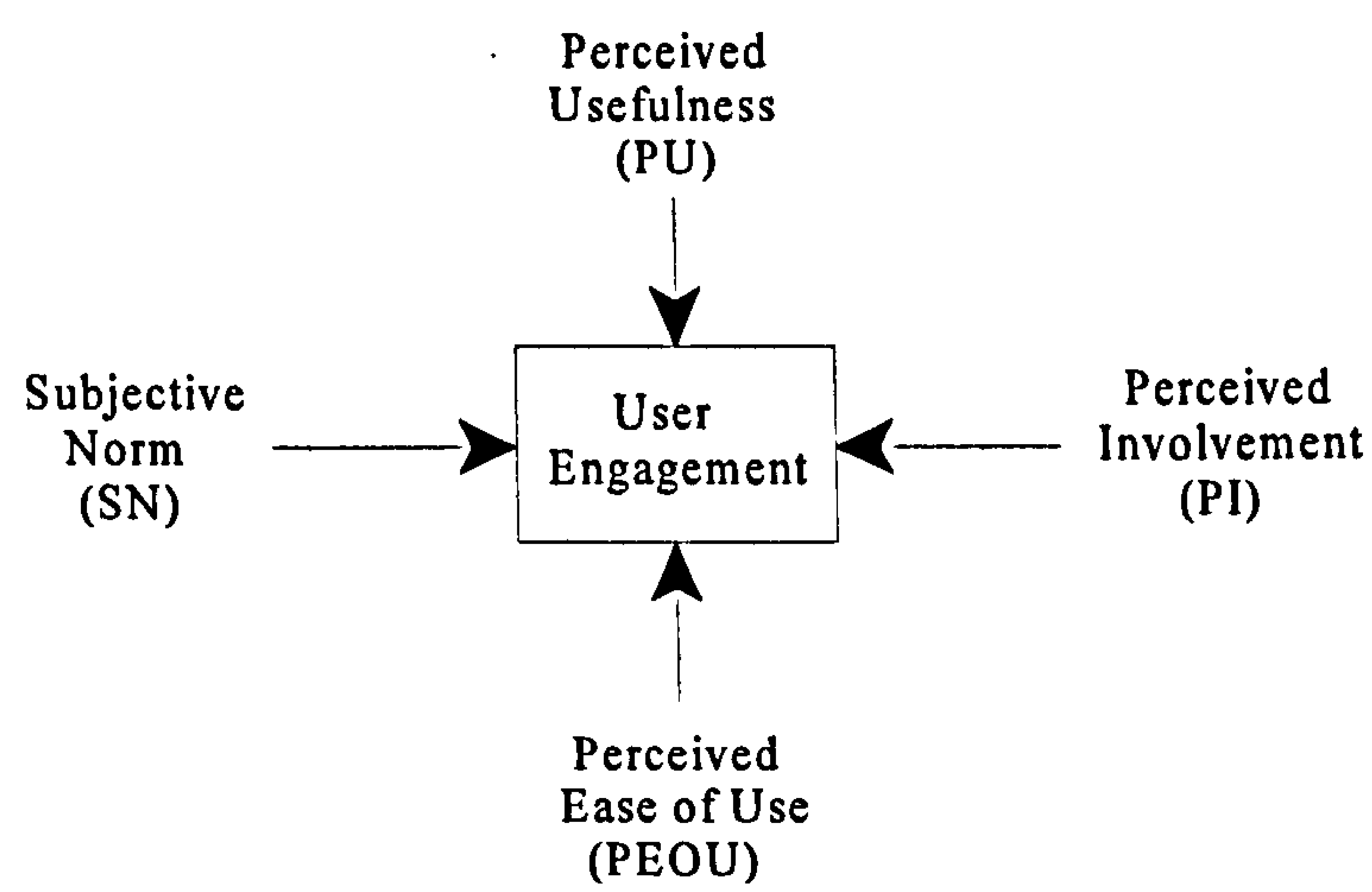


Figure 4.4

The next section describes the approach taken in order to test the model and the research tradition on which the approach is based.

4.6 Research Approach

Five hypotheses have been proposed in the preceding sections of which four are directly related to the model. The first hypothesis i.e. that system characteristics alone are not sufficient to ensure user satisfaction, is a necessary precursor to those of the model: - If the characteristics of a software package are sufficient of themselves to ensure user satisfaction, then it must be possible to create a universally satisfactory package. Research should concentrate solely on the development of the product and ignore any influence of context. The four hypotheses on which the model is founded assume that context is important and propose that people's perceptions will affect their satisfaction and that it is the combined influence of these perceptions that will affect their overall engagement with the product.

All of the hypotheses formulated are drawn from work which has been reported in the HCI literature and this thesis may, therefore, claim to be based in the discipline of HCI. However, that claim raises several questions:

- Is there an HCI discipline?
- On what established theory of HCI have the hypotheses been based?
- What research method is appropriate to the HCI discipline?

4.6.1 The HCI discipline

The fact that HCI draws on several different disciplines raises the question of what the boundaries are of HCI itself as a discipline. Long and Dowell (1989) give the definition of a discipline as "..the use of knowledge to support practices seeking solutions to a general problem having a particular scope"(p.12). In addressing the question of whether Management Information Systems (MIS) or Information Technology (IT) may be classed as a discipline, Farhoomand (1992) cites HCI as the third most popular area of research within MIS and, therefore, as a sub-discipline. He defines MIS as "..an integrated user-machine system for providing information to support operations, management, analysis and decision making functions in an organization"(p.93). He further argues that MIS is a discipline separate from those to which it refers in its search for knowledge, although lacking in a body of substantive theories itself. If the definitions given of HCI in Chapter two of this thesis are considered, they appear to have a very similar scope to that proposed for MIS by Farhoomand (1992) as they are concerned with the problem of creating an integrated user-machine system. It could be argued that the general problem addressed by MIS is that of providing information to support various functions within an organisation, whereas the general problem addressed by HCI is that of how to design computer systems such that people are readily able to interact with them in order to produce information. However, the definition of HCI as given by Long and Dowell (1989) incorporates both the one to one relationship between person and computer and that relationship set within the context of an organisation with the implications that may hold with regard to the behaviour of the individual and the design of the computer.

It is difficult from these definitions to make a distinction between MIS and HCI. However, the argument used by Banville and Landry (1992) that the number of journals and university courses which cover the subject area normally referred to as MIS or IT means that it may be considered a distinct discipline, is also true for HCI. It may, therefore be argued that there is an HCI discipline even though its scope is unclear. The argument over the boundaries of HCI continues, as evidenced by a panel discussion at HCI 2000[7] where it was suggested that the problems addressed by HCI are being hijacked by other areas such as Software Engineering.

4.6.2 The theory of HCI

The problem of establishing a definition and scope for HCI is reflected in the discussions over what constitutes a theoretical framework for HCI as a discipline. Many disciplines have been drawn on from both the natural and social sciences and there is a strong cognitive science tradition in HCI. Indeed, Nardi (1996a) says that it is "the dominant theoretical voice in HCI" (p.13). There is good reason for the cognitive approach. The field of HCI had its nascence in the early 1980s at the time that the PC was being introduced (Alexander,1987). The concepts of short and long-term memory and of processing data mapped well onto both the human and the computer partner in the interaction. It is this conceptual framework which assisted the development of Nielsen's seven layer model (1986) and the GOMS (Goals, Operators, Methods and Selections) model of Card, Moran and Newell (1983). Such models are helpful to designers in evaluating one design against another in terms of efficiency of use. However, cognitive theory alone does not provide a sufficient base from which to increase knowledge in the field of HCI in its broader definition as given by Long and Dowell (1989) as it focuses on the individual and the computer without reference to the wider social context. The GOMS model itself has to

[7]The Panel discussion was called " HCI challenges for the new millennium ". Contributors were Hilary Johnson, Alistair Sutcliffe, Mike Wilson and Chris Johnson

assume an expert user working without distraction.

In suggesting a framework for HCI, Long and Dowell (1989) propose that there should be a dual conception of the field. They take the general problem addressed by HCI according to their definition and divide it into two perspectives i.e. the problem of "the design of humans interacting with computers" and "the design of computers interacting with humans" (p.14). These two areas are addressed by the disciplines of Human Factors (HF) and Software Engineering (SE) respectively. Thus, they argue, the knowledge which supports the practice of HCI is a combination of knowledge from the HF and SE disciplines.

Mackay and Fayard (1997) also propose a cross disciplinary framework drawing on both the natural sciences and design, and give examples of how the problem posed by HCI may be addressed depending on the nature of the problem (i.e. predominantly human or technical).

Other authors have discussed the possibility of using Activity Theory as an over-arching theory of HCI (ref. Nardi, 1996b for supporting papers). In Activity Theory the computer is seen as an artefact which mediates between a person and his activity. It is the activity which is the unit of study and this necessarily encompasses the person doing (i.e the subject of the activity) and the computer (i.e. an object in the activity).

In summary, the discussion of what constitutes the scope and theory of HCI continues. As PCs have become more pervasive in the workplace, the interaction problem has extended from human to computer and computer to human, to human to human interaction via computer and within a social context. This has necessitated reference to an increasing number of disciplines thus further blurring the boundaries of HCI.

4.6.3 Theoretical basis of this research

The first hypothesis proposed in this thesis relates to the software package and, therefore, belongs in the SE aspect of the framework for HCI proposed by Long and Dowell (1989). The remaining four hypotheses are concerned with what influences people's attitudes towards (engagement with) a software package and, therefore, relate to the HF aspect of the framework proposed by Long and Dowell(1989). Three of the hypotheses (H2, H3, H5) have their roots in the Theory of Reasoned Action (TRA, Fishbein & Ajzen,1975) from the field of social psychology. The hypotheses relating to perceived ease of use and perceived usefulness (H2 & H3) are a result of the work based on TRA undertaken by Davis in developing the Technology Acceptance Model (TAM) (Davis et al.,1989; Davis, 1989; Davis, 1993). The difference is that TAM is concerned with the effect of perceived ease of use and usefulness on behaviour (i.e. usage) whereas the model proposed is concerned with their contribution towards attitude (i.e.user engagement). The hypothesis relating to the influence of referent others (H5), is derived directly from TRA (Fishbein & Ajzen, 1975).

The fourth hypothesis (the need for involvement) draws on socio-technical theory and other work in the field of HCI, especially that of Participatory Design. According to these, the involvement of people in change which affects them, is necessary in order that they may gain an understanding of and commitment to the project. In the case of the introduction of a software package the hypothesis is that if people are involved by having a say in which package is selected, they will feel more favourably towards it.

4.6.4 Research Methods in HCI

All of the discussions on what constitutes the theory of HCI recognise the difficulty of exploring a field wherein one object of exploration (the computer) is being constantly developed and the other object (the human) does not fit a standard type or category and whose actions in a given situation may not be readily generalised because the variables that affect the situation may not all be known. This latter problem is the problem faced by all social sciences and is the reason the question is sometimes raised as to whether they may be called a science at all (Willer & Willer, 1973). This in turn has led to a re-examination of what is meant by science and of what constitutes scientific method (Hirschheim, 1992; Banville & Landry, 1992).

Long and Dowell (1989) also discuss the question as to whether HCI should be seen as a Craft, Applied Science or Engineering discipline. In a Craft discipline, they argue, it is practice which maintains and increases knowledge. In other words, the knowledge of the craft takes its form as design principles or heuristics, such as "simple operations should be simple, and the complex possible" (p.18). These are applied to specific instances of the general problem of the interaction of person with computer and retained or discarded according to their efficacy. The method then is to make a proposition based on intuition or experience and to refine that proposition based on the evidence of experience. The drawback with this approach is that it is experiential and not explicit or formal. The people applying the heuristics define what they mean e.g. by saying what is meant by simple in relation to what is being designed or what action is being undertaken. There is no single definition which applies in all cases. This in turn means that it is not possible to generalise the knowledge gained beyond the specific application for which it was used. The expert knowledge of the "craftsperson" might enable him to apply the principle or heuristic in other situations but there is no rule which the non-expert can apply. Whilst Long and Dowell (1989) acknowledge that some heuristics may be refined by iteration to a fixed rule which may be validated, they argue that this would then no longer be craft knowledge. It must also be added that such design rules (e.g. specifying a fixed, standard screen layout for all applications) may actually be restrictive and not add to the ease of use or effectiveness of individual applications (Grudin, 1989).

HCI conceived as an Applied Science discipline also presents problems. In the scientific method a proposition is made based on theory, an experiment designed and data gathered, the results of the data are analysed and the original hypothesis may be refuted, found to be supported or modified. Although Long and Dowell (1989) state that "scientific knowledge is explicit and formal, operational, testable and generalisable"(p.20), they argue that the application of such knowledge derived from the SE element of HCI will contribute only to the design of the computer and the knowledge derived from the HF element will explain only the behaviour of the individual. There is no conjoining of knowledge from which the overall problem of people interacting with computers may be approached. Although Long and Dowell do not say so, this may be seen as a restatement of the lack of an over-arching theory of HCI and the difficulty that arises from the framework proposed by them wherein SE and HF are distinct elements. This difficulty is not so pronounced when HCI is conceived as a craft as the emphasis is on iteration and refinement as the effect of human on computer and computer on human is observed.

The concept of HCI as an Engineering discipline as proposed by Long and Dowell (1989), would overcome the problem of linking the SE and HF disciplines. The scope of one would overlap the scope of the other. The requirement for each would be the interactive work system and when the behaviour of the computer is specified, HF would prescribe the behaviour of the user; when the behaviour of the user is specified, SE would specify the behaviour of the computer. In this approach intuition and experience are tested and lead to design principles which may be applied by inexperienced others. However, as they go on to say, neither SE nor HF currently possess engineering principles of this idealised form.

Long and Dowell (1989) conclude by suggesting that the different conceptions of HCI should be regarded not as mutually exclusive but as mutually supportive. The means by which knowledge is acquired and the methods employed will vary according to the context of the problem being addressed. This view is echoed by Sanderson and Fisher (1994) and Land (1992) amongst others (see Galliers, 1992a).

4.6.5 Research Method used in this Thesis

Scientific method may be said to have two elements:-

- theory building in which the subject matter is explored, concepts developed and hypotheses generated
- theory testing in which the hypotheses are tested by means of experiment. The data gathered in the experiments are generally such that statistical analysis may be performed on them. The results of the experiments are discussed and, as a result, the hypotheses may be modified, refuted or found to be supported (Rivard & Huff, 1988).

This thesis builds on existing theory in order to generate hypotheses and so may be said to be using the scientific method. However, the hypotheses proposed do not lend themselves to an experimental approach. Such an approach demands that the variables may be controlled or at least measured. When what is being considered is the interaction between an individual and a specific interface, a laboratory based experiment, where the environment is controlled and only details on the interface are altered, might be possible. Even then, certain assumptions have to be made about the human side of the interaction if the results are to be generalisable (as in the Key Level Model of Card et al., 1983, where an expert user is assumed). According to the model proposed, the dependent variable is User Engagement and each of H2 through to H5 are the independent variables. In order to refute the hypotheses experimentally, it would be necessary to be able to control and measure each of these and remove the possibility of contamination by other variables.

The problem of applying the scientific method to research in real-world situations and the need for different approaches according to the research question being addressed is generally acknowledged (Gale & Christie,1987; Galliers,1992b; Hirschheim,1992; Sanderson & Fisher,1994). Galliers (1992b) suggests that there are two common approaches: The Scientific approach and the Interpretivist approach. According to Galliers (1992b) "Scientific approaches may be defined as those that have arisen from the scientific tradition - characterized by repeatability, reductionism and refutability ....and which assume that observations of the phenomena under investigation can be made objectively and rigorously...Interpretivist approaches argue that the scientific ethos is misplaced in social scientific enquiry because of..the possibility of many different interpretations..the impact of the social scientist on the social system...the problems associated with forecasting future events.."(p.148). From this, it may seem that the two approaches are mutually exclusive. However, this does not need to be the case. In any study there may be more than one perspective and a different approach may be appropriate to each part. In the case of this research it was felt that a mix of the Scientific and Interpretivist approaches as defined by Galliers (1992b) was needed.

Support is sought for the first hypothesis by comparing the level of satisfaction with three packages, used by three different workgroups, with their level of usability. Satisfaction with the individual packages was measured by use of an industry standard questionnaire and confirmed by discussion with the users. Usability was measured by an independent evaluator completing a series of checklists which have also been researched and used in industry. The scores achieved by each package according to the checklists are compared with the scores achieved on the satisfaction scale. The results of these comparisons for each of the packages are then compared between the packages. This serves to show that a high score on the usability checklists does not necessarily equate to a high score on the satisfaction scales.

The results are merely those for three packages within one organisation. In order to generalise the results it was necessary to obtain the satisfaction scores for the same packages, from a number of different organisations. It would then be possible to see if all the satisfaction results were similar, in which case the cause of satisfaction or lack of it could be simply due to the software. Had all the satisfaction scores been at a discrepancy with the evaluation scores e.g. the satisfaction score was consistently high and the evaluation score consistently low, it would suggest that the interpretation of the evaluation was not correct. It was essential that the same packages were used for this as to obtain satisfaction scores for a number of different packages from different organisations and compare them with their usability evaluations would effectively be a series of one to one comparisons from which no picture could be built.

In the event, as only one organisation was found which was using the same software and agreeable to completing the satisfaction rating, the market performance of the software packages was used as an indication of general satisfaction levels. This places the method used towards the interpretivist side of the scale, as an inference is made from the discrepancy between the satisfaction ratings of the software packages and their scores on the evaluation checklists without the evidence of quantity.

Support for the four hypotheses relating to the model took the form of case study and questionnaire and the method is closer to the scientific than the interpretivist approach. Questionnaires have the advantage of enabling generalisation of results given sufficient quantity of data. However, they provide little insight regarding the cause behind the phenomena under study (Galliers, 1992b). Case studies follow the social research tradition and the emphasis is on qualitative rather than quantitive data (Sanderson & Fisher, 1994). The fact that a case study takes place in a single organisation may be seen to be a weakness as it is difficult to generalise the findings. Although as Galliers (1992b) states, the case study approach is usually placed under the scientific banner, the method used is very similar to that categorised by Long and Dowell (1989) as belonging to a craft: The knowledge gained from one study may be used and refined in another but the approach is experiential. The use of both questionnaire and case study in the research reported here meant that the strength of one helped to balance the weakness of the other.

•Research steps

There were five stages to the research:

a) Assess current level of satisfaction

b) Develop and test questionnaire

c) Case Study 1

d) Case Study 2

e) Comparison of results with each other and with the User Engagement Model

Step a) served a twofold purpose. It provided data both on the current situation with regard to satisfaction with software and the factors which may have contributed to that satisfaction. It also provided the data with which to test H1 within the organisation.

A brief summary of the steps is given below.

a) Assess current level of satisfaction

The first stage was to ascertain the current level of satisfaction with three COTS packages (IM, MM, PM) each used by a different workgroup within the organisation. Data was gathered by use of an industry standard questionnaire to measure satisfaction. Separate data were gathered by means of semi-structured interview. The data gathered was on computer usage background of some of the users of the software, how much they felt they had been involved in the selection of the package and what they felt about the software. Records were checked to confirm the author's memory of how the packages had originally been selected. These data were compared with the results of the satisfaction measure, the factors proposed in the model and the factors not proposed in the model but which other research had found to influence satisfaction. The results were cross-checked by interviewing a sample of users. This is necessarily an interpretivist approach and relies on the "inside" knowledge of the author in order to make inferences from what is said and how it is said. However, it gave a baseline measure of current satisfaction and provided evidence that the proposed factors contributed to it.

The next stage was to see if the same factors were influential before experience with a package was gained and whether the sum of the factors (user engagement) corresponded to satisfaction. An exact correlation would not be expected as, by definition, people could not be satisfied with something they were not familiar with. However, a correspondence between their initial feelings towards the package (engagement) and their later satisfaction with it, was proposed.

b) Develop and test questionnaire

In order to be able to produce a measure of the proposed factors to correlate with the measure of satisfaction, a questionnaire was developed. Full validation of the questionnaire required that it be issued before the introduction of a package and the results correlated with those of the satisfaction measure issued after experience with the software had been gained. As the influence of others is one of the proposed factors, a real world environment was demanded in which use of the software was a requirement. Several organisations were contacted and three agreed to assist. Two of these were multi-national organisations (for convenience called A and B) and the third (C) was a software reseller and trainer. The latter agreed to ask people on training courses to take part. This meant that the respondents would be from various organisations but all using the same software. As the issuing of the questionnaires had to be via an intermediary an explanatory note was written to be handed out with each questionnaire. In these a brief background to the purpose of the questionnaires was given and people were told that the information would remain confidential and they would not be identified as individuals. They were also thanked for their time. In the case of the third organisation (C) some additional questions were included. This was in order to gain some background information on the type of organisation the respondents worked for and the size of the workgroup using the software. They were also asked whether they felt that there was management support for the use of the software. This background information was provided by the contacts at the other two organisations.

As people could remain anonymous each set of questionnaires was numbered in order to ensure that the correct results were used in the correlation and, as an extra precaution, the stamped addressed envelopes for their return were also numbered. The request was that the first questionnaire be completed and returned after first experience of the software (viewed demonstration or hands-on training) and that the second questionnaire be returned once familiarity was gained.

The number of questionnaires sent to each organisation was according to the workgroup size or, in the case of organisation C, the number of people being trained over an eight week time period. Ten questionnaires were sent to A. Eight of the first questionnaire (80%) and zero of the second were returned. Sixty questionnaires were sent to B. Twenty-seven (45%) of the second questionnaire only were returned. Thirty questionnaires were sent to C. Twenty-four (80%) of the first questionnaire were returned and ten (33%) of the second questionnaire. This highlights the difficulty of testing in the messy real-world environment especially when a longitudinal study is required.

The aim of this stage was to test for internal consistency of the questionnaire and overall construct validation by correlation of the new questionnaire with one which had been tried and tested. This would give some confidence for use of the questionnaire in a case study where the processes which caused the responses could be closely observed. Although some background information had been gathered from the organisations this was to place the responses to the questionnaire in some form of context and was not sufficiently in-depth to be used on its own. Nor were follow-up interviews possible.

The results of the questionnaire validation suggested that it was internally consistent. However, the overall construct validation was inconclusive due to the small number of usable returned questionnaires. Even so, it was decided that the results were sufficiently encouraging for the questionnaire to be used in conjunction with interview and observation in a case study so that the results of each could be compared with one another and with the proposed model of User Engagement.

The two packages (MM and PM) which did not achieve a satisfactory score from the initial study (descibed under (a) above) were selected for replacement. The replacements had to be planned within the overall corporate plan of the organisation and took place over a period of three years. The software for the larger workgroup (MM) was replaced first.

c) Case Study 1- replacing the MM software

As a result of the findings from the first study it was agreed with management in the user departments that an effort would be made to increase the involvement of the end-users of the software in its specification and selection. This was done by the IT department asking for their opinions both formally and informally (requirements gathering stage), keeping them informed of what was happening and asking them to cast a vote for their preferred option to replace the existing software (selection stage).

The questionnaire being developed (ITU) was issued after a decision had been reached on the preferred package. The plan had been to issue the satisfaction questionnaire (SUMI) a few months later but it was actually a year later before it was finally issued. Additional data was gathered via semi-structured interview and general discussions with users.

The results of the ITU and SUMI questionnaires were correlated. The result was a negative correlation suggesting that the ITU was indeed measuring User Engagement. Interview and discussion re-enforced the results of the ITU and, thereby, the model of User Engagement.

d) Case Study 2 - replacing the PM software

As a result of the findings from the first replacement package a slight change was made to the steps taken in the selection of a replacement for the second package: People were not only involved in the specification and selection of a replacement package, they were also asked whether they felt that a change to working practices was needed. It was possible to pose such a question because of the small size of the user group.

The ITU questionnaire was issued to this user group after initial training as there had been a gap of several months between selection and purchase due to cost negotiation. It was planned to issue the SUMI questionnaire some months later. However, due to redundancy and reorganisation the software was still not being used seven months later and there was no guarantee that the people who had taken part in the exercise would be using it in the future. Some data was gathered by general discussion and this lent support to the model and the results of the ITU questionnaire.

e) Comparison of results

The questionnaire results from each set of replacement projects were compared with the data gathered by discussion with the users and with each other. The results were also compared with the satisfaction results from the original packages and with the User Engagement model. From these interpretations were made and conclusions drawn.

4.6.5.1 The role of the author as participator and observer

During both the case studies, the author played an active role in ensuring that the end-users of the software were involved first in contributing to the requirements of a new system and then in selecting the one to be used. In the case of the requirements gathering, the author was also the member of the IT department who was principally tasked with gathering those requirements. This dual role of the author as both participant and observer in the case studies might be seen as problematic. However, it sits well within the social research tradition of HCI. Whilst the behavioural tradition follows the classical scientific method and seeks to achieve results which are "as objective, replicable, and general as possible,"(Sanderson & Fisher, 1994, p.271), the social research tradition is concerned with understanding as well as observing. This being the case, the participation of the observer may be said to be of benefit because the aim is to "better understand situations from the domain practitioner's point of view, rather than to impose abstractions from an outsider's point of view" (Sanderson and Fisher,1994, p.288-289). In the research described, the author, as a senior member of the IT department whose principal role is to assist user departments in assessing their IT requirements and to source and purchase suitable software packages in conjunction with the IT Manager, was not a researcher taking on the role of domain practitioner but was rather a domain practitioner taking on the role of researcher.

However, this was not action research to the degree that is advocated by Checkland and Scholes (1990) whereby complex systems are developed by the participant researcher experiencing and learning from that very process of development. In the two case studies described, the role of the participant researcher was to observe and understand the situations which set the context into which a new system was introduced, and the nature and degree of participation was different during different phases of the exercise, as described below.

• The requirements gathering stage: the author as a necessary member of the process

The format of the requirements phase followed the standard IT procedure whereby people were questioned as to their problems with the present software and asked what facilities they would like to see in a new system. The interviews and conversations which took place were exactly like those that took place for other software that had been introduced. Nor was there anything extraordinary about the interviews or follow-up interviews pre and post the SUMI questionnaires. Finding out how well or otherwise a software package was performing was standard IT procedure. Whilst the SUMI questionnaires were novel, the fact that they were followed up due to the poor results was not viewed with surprise. The follow-up interviews themselves were guided by specific questions arising from the SUMI, however, the discussions were allowed to be wide-ranging and were not restricted by the author so that questions arising from other software in use were also covered. In summary, from the point of view of the users there was no difference between the activity of the author as researcher and that of the author as member of the IT department during this phase.

• Ensuring people were informed: the author as proactive participant in the process

Keeping people informed of what was happening took the form of brief, written communications on what had happened (e.g. end of requirements phase) and what would happen next. In the case of the MM2 study, people were told about the other package that had been considered. These communications from the IT department were the only information on progress that the MM2 users received and it is known that it was appreciated. The updates sent to the PM2 users were supplemented by the principal player in that group who gave verbal reports. The communications themselves were factual and people were told that they were welcome to ask for further information if they wished. The aim of the author was to ensure that people were involved by keeping them informed. From the point of view of the users this was a passive involvement in that they were not demanding information but were being offered it and the opportunity, rather than active encouragement, to ask for more.

• The selection stage: the author as facilitator

The involving of the users in the selection process was an alteration from previous practice for those users in both case studies. However, the precedent had already been set in the organisation with the IM software. In the instance of the MM2 case study, the author took a background role once the sessions had been organised. The suppliers of the software led the sessions by demonstrating what the software was capable of and prompting for questions. The author was occasionally referred to in her capacity as a member of the IT department, if there was a question about hardware or training but was otherwise ignored. A more active role was taken in the PM2 case study. Whilst the supplier overviews took the same format as that for the MM2 software, with the author very much in the background, the second session, when the results of the reference site visits were reported, was presented by the author. This inevitably meant that the author drove the session and, indeed, asked the question about the change of working practice. However, having sparked the discussion the author was again able to fade into the background. Thus, in the selection process for both the MM2 and PM2 software, the author enabled discussion to take place but did not participate in it. The opportunity for involvement was given to others and the role of the author was that of facilitator rather than that of active participant.

In conclusion, whilst it could not be said that the participation of the author in the replacement of the two packages did not have any effect on that process, the nature of her role as practitioner meant that influence was limited to facilitating the involvement of the users in certain parts of the process.

4.6.6 Difficulties encountered with the research approach

It is acknowledged within the social research tradition that completely objective observation is not possible (Sanderson & Fisher, 1994). Although it might be argued that the position of the author in the organisation (being a participant as well as observer) might lead to an interpretational bias, this must be balanced by the fact that it also enabled the author to be fully aware of the interactions between the users in the different workgroups and their managers, whilst not being a part of those interactions. Indeed people were quite open, even to putting their names on the questionnaires, in a way which they may not have been to an outsider due to loyalty to the organisation. Thus a certain amount of objectivity combined with "insider" knowledge was possible.

There is a danger that people's awareness of being part of a research project can alter their behaviour (O'Neill, 1998). However, the user departments involved in the two case studies were familiar with the author as practitioner and were not made formally aware that the replacement of the software packages was being used as part of a research exercise, although research was mentioned in the written request to people to complete the ITU. The fact that any awareness that research was being undertaken was minor combined with the familiarity of the users with the author and with the already standard procedures that were followed, would seem to negate any effect that awareness may have had.

Validating the ITU questionnaire was difficult and not fully achieved. People are very pressured in the workplace and any additional work is not generally welcomed. In this instance the issuing of the ITU questionnaire is required at a time which is very fraught for those people who are responsible for software purchase. In the case of one government body approached for assistance there even seemed to be concern at what the questionnaires might reveal with the result that the organisation did not participate.

The need for a longitudinal study was also problematic. In Case Study 1 the time span between the issuing of the ITU questionnaire and the SUMI was longer than planned due to illness. Fortunately, the workplace in the organisation was relatively stable in terms of staff mobility and the reason for the lower number of respondents to the SUMI questionnaire was mainly because of the lower number of people actually using the software than because of staff change round. The start of Case Study 2 and, therefore, its completion, was also unavoidably delayed. However, the stability that was present during Case Study 1 was disappearing and a major re-organisation and redundancy situation took place in the workgroup concerned in Case Study 2 just after issuing the ITU questionnaire. This naturally caused some disruption with the result that the SUMI questionnaire was not issued. Such occurrences make life more difficult for the researcher but they reflect the context within which most software is used.

The next chapters describe the above steps in detail. These are shown schematically in figure 4.5 below. This represents what actually happened. The original plan would have seen completion of the PM study in Year 4.

Schematic representation of research steps

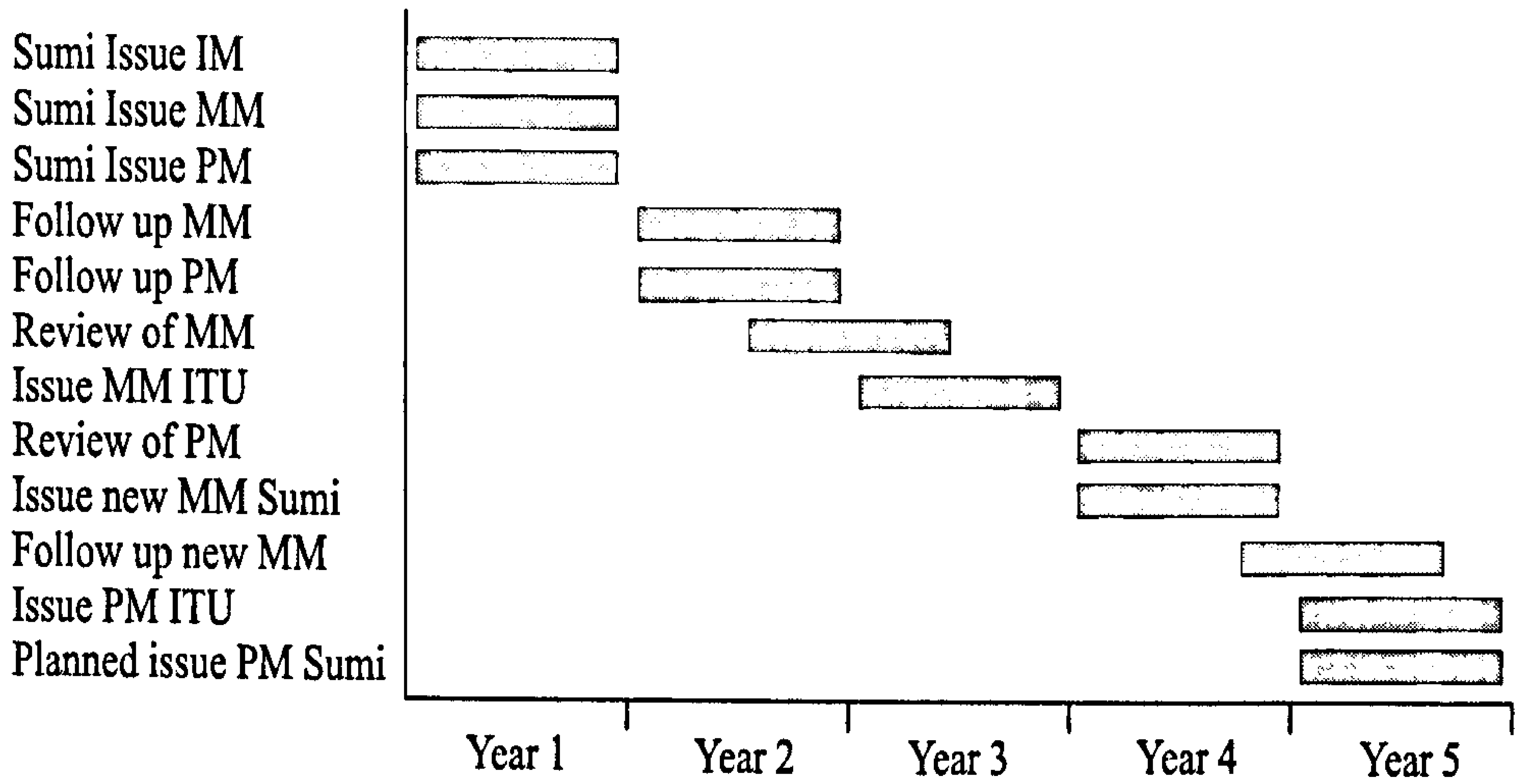


Figure 4.5

Chapter 5

Testing the User Engagement Model

Various factors have been suggested in the literature as having an effect on user satisfaction with software. However, the literature has concerned itself mainly with software which is designed and developed in-house. Most small to medium sized organisations do not develop in-house and purchase commercial off-the-shelf (COTS) software to meet their needs. It is the purpose of this research to determine which factors affect user satisfaction with COTS software. However, account has also been taken of the fact that such organisations do not have the resources to investigate and analyse all the potential influencing factors. It is the contention of this research that it is the perceptions of the users about the software and of the beliefs of those around them, plus their involvement in the process of selecting the software, which will dispose them favourably or otherwise with a package. Other factors may have an effect on those perceptions but it is the perceptions themselves which are important in determining whether people have engaged with the software. Engaged is used in the sense that people have a favourable attitude or commitment towards the software. The idea that it is people's perceptions, however they arose, that are important is similar to the Technology Acceptance Model (TAM) (Davis et al., 1989) except that TAM assumes that usage of the software is voluntary and, therefore, does not take account of the perceived involvement of the user nor of the subjective norm.

The first stage in testing the model proposed in Chapter 4 is to assess the current level of satisfaction with packages already in use. A working definition of satisfaction was given in Chapter three for the literature investigation. In order to test the model of User Engagement proposed in Chapter four, it is necessary to be able to measure the satisfaction resulting from use of a software package. The first section of this chapter describes the selection of a tool with which to carry out that measurement. Section 5.2 then describes the organisation in which the tool will be used. The process of selecting three different software packages and the later administration of the tool to measure satisfaction is described in sections 5.3 to 5.6. Similarities or differences in the results are then discussed and are compared with the hypotheses proposed in the model in section 5.7. As differences could be purely due to the characteristics of the software, an independent evaluation of the software is carried out and the results are compared with the level of satisfaction reported. This is described in section 5.8. Section 5.9 then summarises the "current situation" picture and describes the next steps taken.

5.1 Measuring Satisfaction

In the studies discussed, where the term "acceptance" has been used, it is in a context which implies that people are satisfied with the software. The degree of acceptance is often measured by the amount of use of the software. Martinko et al. (1996) suggest that rejection is shown by lack of use and Hirschheim and Newman (1988) use the same basis. Rice and Tyler (1995) take usage extent as an indication of acceptance. However, a distinction should be drawn between individual use and use within an organisation. An individual whose work is autonomous may choose not to use a computer system which is felt not to be satisfactory. In the case of software which is used throughout an organisation the individual may have little choice whether to use the software or not. The requirements of the organisation may override those of the individual. Rice and Tyler (1995) note the fact that the more conservative of the organisations was found to have a high reported intentional usage of the v-mail system but that this could be due to the likelihood that the organisation was more procedurally oriented. Intentional usage, therefore, may not be the same as voluntary usage. The fact that software is used in an organisation does not necessarily mean that it has been totally accepted. It may be used because company procedures dictate that it must be. However, it is unlikely that it will be used to its best advantage if the people using it are not happy with it. Whilst it is not sufficient that people are happy to use a piece of software (they may be happy to use games software but that will not lead to high productivity), it is generally held that people who are happy in the work they are doing will be more productive.

If usage on its own is not necessarily a good indicator of satisfaction and, therefore, the measurement of usage alone may not be used as a measurement of satisfaction, there is the problem of how satisfaction may be measured. In measuring satisfaction, the focus shifts from the capabilities of the software to how people feel about using it. This means that an evaluation process which determines how well software meets specific guidelines or standards is not measuring the right thing, even if the achievement of those standards should help to ensure user satisfaction. It is the *perceptions* of the people using the software which have to be measured.

Once these perceptions have been measured, then is the time to try to discover why the results are as they are. Perceptions, unlike usage, are not easily quantifiable. Interviews may be used to elicit information from people and may be very revealing as the interviewer is able to discover more about why an individual has responded in a certain way. However, even when a formal, structured interview is used (which may lose some of the benefits of an interview by restricting the interviewee's responses), it is no simple matter to collate the information from several interviews in a standard manner such that it may be presented as a formal metric. The questionnaire approach helps to overcome this as it facilitates the production of a numeric measurement. This ability to formalise a measurement is necessary if results are to be compared or to see if the presence or absence of a factor has an effect on the end result. As Preece et al. (1994) say "Once something can be measured numerically, you move away from the world of opinion and intuition" (p.517). However, questionnaires themselves are not without problems. First of all, the people who respond to them may do so because they have particularly strong feelings which are not representative. Alternatively, some people may just tick the boxes without any thought for what they are being asked. If the individuals who do respond remain anonymous, it is impossible to check what they have understood by the questions and whether their responses do indeed reflect their true feelings. Finally, the length of the questionnaire may influence how and whether people will complete it. It needs to be long enough to extract the necessary information but not so long that people feel daunted by it (Kirakowski, 1995b). However, the questionnaire approach does give a means of measuring and comparing responses.

It was decided that a combination of interview and questionnaire would help to overcome some of these difficulties and that the results of one could be checked with the results of the other to see if they supported each other. In this research, the personal knowledge the author has of the people involved is also of some benefit, although that is not without its own danger of reducing objectivity. However, the combination of interview, questionnaire and personal knowledge combine to form a triangle of supporting data.

Various questionnaires to measure satisfaction have been developed in the past. Those for which there is some solid support in the literature are now discussed and the reasons given for choosing the one selected for this research are given.

Ives, Olson and Baroudi (1983) define user information satisfaction (UIS) as "the extent to which users believe the information system available to them meets their information requirements" (p.785). They further assert that the measurement of this satisfaction acts as a surrogate for measuring changes in organisational effectiveness as a result of the introduction of a computer system. Their review of the literature highlights the different ways in which satisfaction has been measured, from a single item construct to one which is broken down into either measures of user perceptions of the output of a system, or both measures of the perceptions of the surrounding support for the system as evidenced by training, documentation etc., and items relating to the content of the system. The latter multi-item construct assesses satisfaction with the overall quality of the service provided by the DP department as well as with the output of the system. Ives et al.(1983) selected a questionnaire for which there was some empirical support and which measured both the product of the information system and the general system services, for further refinement. They found the original questionnaire to have good reliability as measured by Cronbach's alpha; content validity, as measured by correlating each of the item scales with a separate four item User Information Satisfaction (UIS)[8] measure (these four items consist of how well the DP group meets the needs of the user's area of work and the needs of users generally and how effective and efficient the DP group is); predictive validity, by correlating the total score of the questionnaire with the total score of the four item UIS measure; and construct validity by virtue of replicating the findings of the original study (although Ives et al. point out that further testing is required). The shorter form questionnaire which they developed from the basic structure of the original has since been used in other research. For example, Ramamurthy et al. (1992) used an amended version of the questionnaire to assess user satisfaction.

In the factors such as "reliability of output" measured by the questionnaire, the perceived usefulness of that output is being measured. If the output is not reliable, accurate, timely etc. then its usefulness is diminished. Factors such as "convenience of access" and "relationship with EDP staff" may be related to perceived ease of use as they suggest how readily an individual is able to obtain the information he requires and "User's participation" measures perceived involvement. However, these factors, which are measured to determine satisfaction, require some consideration. They all relate to the result of using a computer system in the format of output reports or to factors concerning the data processing (DP) department. This is reflected in the scales measured in the questionnaire such as "Reliability of output"; "Accuracy of output";

[8] The UIS measure was created in an earlier study by the same authors

"Communication with EDP staff"(p.793). Ramamurthy et al. (1992) removed the questions relating specifically to DP functions, as their study was undertaken under laboratory conditions, but they were still measuring the output of a system to assess satisfaction. The four item UIS measure used by Ives et al.(1983) and against which they correlate the results of the detailed questionnaire, also measures the perception of the performance of the DP department in meeting the needs of the users. This focus on the output of a system reflects the context in which the questionnaire was developed. Firstly, in the early 1980s, use of mainframes and minis with centralised DP departments was the norm. The idea that the actual interaction between an individual and a system should be considered was only just taking hold. There were people who keyed data into a computer system via a terminal and there were people who made use of the information produced as a result of that data collection. This latter group, the managers and decision makers, are the people whose satisfaction is measured as a surrogate for improved efficiency. This is not a measure of satisfaction with interaction with the computer system. It is a measure of the satisfaction with the output of that system and the interaction with the DP department. The context of today's use means that there is much more of a focus on the interaction of the individual with the computer system.

Similarly, the questionnaire used by Rushinek and Rushinek (1986) is designed to determine people's satisfaction with computer systems in general, not with the use of a particular piece of software. This questionnaire uses a single item construct (of which Ives et al.(1983) question the validity) to determine satisfaction. The responses to 17 other questions are correlated against the question on overall satisfaction to determine their importance relative to each other and reasons for the results are suggested. The survey was sent out to a large number of people from the mailing list of a computer magazine and asked, amongst other things, the percent of computers which were mainframes, micros or minis used in the respondent's workplace. The responses they received covered 179 computer systems between them. Whilst Rushinek and Rushinek (1986) claim that the "microcomputer..induces satisfaction and "user friendliness" by providing users with a sense of control and independence.." (p.597) there is nothing in the questionnaire to indicate that users do feel control and independence. The questionnaire concentrates on the availability, cost and compatibility of the systems rather than the interaction with them. One question does ask to what extent the system meets the user's expectations and another asks about the efficiency and effectiveness of the database language, but what the expectations were, or how efficiency and effectiveness are defined, is not described. Although the coverage in terms of users is greater than that of Ives et al.(1983) and the computer systems range from single to multi user, the Rushinek and Rushinek questionnaire serves only to show a difference in overall satisfaction rating between people using different systems. It does not measure how satisfied a user feels with the actual interaction with a system, nor does it investigate the surrounding context of use.

Igbaria and Nachman (1990) define satisfaction as "the affective attitude toward a specific computer application by any hands-on user of microcomputers or computers" (p.74). They make use of the Ives et al. (1983) short form measure of satisfaction as tested and reported by Baroudi and Orlikowski (1988). This questionnaire measures the three factors of information product, electronic data processing staff and services, and knowledge and involvement. The sum of these three being an indication of level of satisfaction. The aim of the study by Igbaria and Nachman (1990) was to correlate the various factors of leadership style, hardware and software accessibility, computer anxiety, attitude towards end user computing (EUC), system utilisation, computer background and demographic indicators, with overall satisfaction as measured by the Baroudi and Orlikowski questionnaire and thus to show the effects of the various factors on satisfaction. Leadership style was assessed according to whether it was authoritarian or supportive. This would form part of the subjective norm (SN) of an individual and a significant correlation was found between leadership style and overall satisfaction. The more supportive style correlated with increased satisfaction and the less supportive correlated with reduced satisfaction. Attitude was also found to have a significant correlation with satisfaction. The Attitude construct corresponds to the ex ante attitude of Ramamurthy et al.(1992) and measures pre-conceptions about EUC. This, too, combined with computer anxiety, could be said to form part of the SN of an individual as the attitude is not formed purely on the basis of experience with the current system. The accessibility of the hardware and software should rarely be a problem today when they are provided as part of the tools to perform the job. However, the construct may be related to that of ease of use. If something is readily accessible when needed, then it may be perceived as easier to use than when it is not so readily accessible.

Igbaria and Nachmann (1990) also found a significant correlation between utilisation and satisfaction. However, utilisation was measured not by the frequency of use of one particular system but by the number of different systems used. Whilst this might show that people who are satisfied with one system will be more inclined to use another, it might also be that other tools are not available to them and they use the software because it is what is provided and not because they are satisfied with its use. No significant correlation was found between demographic variables and satisfaction.

The questionnaire tested by Baroudi and Orlikowski (1988) and used by Igbaria and Nachmann (1990) measures satisfaction with the three areas of DP staff and services, information product, and user knowledge and involvement, which combine to give a rating of overall satisfaction. These are naturally very similar to the original questionnaire devised by Ives et al. (1983) described above, from which the short form measure was derived. Although Igbaria and Nachman (1990) use it with reference to EUC, which includes use of PCs as well as central systems, satisfaction is still being measured partly by reference to a central DP function. They are, therefore, suggesting that an individual's satisfaction with the system could be affected by their relationship with the DP department. This could be because the study aims to include all forms of EUC (i.e. PC and central computer) but it does suggest that the satisfaction being measured extends beyond satisfaction with the computer application itself. In fact, when the areas of the Baroudi and Orlikowski measure are considered, they do not focus on the software application at all: The relationship with DP staff and services, shows a dependence on others for the provision of information or assistance, which relates to the days of a centralised DP function, where people had no independent access to data. Even where an application and data reside on a central computer, it is rare today that people are not able to create their own enquiries and reports without reference to the DP department. The change in name from Data Processing to Information Services (IS) or Information Technology (IT) is a reflection of the changing role of this group. The second area, that of information product, whilst it can be said to correspond to the perception of usefulness, narrows the scope of that usefulness. Although a system is only really useful due to the information it supplies, access to and manipulation of the data which forms the information may be via direct interaction, which would not be perceived as output from the system. Output tends to suggest something of a more static nature, such as a report. This again is a reflection of the circumstances of the period when the satisfaction measure was created. Finally, there are the two elements of user knowledge and involvement. As has been discussed, the evidence for the influence of user knowledge is mixed (Ramamurthy et al.,1992; Zmud, 1979) and not consistently defined. Whilst it surely has an influence on satisfaction, it is questionable as to whether its inclusion in the Baroudi and Orlikowski measure (1988) in the form of the question "Users' understanding of systems" is justified. Is the user responding about his understanding of systems in general or about his specific knowledge of one system? Does "understanding" mean that the user understands the applicability of the system or does it refer to his ability to use the system? Baroudi and Orlikowski (1988) state that the short form measure is a "way to obtain information about overall user satisfaction with information services" (p.55). This means that the satisfaction that is being measured is that of a general satisfaction with the computer systems provision within an organisation rather than satisfaction with a specific application. This general perception of the computer systems in an organisation may, of course, influence the reception of any new introduction, such as would be the case with the introduction of a packaged system. However, what is required for this research is a measure of satisfaction with the packaged application itself. This measure of satisfaction needs to be broader than a simple measure of satisfaction with the interface but more confined than a measure of satisfaction with the totality of the computer systems available.

The Questionnaire for User Interaction Satisfaction (QUIS) was developed in 1987/88 by the Human Computer Interaction Laboratory at the University of Maryland, College Park, USA and concentrates on satisfaction with a specific application. Since its development, it has continued to be revised and improved upon. It has been licensed to commercial/industrial users as well as to academics. The paper format of version 5.5 of the questionnaire was the one considered for use. The questionnaire consists of 7 sections. The first three ask for information on the user's current usage, previous experience, and overall perception of the system. The next four sections cover different aspects of the system - Screen, Terminology and System Information, Learning, and System Capabilities. A short and long version of the questionnaire is available and it is possible to tailor it by removing questions which may not be applicable to the system to be evaluated. The questionnaire may be administered in paper or computer based format. A license fee has to be paid for the use of QUIS, the fee increasing according to the number of locations at which it is to be used. This commercial aspect of its use would suggest a product which has been well tested in the field. However, apart from the first general section, it is very focused on aspects of the system interface. The underlying principle is that user perceptions of the interface will lead to satisfaction or dissatisfaction and that the questionnaire can assist in identifying particular problem areas (Chin, Diehl & Norman, 1988). It is for the distributor of the questionnaire to analyse and interpret the responses and there is no given baseline against which to perform this (Harper & Norman, 1993). The questionnaire was developed and tested in the USA. At time of consideration, there was not a UK version available. As English is a common language between the two nations, this may not be thought to be necessary. However, part of the validation of a questionnaire is to ensure that the meaning of the questions is fully understood and interpreted in the same way by the respondents and there are some differences between US and UK English which may affect this. A student license for QUIS was obtained but it was not used as the Software Usability Measurement Inventory (SUMI) was finally selected as the questionnaire for this research.

SUMI was developed by the Human Factors Research Group, University College Cork, Ireland as part of the Metrics for Usability Standards in Computing (MUSiC, ESPRIT Project P5429) and is now used to assess the user satisfaction aspect of usability in the MUSiC suite. In developing SUMI the HFRG took account of other questionnaires measuring satisfaction (Kirakowski,1995a). One of these was the Computer User Satisfaction Inventory (CUSI) which had been developed at about the same time as QUIS. It measures the degree to which users express a liking for the system and the degree to which they feel supported by it. A close correlation between the Affect of CUSI and the overall satisfaction rating of QUIS is reported. The original pool of items used in SUMI to measure satisfaction was over 150. Further refinement reduced the number first to 75 and thence to 50. Testing in the field reduced the Lickert type scale from 5 points to a 3 point scale of Agree, Undecided, Disagree. The 50 items grouped into 5 scales which are:

Efficiency - the user feels that work is being done without any extra mental workload being imposed by the software

Affect - how the user feels as a result of interacting with the software; good, warm, happy or the opposite

Helpfulness - the user's perception about the helpfulness or otherwise of messages and warnings given by the software

Control - the user feels that the software responds in a normal and consistent manner to input and commands. It isn't difficult to understand how to make it work.

Learnability - the user feels that it is relatively straightforward to become familiar with the software and accompanying manuals, tutorials etc. are easy to follow and instructive.

Overall satisfaction is measured by the Global scale, which is calculated from within the 5 sub categories. The Global scale should be seen as the measure of satisfaction, with each of the 5 sub scales measuring specific areas of that satisfaction. The questionnaire itself was developed using the method of item generation, item allocation, factor analysis, item revision, standardisation and validity and reliability testing (SUMI Handbook, section 8) and has been tested in the commercial and academic sectors. In addition to the development of the questionnaire a standardised database of the results from a large number of different users to a wide range of products has been created. This means that SUMI enables a satisfaction rating of a product against a standard base to be made. A package scoring between 40 and 60 on the Global scale is said to be within the satisfactory range, with 50 being the "average" point. Thus a score above 50 is above average and below 50, below average. A score below 40 is a definite indication of problem software. The professional edition of SUMI (for which there is a one off fee, additional payments being made only for additional questionnaires) consists of a paper based questionnaire, the SUMI handbook, manual scoring tools and a computer program which may be used for initial data analysis. This latter produces a table and graph of each of the 5 subscales and the Global score; a table of scores on each scale for each user; and an Item Consensual Analysis (ICA). The ICA is particularly useful as it calculates the number of responses against Agree, Undecided and Disagree for a question and the statistically expected number based on the standardisation data for that question. This means that it is possible to highlight specific areas of good or poor performance.

The Efficiency and Control subscales may be used as an indication of usefulness although they are not specified as exactly measuring this. Efficiency measures the mental effort required by the user in order to achieve a task and is, therefore, partly a measure of how well the software matches the task to be performed. Control measures the confidence the user feels in using the software, which is also an indication of how the software assists in the accomplishment of that task. The Helpfulness and Learnability scales may be used as an indication of ease of use. If the software itself indicates the sequence of steps to be followed to perform a task and if it is easy to remember what those steps are, it will be easier to use. It may, therefore, be said to be measuring two of the factors in the proposed model i.e. perceived usefulness and perceived ease of use.

However, SUMI is designed to be completed by someone after they have been using the software and is intended to give an overall satisfaction rating. Although the scales may be considered individually, they should not be taken apart from the context of the whole. The SUMI handbook suggests that a user group should not be smaller than ten people, although SUMI has been used successfully with smaller groups. The developers stress the fact that the evaluation is specific to the product, users, task and the setting in which it is used. In other words that the perceived usability of the software leading to satisfaction is not an attribute of the software alone. As well as the administration of the questionnaire and analysis of the results, it is recommended that there be follow-up interviews with a sample of users.

The present study is not intended to compare the strengths of one questionnaire over another but to determine and then compare the satisfaction felt by different groups of users with the software purchased for them. This means that one questionnaire must be selected and used for all packages. The SUMI questionnaire was chosen as the means of determining the user satisfaction largely because it had been developed and tested as part of a European programme (ESPRIT) and the software in question and the organisation using it are all Europe-based i.e. each of the software packages had been developed in Europe. The computer program provided, which performs the initial data analysis, also ensures consistency of that analysis and that the satisfaction rating is as a result of comparison with a standard base.

However, it was felt that the initial questions asked in QUIS relating to the users' prior experience might have an effect on their perception of software and a request for this information was made during the interviews.

The following section describes the process of selection of three commercially available software applications which was undertaken over a period of years by one organisation. The method of administering the SUMI questionnaire to the different groups of people using the software is described and the results are discussed and compared with reference to the hypotheses stated.

5.2 The Organisation

The organisation chosen for study is a small company of fewer than 1,000 employees. It is not funded by Government but is structured along the lines of Public Sector organisations. The Organisational Type is conservative and service-centred. Its function is to provide and maintain equipment across England, Wales and the Channel Isles. It is divided into six geographically separate locations. One location is mainly concerned with the legal and administrative aspects of the organisation (HQ), another develops new equipment and carries out modifications to existing equipment (PM) and the third acts as a central warehouse of spares (CS) supplying the remaining three, all of which are primarily concerned with maintaining the equipment, using stock held at the local level (MS). The general administrative functions required for any business, including the provision of IT, are situated at one of these locations. This location, which is made up of several sites spread over a small area, also houses the Central Maintenance (CM) group. The three locations which have their own stores holdings are known as maintenance depots. The maintenance depots are semi-autonomous but answerable to the CM group for how their budget is managed and the work performed.

All six locations in the organisation to be studied are linked together over a Wide Area Network (WAN) with a central mini system and Local Area Networks (LANs) at each of the sites. Software is held on the mini system, fileservers and PCs. The size of the organisation means that in-house development of software applications is not readily undertaken. Figure 5.1 gives an overview of the organisation and how the sites are linked over the WAN.

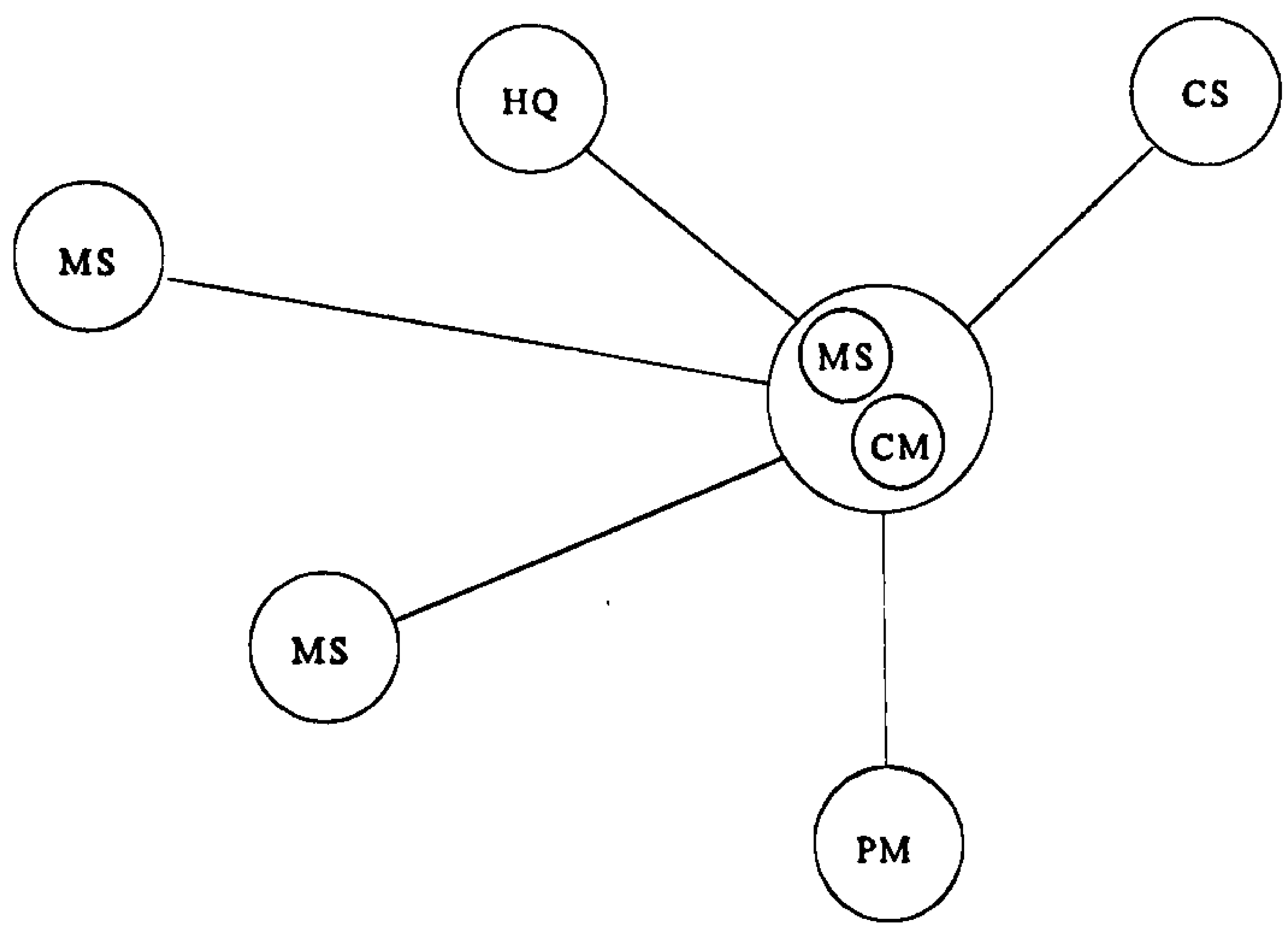


Figure 5.1

The drawback to using such an organisation for study is that it is classed as being part of the Public Sector and its "culture" may, therefore, be said to be unlike that of commercial organisations. However, it is totally self-funded and its structure is not unlike that of many in the Private Sector. The benefit of the study itself is that the researcher works within the organisation and is completely familiar with the process of software selection through to its use in a live context. The study does not rely on taking sample users or testing away from the true environment. A comparison may be drawn with the Cafe of Eve project (an acronym for Controlled Adaptive Flexible Experimental Office of the Future in an Ecologically Valid Environment) which made the case for developing and testing software design within a working organisation so that usability could be evaluated in context (Gale & Christie, 1987). However, in this instance, the users involved have not been aware of being involved in a research project. A possible drawback to this approach is the potential for bias on behalf of the researcher by virtue of being actively involved in the organisation.

The different approaches in the process of software selection described in this chapter were not taken as part of the research project but were the standard method used. They are described as they are used in the interpretation of the SUMI results. The issuing of the SUMI questionnaires and the interviews and discussions were undertaken as part of the research.

5.3 Software selection

Three different software packages were selected for administration of the SUMI questionnaire. Of these, one (**IM**), had recently been introduced and ran on the central mini-system. Access was via PC in terminal emulation mode. The second (**MM**), ran on fileservers with PC clients using DOS and a menu-type interface, and was due for review. The third (**PM**), was also due for review and ran on PCs under the Windows operating system with data held centrally on a fileserver. Between them, these three gave a sample of operating systems, interface types and user groups.

Each of the packages had been selected and introduced outwith this research. However, the author was involved in the selection process of all but one (for which the process followed is still known). In selecting a software package a fairly standard approach is followed and this is outlined below. Although this process does not directly form part of the current research, the steps are important when the SUMI results are considered.

a) The prime purpose of the software purchase is established - e.g. is the intention to improve existing work practices by computerisation; is an existing computerised system felt to be deficient in what is needed etc..

b) Once it has been established that a new software purchase is required, an investigation into the current methods of working is undertaken. The prime aim is to ascertain the functionality that the software must provide and any that would be useful but not essential.

c) The information gathered is written up as a formal requirements specification.

d) During the information gathering period, details of commercially available software on the subject (e.g. payroll) is obtained and demonstrations may be viewed (generally at the supplier's site or at an exhibition).

e) The detail in the requirements specification is summarised, agreed to at management level, and a weighting given to each item.

f) Suppliers of software which is likely to provide the required functionality, are written to, sent a copy of the specification and asked to detail how their product meets the requirements, if they wish to be considered.

g) The responses are matched against the summary table and each item is given a score.

h) All item scores for a product are totalled to give an overall mark. The suppliers of the top three scores are then invited to visit the organisation and give a demonstration of the software.

i) The final decision on which product to purchase, is made after the demonstration and, at this stage, comparative cost may be weighed against functionality.

Steps a) to h) are undertaken by the IT department, with reference to managers and users at different stages. The final choice is made by the managers of the user department.

Each of the packages is now described in turn. The full background to how the software was selected is given even though it did not form part of the research project as this is needed when considering the results of the SUMI questionnaire.

Following the return of the SUMI questionnaires for each package a Profile Analysis (the table of overall scores for each scale), User Scale (the individual user scores) and Item Consensual Analysis (comparison between the results and what would be expected from the standardised database) was produced. An example of the questions from the SUMI is given in Appendix 1.

5.4 The IM software

5.4.1 The IM function supported by the software

The IM function is divided into two areas: The purchase of items for stock and the recording and issue of those items as required. Four separate locations hold and issue stock. These locations are physically remote from each other. Three of the locations service their Maintenance depot. These are shown as MS on the organisational diagram (figure 5.1). The fourth location issues stock to those locations which do not have their own IM service and is also the sole holder of clothing requirements which are issued directly to individuals. It also performs the stock purchasing function on behalf of all locations. This location is known as the Central Stores (CS). Items which are common to all locations will be stocked by the CS and issued, on request, to the other locations. Items which are special to the maintenance requirements of one or two locations will still be purchased by CS (on request by the locations) but are not stocked by CS and may be delivered direct from the supplier to the requesting location. Deliveries of stock items are made from the CS to the other locations on a regular basis (transport of items other than stock is needed between locations and this is also handled by CS). Return of faulty items is handled by the CS as is the matching of invoices against goods received.

The (MS) locations are responsible for ensuring correct stock levels for their area. When more stock is required, a request for a transfer is made on CS. Sometimes more items than are finally needed for a job are requested and issued and are later returned to stock. Maintenance work may result in items being returned which need repairing before they may be returned to stock. Alternatively, they may not be worth the cost of repair in which case they have to be written off and replacements ordered. The process for this requires communication between the MS and CS. If items are to be disposed of, they are transferred to the CS. Requests for goods, actual transfers and the disposal process are all managed via the software. In an emergency, an MS location may transfer items from its stock to another MS location. In this instance communication on whether to transfer is external to the system and will involve the CS as well as to the two Maintenance stores.

- **Background to software purchase**

The software to be evaluated replaced an existing inventory management system for which the purchasing module had not been invoked. Both the software which was replaced and the current software ran on a mini computer, accessed via terminal or PC using terminal emulation. The interface is character based and menu driven. All users of the replacement system were familiar with the previous system and, therefore, with the use of terminals or PCs using terminal emulation. The previous system had been in place for 6 years and general dissatisfaction had been expressed at management level regarding the difficulty in extracting information and at user level which gave no specific problem just a general "it's not friendly".

- **Selection of the software**

The standard steps previously outlined were followed. A more detailed description is now given in order to set the context in which the software package was selected.

Semi-structured interviews were conducted at each of the locations where a stores service was provided plus at those sites who requested items from stock but did not have a local stores. People at all levels were interviewed - from those who required management information for accounting or other purposes to those who were primarily concerned with obtaining items from stock in order to perform their job. A total of twenty-five people were interviewed. The interviews with the eight people at management level - a mix between local and central management (all male, with an age range from late thirties to mid fifties) were geared to finding the type of information they required from the system. The interviews with the six general users (representative of the people who requested stock to be issued and all male with an age range of mid thirties to mid forties) were geared to understanding how they felt about the process required to obtain stock. The interviews with eleven actual software users (five male and six female, with an age range between mid-twenties and early forties) were aimed at gathering information on how the task was currently done, what the user understood the aim of the job to be, how they felt it could be improved and the good and bad points of the current system of operation. Some time was also spent in exploring the good and bad points of the existing software. The primary operators from the latter group at each location were also asked to do a walkthrough of the method they used when issuing stock - from receipt of the request through to its fulfilment and also when receiving goods into stock. Examples of all paper documents maintained which related to the whole stock management process were also viewed at each location. Notes were written up following interviews and walkthroughs and sent under confidential cover to the participants with a request that they should make any corrections they felt necessary and return. All notes were treated as confidential.

A requirements specification was drawn up for agreement at management level incorporating the results of the interviews. It was noted that there was a conflict of operation between the CS and the MS in the issuing of items. At the MS the same person who received the request to issue updated the computer system. At the CS, the computer operation was performed in the office and the physical issue performed in the stores warehouse. Some of the dissatisfaction at the local level with the system to be replaced was caused by the system operating in a way more attuned to the operation at the CS. Management felt that it would not be possible to align the way of working at each location and that there would have to be a compromise.

Potential software suppliers were identified by the IT department and invited to respond to the specification drawn up. A list of 39 key points was made and given a weighting factor according to importance (as defined by management). The replies were measured against this list by the IT department. Cost and hardware platform did not feature in the list as the primary aim was to find software suitable for the task and it was felt that the field should not be narrowed at this stage.

The suppliers of the top three scoring packages were invited to give a demonstration of their product, showing the process for a typical request to issue and request on CS to transfer process. They were then free to point out any benefits they felt their software had. Direct users of the system from each location and users of management information were invited to see the demonstrations and to ask any questions they felt had not been addressed. The demonstrations were given over three consecutive days. Before the first demonstration, the attendees were each given three copies of a list of key points and asked to put a score against each of the them at the end of each demonstration. In the event, few of the attendees were prepared to commit themselves at the end of the first demonstration, wishing to see the others first. In the end most of the checklists were not returned.

The vote was finally taken by a show of hands. This showed a split between the users of management information produced by the system and the hands-on users. The latter favoured a package which had various features not listed in the requirements but which the users felt would be useful. However, it was weaker in the aspect of integrating with existing systems. The management users favoured the package which, whilst having fewer additional features, was obviously going to present few integration problems. It was this package that was selected but only after consultation with the direct users and an agreement that the package could be tailored in the future to add features they felt were lacking, provided the cost was not too great compared to the benefit. In fact, the suppliers of this package were asked to give a further demonstration of their software before the final decision was agreed. This was because both the hands-on users and management users felt they had learned how to assess the demonstrations over the three day period and might be misjudging the software because it had been the first to be demonstrated. In the event, some of the features that had led the hands-on users to prefer the alternative software were also available in this package but had not been shown. This did not change their preference for the other software but they did agree to the management decision.

Full training, using familiar data, was given to all hands-on users in advance of commencing a parallel run of the software. Some changes to the software were agreed at the time of training and further minor enhancements requested by the users were made during the four months of parallel running. These changes did not affect the underlying structure of the software but did add to its ease of use by changing field labels on the screen to something more familiar to the users.

The software had been in live use for just over nine months when the SUMI questionnaires were administered. The steps described below were undertaken as part of the research project.

5.4.2 Informal interviews and the administration of SUMI questionnaires

•Informal Interviews

In order to gather some general information on the users' perception of the system, informal, semi-structured interviews, taking the form of a general discussion in the individual's normal environment, were carried out at three of the MS locations before issuing the SUMI questionnaires and at one MS location after the issuing of the SUMI questionnaire but before the returned data had been analysed. Two people (one male and one female) were interviewed at one location and one (male) at the second and third locations. All four had been involved in the information gathering interviews which had taken place before the specification was drawn up. All had been invited to the software demonstrations and two had been present. The four were selected because their work required daily use of the software and they were in regular contact with colleagues in other locations who also had daily use of the software. Although only a small sample, it was felt that they were fairly representative of the "common" view and would be aware of the views held by others. It was not possible at the time to hold such interviews with the people at the CS location although note was made of general comments that were made during other dealings with them.

Brief notes were taken during the time of the informal interviews and fully written up immediately afterwards. The information being sought was:

a) how much previous experience of computer systems the individual had prior to the system in question;

b) the amount of involvement they felt they'd had in the selection process;

c) how long they'd been using the system;

d) how they felt about the implementation process (too fast, slow etc.);

e) whether the right amount of training was given and at the right time;

f) whether the software was useful to them in their job;

g) whether it was easy to use;

h) whether others felt as they did.

The questions were asked in order to encourage people to express their opinion generally as well as to gather specific information. The reason given for the discussion was that the IT department was reviewing the process of selecting and implementing packages.

•Distribution of questionnaires

The SUMI questionnaire was given to users for whom use of the software is a key element in their work, numbering twelve in all. Of these, six (five female and one male) were located at the CS location one of whom had not been included in the investigation interviews prior to the software selection. The remaining six were made up of two people from each of the other Maintenance locations (five male and one female). Two of these had not been included in the initial investigative interviews. SUMI questionnaires were not given to any of the management users as their interest is in the results of the system operation not in its actual use and they did not have daily hands-on use of the software. Similarly for the general users.

Following the guidelines given by the producers of SUMI, each questionnaire was accompanied by a typed note giving a brief explanation of what it was and how to complete it. Where possible, the questionnaire and note were handed personally to the user, with a verbal explanation as to the reasons for it. In one case the user was spoken to by telephone before the questionnaire was forwarded and, in the case of CS, one person was given the verbal explanation and asked to pass on the questionnaires and notes to the others. It was stressed that the questionnaires could be completed anonymously. Eleven of the twelve questionnaires were returned. The twelfth was not returned due to longterm sickness. None was returned anonymously. The questionnaires were handed out and returned over a period of six weeks.

Once all the questionnaires had been returned and it had been confirmed that the twelfth would not be, the data was keyed in as a table. The computer program supplied with SUMI was then used to analyse the data and to produce various reports.

5.4.3 Results of informal interviews and the SUMI questionnaires

- **Informal interview**

a) Experience of computer systems prior to the one in question

All four had been daily users of the previous IM software for the six years of its life. Prior to that, one had limited use of computers in previous employment and two had a home PC. The fourth had no experience prior to the previous software.

b) Involvement in selection process

All said they felt they had been involved in the process. Two expressed the view that the choice of software was biased towards the needs of management, but understood there had to be compromises. Another felt that management knew what was needed and trusted their decision. None felt that more should have been done to involve them.

c)Number of years using the software in question

All four people had been using the system since it was introduced, just over nine months prior to the discussions taking place.

d)Implementation

Two felt that the implementation process had been hard work but that people had been kept well-informed as to what was happening. Another also expressed the view that knowledge of the previous system had helped the process. The fourth was non-committal but did not have any complaints about the process. Rather he regarded it as something that had to be done.

e)Training

Three people felt that the training provided had been very good and sufficient for their needs. The fourth was happy with what had been given but would have liked it to have been a little more in depth in one area.

f) Useful Tool

None could say whether they felt that the software helped them to operate more efficiently but they felt that the method of operation had improved by virtue of its use. However, one did say that further improvements could be made and that, because his way of operating was slightly different from the others, due to different work practices at that location, the software didn't support him to the full extent he would like. Even so, it was an improvement over the previous package.

g)Easy to Use

The feeling was that the software was very straightforward and that it was very clear what steps were required to achieve an end.

h) **Feelings of others**

All reported that their co-workers had similar opinions. Although this could be due to false consensus, this impression was supported by general feedback on the system which had not been formally sought.

Overall, the discussion around the questions suggested that all four people felt that they had been involved in the process of software selection even though two of them had not attended the demonstrations (they had been invited). Even though the discussions were brief and with a small sample of users, a general impression of satisfaction with the whole process of software selection and implementation was given.

- **SUMI questionnaires - summary of results**

The SUMI program was used to produce the following tables:

a) Profile Analysis

b) User Scales

c) Item Consensual Analysis

When the above tables have been generated, it is suggested in the SUMI handbook that any Outliers i.e. users giving particularly high or low scorings are removed and the analysis re-run. In this instance, there was one user who fell into this category. As that user could be eliminated without the sample falling below the minimum ten recommended, the data was removed and the analysis re-run as suggested. Both sets of results are discussed here as the software forms an integral part of the user's work and his views are, therefore, important. The tables for all 11 users are the ones presented within the text below but the discussion itself refers to the results both including and excluding the Outlier. When the data was run the second time, a further Outlier appeared. The fact that the removal of one Outlier flagged another, might suggest a wide spread of views amongst the respondents. However, this is not supported by the spread between the Upper and Lower Confidence intervals. The Outlier from the second run was flagged because of a particularly low score on Learnability. This is the scale which achieves the highest score on both runs (62 and 61 respectively). This, combined with the fact that the removal of the extra Outlier would result in a sample of 9, led to the decision not to rerun. Only items from the Item Consensual Analysis which were asterisked (i.e. between 95% and 99.99% likely to be different from the standard database) are discussed below. Five responses fall into this category from the set of results with 11 users. When the analysis was run with ten users, seven were flagged as different from the standardised database of which four were the same as on the analysis with eleven users.

a) **Profile Analysis**

The Profile Analysis (Global Scales) results are presented in table 5.1 and are discussed first.

On both the analysis with 11 users and the analysis with 10 users the median of the Global Scale indicates that the software is in line or slightly above what would be expected for commercial state of the art software.(The Global scores being 54 and 53 respectively).

Profile Analysis

Scale	UF	Ucl	Medn	Lcl	LF
Global	74	59	54	49	35
Efficiency	87	64	58	52	30
Affect	75	62	57	52	30
Helpfulness	66	58	54	50	45
Control	77	57	52	47	29
Learnability	81	67	62	57	39

Table 5.1

Looking at the Profile Analysis table it can be seen that the Upper and Lower Confidence levels (Ucl, Lcl)bracket an area of 5 points either side of the median. For the sample size of users, this indicates close agreement between them. The widest span between the Ucl and Lcl is on the Efficiency scale, where the span is 6 on either side of the median. Efficiency is also the highest overall score after Learnability at 58. The slightly wider span of scores on this scale suggests that agreement is not quite so solid with regard to Efficiency as it is on the other scales. Efficiency is the SUMI scale which relates to how well the user perceives the software to assist in the job he has to perform and hence to perceived usefulness. It is possible that the slightly wider variation in views on this scale is due to the fact that the software is seeking to compromise between the way in which the CS operates and the way in which the other locations operate.

The lowest variation in scores is on the Helpfulness scale, where the span is 4 points on either side of the median. The implication is that the interface is fairly intuitive. The highest score is on Learnability. This is interesting as the manual provided with the system is rather scant and has been supplemented with user-written guides since the SUMI was distributed. There is no on-line help. This, combined with the results on the Helpfulness scale suggests that the users find the software easy to use.

The Control subscale achieves the lowest score, although it is still above the required global average of 50. The SUMI handbook suggests that the Control subscale should always be considered in tandem with the Efficiency scale. The lower score on Control could help to explain the slightly wider variation on the Efficiency scale as some people may feel that the software does help them to perform their job, but perhaps not always using the method they would wish.

b) **User Scales**

The individual user scores are presented next (Table 5.2). This table gives the individual scores for the Global scale and each of the subscales. Personal knowledge of the work context of each person is very useful when considering these results. The (H) by user 6 shows that the Help score given by this person is rather higher than that given by the others.

Individual User Scores

User	Global	Efficiency	Affect	Helpfulness	Control	Learnability
1	61	68	60	55	61	59
2	48	58	45	52	52	53
3	51	49	39	52	48	67
4	54	52	59	54	57	65
5	48	53	45	53	44	56
6	71	71	63	71	71	70 (H)
7	45	44	57	47	45	39
8	58	62	60	57	41	62
9	42	37	50	48	49	51
10	60	58	56	59	61	70
11	65	71	68	61	59	64

Table 5.2

User 6, the Outlier on the Helpfulness scale, can be seen to also have presented high scores on each of the other scales. This person did not at all like the previous software, although, as one of the interviewees, he had said that use of that software had assisted in being able to learn the new system. It is worth noting that the software is an integral part of his job and any queries or problems about how that job should be performed are not asked of his local line manager and go direct to the senior manager in charge of the stores function. This manager took a keen interest in how people performed their tasks and was always willing to listen to suggestions they had and to discuss why work was to be carried out using one method rather than another.

Four people gave a global score below 50. Two of these are female and two male so there does not seem to be a gender split between those who are very satisfied with the software and those for whom it only just meets their requirements. However, it is noteworthy that three of them are from the CS. The previous software had been completely geared towards their way of working (although they generally disliked it) and the replacement software required a slight change in this, in order to better accommodate the MS locations. The fourth person presenting low scores worked in an MS location where the line management was indifferent to IT in general and, in practice, took precedence over the senior manager in charge of the IM system even though, in theory, this was not supposed to happen.

c) Item Consensual Analysis

This report gives a comparison between actual response and the expected response to each question, according to the standardisation database. The Goodness of Fit between these is summarised using Chi Square and the computer program flags those items where it is most likely (i.e. 95% or above certain) that the actual response differs from the expected. In the first results (i.e. those with user 6 present) 5 items are flagged in this way and, therefore, have been selected for consideration. All other items return a low Chi Square value and, therefore, fall within the expected range. These are not discussed here. The second analysis, excluding user 6, returns 7 items as having values different from the standardisation database. Four of these are the same as the first analysis. Table 5.3 shows the items from the questionnaire which were flagged on each run. They are presented in the table in the same order as shown on the ICA itself, the first being the most likely to be different from the standardised database.

- **Efficiency** Two Items (11 and 6) are flagged from this scale.

Only one person agrees with the statement to Item 11 in both sets of results and the remainder disagree. It is this positive response that leads the Item to be flagged on the ICA. The high number of people who disagree suggests a high level of confidence with the software.

Item 6 is flagged only from the results with 11 people although the Chi Square value is still quite high on the results with ten people. The reason for it being flagged is the high number of people who disagree with the statement and are hence giving a positive response. In fact, only one person is undecided and none agree. The one person who is undecided is the same person who agrees with Item 11 (so there is at least consistency of response!).

Item Consensual Analysis

Item	**11 users**	**10 users**	**Scale**
42 The software has a very attractive presentation	Y	Y	A
8 I find that the help information is not very useful	Y	Y	H
11I sometimes wonder if I am using the right command	Y	Y	E
3 The instructions and prompts are helpful	Y	Y	H
6 I sometimes don't know what to do next	Y	-	E
2 I would recommend this software to my colleagues	-	Y	A
4 The software has at sometime stopped unexpectedly	-	Y	C
39 It is easy to make the s/w do exactly what you want	-	Y	C

Table 5.3

- **Helpfulness** Items 8 and 3 are both from the Helpfulness (H) scale and are flagged in both sets of results.

Item 8 is flagged due to the large number of Undecideds. Referring back to the brief discussions with three of the users, they all felt that the amount and timing of training had been right. Perhaps as a result of this they have not had to refer to the manuals. It should be noted here that the software does not have on-line help screens. Manuals were given to all user groups. However, the tendency in the organisation is for users to telephone the IT Help Desk if they have any queries, in preference to looking at manuals or Help Screens. The number of calls from users of the software to the IT Help Desk was initially quite high but gradually fell over time. Whether as a result of training or there being a different method of obtaining help, this result suggests simply that the help information has not generally been referenced.

Item 3 is flagged because there is total agreement that the instructions and prompts are helpful, whereas the standardised database would expect there to be some Undecideds or Disagrees. Perhaps it is the helpfulness of the instructions and prompts that give confirmation to the user that they are using the right command. This could also help to explain the high number of Undecideds to Item 8 as people do not need to look for help.

• **Affect** Items 42 and 2 are both from the Affect (A) scale. The high Chi Square on Item 42 is evidently due to the large number of people (9 out of 11 and 8 out of 10) who have ticked Undecided. Perhaps because the work performed requires specific procedures to be followed and actively discourages creativity (issuing and recording of stock is subject to strict audit requirements), the attractiveness of presentation of the software is not something that users are too concerned with - they are more concerned with the functionality.

Item 2 is only flagged on the second set of results. It appears there because the balance between the Agrees (already low compared to the standardised database) and the Undecideds has shifted to the Undecideds with the removal of User 6. Generally speaking, if we are well satisfied with something we are happy to recommend it to others. However, in this instance the software is for a specific purpose and anyone who needs to use it, already does so. This may account for the high number of people who ticked undecided. The box being ticked not because they are really undecided about whether they would recommend it, but because they do not see to whom it could usefully be recommended.

• **Control** The two items flagged on this scale (4 and 39) are both on the set of results for 10 users only, so the Outlier (user 6) who disagreed with Item 4 and agreed with Item 39 has tipped the balance on both of these.

With regard to Item 4, the system was run over a wide area network and there were occasional problems with the communication lines. This could be what is being reflected in the response to this question

Regarding Item 39, the reason for the high Chi Square is again the large number of people who are undecided. As with the result from Item 2, the reason for this response could be that the job dictates what is to be done and the question has been interpreted as the user being able to choose what to do with the software away from the job activity.

5.4.4 Summary of the IM situation

Overall, it may be seen that this software has been well accepted and is well liked. The high scoring on Efficiency suggests that people do find that it helps them as a tool and is, therefore, useful. The slightly wider variation on scores on this scale could be due to the fact that, with the different ways of working between the depots and the CS, it would not be possible for once piece of software to fully match the requirements of both. When the results of the Control scale are also considered, the results of the ICA suggest that some early problems with software bugs may have influenced people's perceptions but also that the "restrictive" nature of the software i.e. that it is designed for a particular task and is not flexible in the sense that a spreadsheet package would be, might also colour some people's view, hence the high number of Undecideds in response to item 39. The scores on Helpfulness and Learnability suggest that people find it easy to use. Perhaps this is partly why the Affect score is high. The interface is character based and menu driven and even at the time the SUMI was distributed was starting to look "old-fashioned". However, many features associated with a graphical user interface (GUI) were not required for the tasks and the method of working the software supported.

5.4.5 Comparison with the User Engagement Model

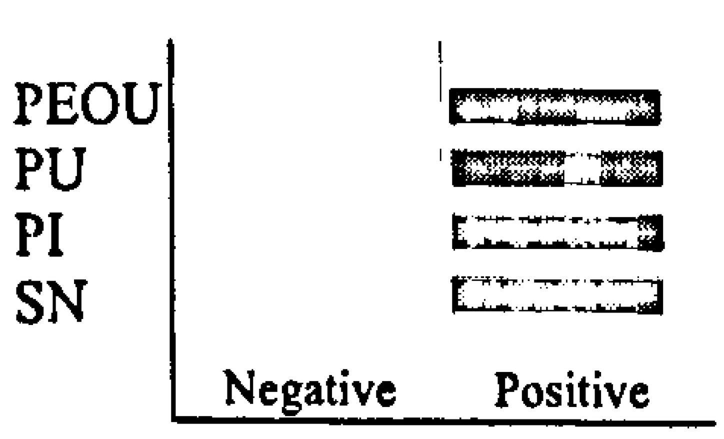
According to the hypotheses of the proposed model, perceived ease of use, perceived usefulness, perceived involvement and perceptions about co-workers and managers will all influence satisfaction.

- **PEOU** The Helpfulness and Learnability scales of the SUMI questionnaire have been used to give an indication of perceived ease of use. Both of these obtained a good score. It is possible that one reason for this is the training that was given when the package was first introduced. Also, the fact that an interface is character based and menu driven, as this package is, does not mean that it is not easy.

- **PU** The SUMI results on Efficiency and Control also suggest that the software is perceived to be useful for the tasks being performed. The slightly wider spread between the Ucl and Lcl and the lower score on the Control scale could be a reflection of the fact that working practices are slightly different between locations and the software has to fit all.

- **PI** SUMI does not measure perceived involvement. The evidence of the interviews suggests that people did feel they had been involved. Before the IM software was purchased people were interviewed as part of the requirements gathering process and they were then invited to see overviews of the shortlisted packages. The package purchased and introduced was not their first choice. However, the view of the four people in the informal interviews was that there had been sufficient involvement in the whole process. There was an underlying suggestion that the final choice always lay with senior management but there was no obvious resentment about this. Certainly, the senior manager in charge of the system would regularly 'phone or visit people to ensure they were happy with what was happening. Perhaps this helped people to feel that their views were important and made them feel more positive about the system.

●SN The attitude of co-workers and the manager in charge (SN) seems to have had a positive effect. Given that the work practices continued to differ between the locations and that the software was a compromise for everyone, it would be possible for everyone to have disliked it. However, some users were very positive about the software and the senior manager in charge took care to discuss issues and explain decisions with people. It would seem likely that this helped people, who might otherwise have been disgruntled, to take a more positive view. Indeed, the overall scores obtained suggest that people are accepting a compromise.



All the factors proposed by the model as contributing to a disposition towards satisfaction were present with regard to this software. The SUMI results also suggest general satisfaction which was supported by the initial interviews. As no corrective action was deemed to be necessary with this software, people were not interviewed about the results of the SUMI.

The distribution of the SUMI and the results for the MM software are discussed in the next section.

5.5 The MM software

5.5.1 The MM function supported by the software

The maintenance function is to ensure that equipment across England, Wales and the Channel Islands is kept in good working order. It is performed from three separate locations, each being responsible for a different section of this area. Each location is semi-autonomous, the workforce being managed at the local level. However, the overall strategy is set by one section (CM), which also reports on how well the maintenance operation as a whole is functioning. The CM section is local to one of the Maintenance locations. The Maintenance locations are shown as MS on the schematic of the organisation in figure 5.1 as they are physically the same locations as the Maintenance Stores.

Each of the Maintenance locations schedules preventive maintenance work on the equipment in its area. This involves co-ordination of transport and supply of spares as well as management of manpower. In addition to planned work, they have to perform corrective work. This latter is deemed more or less urgent depending on the effect it may have in the provision of the service the equipment is used for. In some cases the work can be added to the preventive maintenance schedule as an additional one-off job, in others an immediate response is needed. Work required for preventive maintenance purposes is keyed into the system with a planned date and produced in the form of a Work Order for the maintenance staff to perform. Once the work has been carried out, and the physical Work Order signed off, its status on the system is changed to reflect this as part of the feedback process. The system will produce another Work Order according to the periodicity placed against the job. Non preventive work is raised as an Unplanned Work Order. Feedback can consist of a simple status change to Completed or there may be a report on what was done and why. The time spent on the job and any materials used for the work should also be recorded. The people who raise the Work Orders might also key in the feedback but there is an assistant to help with this at each of the locations. CM staff collate information from each of the locations and report to senior management on how successfully the maintenance function is being performed in terms of providing continuous operation of equipment and the number of corrective action jobs that have had to be undertaken with the reason and estimated cost of these.

- **Background to software purchase**

The software to be evaluated was purchased two years after the second re-organisation of the whole maintenance operation. Originally, all maintenance staff had operated from a central location. In the first re-organisation this had been disbanded and the total number of staff reduced, but the staffing at local level was increased slightly to take account of the full maintenance function now required to operate there. Following the re-organisation, these locations had developed their own methods of working and it was felt that the overall view of the cost and effectiveness of the operation as a whole had been lost. The second re-organisation saw the establishment of the CM section to ensure common practices and reporting procedures at each location. A software package, which had been purchased some years previously but little used, was resurrected and installed at the maintenance locations as an aid to ensuring conformity of working. This had been in place for approximately one year when it was decided that there should be an investigation into what was required for the job and the system replaced, as it was evidently not successfully fulfilling its purpose: There were complaints at user level that the software was "unfriendly", time-consuming and of no benefit to them. At management level it was felt that the reports generated were not accurate as insufficient or incorrect data had been entered. The selection of the software to replace this system is described below. This did not form part of the research project but is included here as it is needed in order to properly interpret the results of the SUMI questionnaire which was issued as part of this research project.

- **Selection of the software**

The standard steps as now described were followed. Semi-structured interviews were conducted at each of the maintenance locations. These were with people responsible for planning the work required of specific maintenance disciplines, with the managers responsible for the whole maintenance function at the location and with the people who actually keyed data into the then existing system. This amounted to nineteen people in all, three of whom were female. Ages ranged from early twenties to early fifties with the majority being in the mid-thirty to mid-forty range. The interviews were geared to gathering information on how the task was currently done, what the user understood the aim of their job to be, how they felt it could be improved and the good and bad points of the current system of operation. Separate questions relating specifically to the software in use were not posed. The people responsible for planning the work required were also asked to do walkthroughs of how they determined what work was required, when it should be done and who should do it and how what had been done was recorded. Examples of paper documents were taken where possible. The CM staff were also interviewed. This consisted of three people (two male and one female with an age range of early forties to mid fifties). These interviews focused on how they saw a future system operating and the type of information they hoped to get from it. Notes were written up following the interviews and walkthroughs and sent under confidential cover to the participants with a request that they should make any changes they felt necessary and return them. All notes were treated in confidence.

There was an evident disparity between the amount of management information perceived to be required at local and at central level. It was also clear that the methods of planning work varied at each location, use being made of manual filing systems, notebooks and a wallchart. The existing software was used purely as a means of recording work done. As it was accepted that CM had a global perspective and that they were driving the need for new software, the requirements specification for the system was written mainly to their requirements. The specification was agreed by senior management and the IT department invited potential suppliers to respond to it.

A list of 23 key points was drawn up and each was given a weighting factor according to importance (determined by the IT department on the basis of discussions and the results of interviews with both CM and local level management). The replies were measured against this list by the IT department. Cost of hardware and platform did not feature in the list as the primary aim was to find software suitable for the job. However, once the scorings had been totalled, those packages which were more expensive by a large margin than another with the same score were rejected.

The suppliers of the top three scoring packages were invited to give a demonstration of their product, showing a typical process from the raising of a Work Order (instruction to do a job) to the entering of feedback from the job and describing what information could be obtained from the system. Local management and CM were invited to see the demonstrations and to ask any questions they felt had not been not addressed. The people who would be responsible for raising jobs on the system and entering feedback were not invited. The demonstrations were given over a period of three consecutive days. The software was chosen by a show of hands after some discussion and the decision was virtually unanimous. The exception being one of the CM team who expressed a concern on the support the supplier company could offer, not on the software itself. The software was PC based running under the DOS operating system with a top bar and pull down menus and selection by highlighting the chosen option.

It was agreed that each location would have its local system but that they must conform to a standard set by the CM group on data structure. The CM group would have access to the local systems in order to compile reports on a global basis. Two local users from each site and the CM team (not local management) were given training in the system. The training was away from their usual place of work and used unfamiliar data. The system was introduced on a single PC at each site with the intention that the base data should be keyed in before the software was used live. However, this took longer than had been anticipated (largely because the data did not exist in a readily available or consistent format) and pressure from the CM team, who felt that the system would be better used if more people had access to it, resulted in two more PCs being introduced at each location and networked together before smooth operation had been achieved. Use of the previous software had ceased at the time the replacement was purchased.

The sections below describe the work undertaken as part of the research. At the time that users' views were sought, the software had been in place for four years. During this time an enhancement to the system had been made. This was requested by the users to make data lookups in one part of the system easier. In the same time period a key player in the CM team had left and been replaced by someone new to the organisation whose views were sometimes at variance with one of the other members of the team. It had been reported that people were not happy with the software and it had been decided to review its use.

5.5.2 Informal interviews and the administration of SUMI questionnaires

- **Informal Interviews**

Informal, semi-structured interviews, taking the form of a general discussion in the individual's normal environment, were carried out at two of the Maintenance locations before issuing the SUMI questionnaires in order to gather some general information. Two people (one male and one female) were interviewed at one location and one (male) at the second. None of these were users of the IM software described in section 5.4. All three had been involved in the initial interviews and walkthroughs before the requirements specification had been written. None had been present at the software demonstrations prior to the selection. The reason given for the discussions and handing out of the SUMI questionnaires was that IT were concerned about the poor feedback which was being received about the system. Brief notes were taken during the time and fully written up immediately afterwards. The information being sought was the same as for the IM software in order to enable a comparison between the user perceptions of the two packages.

•Distribution of SUMI questionnaires

The SUMI questionnaire was given to each of the users for whom use of the system formed the main part of their work. This numbered fifteen in all, of whom 10 were male and five female with ages ranging from early twenties to early fifties. Of these three had not been in post at the time of the introduction of the software. SUMI questionnaires were not given to anyone who used the output of the system only and did not have direct contact with the software.

Following the guidelines given by the producers of SUMI, each questionnaire was accompanied by a typed note giving a brief explanation of what it was and how to complete it. Where possible the questionnaire and note were handed personally to the user and a verbal explanation given as to the reason for it. All fifteen questionnaires were returned. None was returned anonymously.

Once all the questionnaires had been returned the data was keyed in as a table and the program supplied with the SUMI questionnaire was used to analyse it and to produce various reports.

5.5.3 Results of informal interviews and the SUMI questionnaires

• Informal interviews

a) **Previous experience**

One person had used the previous MM system on a regular basis and had prior computer experience. The second had very little prior experience of any system. The third had built a PC from a kit and written some code using assembler language but had little use of application software.

b)**Amount of involvement**

None felt they had been involved in any way prior to the software in question being introduced. When reminded that their views had been sought at the requirements gathering stage, each agreed but said that they did not regard that as being involved. They had not seen what alternative software was available and had not been involved in the selection process. One even said that he would not have chosen the package selected because he did not like it.

c)**Number of years using the software in question**

All three had been using the software since its introduction nearly fours years before the discussion.

d)**Implementation**

One person felt that overall there was no problem with the implementation process, although it had been hard work and the pace rather too fast at one stage. The second felt that there should have been more information from the CM group at this stage and involvement by them in the design of the data structure (a coding structure in order to track movement of pieces of equipment was needed and this had been designed by the CM team). People were told to enter data and "we didn't know why we were doing it". The third person was not happy with the process but, when asked how it might have been improved, said that it was "probably the best that could be done in the circumstances".

e)Training

All three felt that the training itself had been good. However, each also went on to say that, whilst they were trained in how to use the software, it would have been useful if they had been shown its application in their day to day work.

f) Useful Tool

None could (would?) say how they really felt about this. The direct question was responded to by either comparing the features available in the software with those in the previous software; by saying that you do what you are told to at work; or by the statement that it wasn't yet clear what the job was.

g) Easy to use

All felt that the software was awkward to use but none could point to any areas that were difficult. Neither were any suggestions on how it could be improved upon forthcoming, except that one person felt it should have a "Windows" interface.

h) Views of others

All felt that their views were shared by others. This concurs with the general feedback and information from chance discussions.

The general feeling of dissatisfaction apparent in the discussions gave further weight to the poor feedback received about the system. However, only one gave any specific aspect of the software which was disliked. In this case, it was that the user preferred a Windows interface whereas the system was DOS with pull down menus. Evidently none felt that inclusion in the requirements gathering interviews and walkthroughs equated to being involved in any way. The first time they saw the selected software was when they went on the training course. Whilst the pace of implementation seems to have been acceptable, again there seems to be a feeling that the method was dictated without reference to them and their views. Even though the training itself seems to have been acceptable, the impression given was that the software was alien to the day to day work.

There was a feeling that a different way of working was being demanded without clear guidance on what that was or how it should be achieved.

- **SUMI questionnaires - summary of results**

The SUMI program was used to produce the following tables:

a) Profile Analysis

b) User Scales

c) Item Consensual Analysis

When the analysis was run, there was one user who was flagged as an Outlier. As that user could be eliminated without the sample falling below the minimum ten recommended, the data was removed and the analysis re-run as suggested. Removing the Outlier reduced the Global score from 33 to 31. No further Outliers appeared on the second run. The tables presented in the body of the text include the Outlier but the discussion refers to both sets of results where the removal of the Outlier has made a difference. Only those items from the Item Consensual Analysis (ICA) which were asterisked (i.e. between 95% and 99.99% likely to be different from the standardised database) are discussed. However, that is a large number. There are 33 items from the first run and the removal of the Outlier resulted in a further two on the second run.

a) **Profile Analysis**

The results for each scale from the SUMI are shown in table 5.4.

Profile Analysis - first run

Scale	UF	Ucl	Medn	Lcl	LF
Global	64	39	33	27	-2
Efficiency	56	34	28	22	2
Affect	81	47	41	35	0
Helpfulness	71	47	41	35	10
Control	52	33	29	25	7
Learnability	87	43	35	27	-12

Table 5.4

The Global scale of the Profile Analysis on both sets of results (33 and 31 respectively) indicates that the software falls well below what would normally be expected for commercial state of the art software. The upper and lower confidence limits (Ucl, Lcl) span 6 on either side of the median on this scale. Given the sample sizes of 15 and 14, this is perhaps greater than one would expect and might suggest a fairly disparate view. (The sample for the IM software, which was smaller, spanned 5 either side of the median). The Affect and Helpfulness subscales are the highest scores but these still fall below 50 although just within the satisfactory band, at 41, on the run with 15 people. Affect falls to 37 and Helpfulness to 40 on the run excluding the Outlier. The spread around the median of the Learnability scale, which is 8 on either side, suggests quite a wide variety of opinion on this scale. Interestingly, the spread for Control is the smallest, being just 4 either side of the median showing a close agreement of opinion. Control and Efficiency are the two lowest scoring scales in both sets of results and are very close to each other (28 for Efficiency and 29 for Control on the first sets of results and 27 and 28 respectively on the second run). This suggests that the software is failing to support people in their work. It is 95% certain that the true score lies somewhere between the Ucl and Lcl. Even the Ucl falls below 50 on all scales so it is clear that the software is not satisfactory.

b) **User Scales**

The scores for the individual users against each scale are presented in table 5.5. It may be seen from these scores that the high 58 against Efficiency by user 14 is why this person has been flagged as an Outlier. Control also achieves a high score with this person. The job performed with the software is no different from that of users 3 and 9, yet there is a great difference in scores between these three. This disparate view is, to a lesser extent, also reflected in the Control scale (the lowest being 22 and the highest 38). Three other people (users 7,8 and 13) who have similar job functions for which they use the software, are somewhat closer in their views although there is still a difference of ten between the three of them on the Efficiency scale, which falls to a difference of six on the Control scale. If the software is being used to perform the same job in the same manner at the three locations, it is perhaps surprising that opinion about it should differ so much.

The relationship between the local line management and the CM team is worth considering here. At one location, the local manager in charge felt that the CM team were trying to usurp his authority through the system and that he could report the necessary information rather than them obtain it directly from the system, which was not always accurate. The line management at the second location was of the same opinion but side-stepped the issue by giving responsibility for the system, but not for the allocation of work, to someone else. At the third location, the line

Individual User Scores						
User	Global	Efficiency	Affect	Helpfulness	Control	Learnability
1	20	16	32	41	22	43
2	16	16	26	26	20	47
3	13	11	42	23	22	14
4	28	20	27	34	23	32
5	25	25	30	49	22	21
6	37	20	41	54	29	35
7	44	38	42	55	37	63
8	37	35	49	38	38	55
9	40	44	57	30	38	48
10	33	28	33	47	26	21
11	26	29	23	36	34	31
12	16	23	19	20	13	17
13	57	48	58	68	43	54
14	42	58	54	44	33	62 (E)
15	42	31	54	50	34	32

(Table 5.5)

manager felt that there was a need for a system but did not always agree with the CM team on how it was operated. The result was that the three locations did not operate in exactly the same way but were being asked to use the software in a standard manner. The CM team themselves expressed frustration because they had been given a task but seemed to be in constant battle with the three locations in order to carry it out.

c) **Item Consensual Analysis**

The items flagged from the IC Analysis in both sets of results are shown in table 5.6. They are listed in the same order as they were reported, those at the top being the most likely to differ from the standardised database. In some cases the descriptions have been altered slightly from how they appear in the SUMI in order to shorten them.

The ICA gives a comparison between the actual responses given and what should be expected according to the standardised database. As so many items were flagged as likely to be different (33 from the first run and a further two from the second run) the results have been grouped according to the subscale they represent in the discussion below. Appendix 2 contains examples of some of the items in the order in which they were flagged on the ICA, with a comment against each. This gives more detail on each item than is contained in the body of this text.

Item Consensual Analysis

Item	15 users	14 users	Scale
44 It is relatively easy to move from one part of a task to another	✓	✓	C
1 The software responds too slowly to inputs	✓	✓	E
47 This software is really very awkward	✓	✓	A
21 I think this software is inconsistent	✓	✓	E
49 Getting data files in and out of the system is not easy	✓	✓	C
34 The s/w allows the user to be economic of keystrokes	✓	✓	C
42 The s/w has a very attractive presentation	✓	✓	A
45 It is easy to forget how to do things with this s/w	✓	✓	L
26 Tasks can be performed in a straightforward manner	✓	✓	E
41 The s/w hasn't always done what I was expecting	✓	✓	E
2 I would recommend this s/w to my colleagues	✓	✓	A
36 There are too many steps required to get something to work	✓	✓	E
13 The way that system information is presented is clear and understandable	✓	✓	H
24 The s/w is awkward when I want to do something which is not standard	✓	✓	C
5 Learning to operate this s/w is initially full of problems	✓	✓	L
39 It is easy to make the s/w do exactly what you want	✓	✓	C
15 The s/w documentation is very informative	✓	✓	L
46 The s/w occasionally behaves in a way which can't be understood	✓	✓	E
31 It is obvious that user needs have been fully taken into consideration	✓	✓	E
27 Using this s/w is frustrating	✓	✓	A
48 It is easy to see at a glance what the options are at each stage	✓	✓	H
32 There have been times in using this s/w when I felt tense	✓	✓	A
4 The s/w has at some time stopped unexpectedly	✓	✓	C
7 I enjoy my sessions with this s/w	✓	✓	A
29 The speed of this s/w is fast enough	✓	✓	C
37 I think this s/w has made me have a headache on occasions	✓	✓	A
19 I feel in command of this s/w when I am using it	✓	✓	C
28 The s/w has helped me overcome any problems I have had in using it.	✓	✓	H
30 I keep having to go back to look at the guides	✓	✓	L
8 I find that the help information is not very useful	✓	✓	H
50 I have to look for assistance most times when I use the s/w	✓	✓	L
25 There is too much to read before you can use the s/w	✓	✓	L
40 I will never learn to use all that is offered by this s/w	✓	✓	L
20 I prefer to stick to the facilities that I know best	-	✓	L
22 I would not like to use this s/w every day	-	✓	A

Table 5.6

•**Efficiency** This subscale measures whether the user feels that the software imposes extra mental effort in order to achieve the desired tasks. The median of the Efficiency score is just 28. This is the lowest rated of the five items measured by the SUMI questionnaire. The spread between the upper and lower confidence levels is 12 which suggests quite a large variation between the 15 people involved. User 3 returns a particularly low score in this area with a total of 11 which contrasts with the score of User 14 (highlighted as an Outlier in this area) who returns a score of 58.

Of the ten items in this section, five are flagged as 99.99% likely to be different from the norm when compared to the standardised database and two as 99% likely to be different. These are items 1,21,26,41 and 36 and items 46 and 31 respectively. Such a poor rating suggests that a lot of effort has to be put into using the software. This could be because it is not easy to use or because it does not help the job in hand but imposes an extra workload.

Item 31 ("It is obvious that user needs have been fully taken into consideration") has been ticked as disagree by the majority of people. Although the IT department spoke to the people who would actually be keying in to the new system at the requirements gathering stage, the list of requirements was based primarily on those of the CM team who were driving the need for a new system and it was agreed to by senior management without reference to the users. Only the members of the CM team and the local line managers were present at the demonstrations given of potentially suitable software. It is possible that the fact that the key users of the system were not present at the demonstrations meant that features of the software which were needed but missing, or which were present but demanded a different way of working, were not highlighted. However, as the output from the software seems to match the requirements of the job as laid down by those in charge, it could also be that the dissatisfaction stems more from the fact that there is a perception that the software was imposed without the users themselves being consulted and that it does not match the task as perceived by the users. This is supported by the feedback in the interviews that people didn't know why they were being asked to do certain things. The Efficiency score suggests that some of the dissatisfaction being expressed has arisen as a result of the lack of perceived usefulness of the software.

In the three depot locations, as described earlier, the line management displayed differing but never supportive attitudes towards the system. None of them used it on a daily basis for information extraction and the procedures for use were laid down by the CM team. These managers insisted that the system be used for information gathering and saw it as a means of ensuring a consistent method of working and reporting at each of the locations. In effect, the local line management could not openly reject the system but did not require its use. The CM team had the authority of senior management to enforce the use of the system. For them the system was a means of ensuring greater control over how work was carried out and reported. Local management felt that they knew what was going on in their own area and had no need of further help. The CM team, being remote and needing to know what was happening at all three locations, had more need of a tool to assist with this. However, the users at the depots did not see why they should be using software and producing information for which their local management appeared to have no need. This contrasted with the users of the IM system who have a central manager who talked directly with them and supported them and their use of the software.

• **Control** measures the perception the user has that the software responds in a consistent manner making it easy to understand which steps are required. The score for this subscale is just one point above that of Efficiency at 29 (28 on the second run). User 12 gives the lowest score at 13 with the highest score, 38, from Users 8 and 9. Again the difference in scores is very broad. Five of the ten items are marked as 99% or above likely to be different from the standardised database (44,49,34,24,39) and three are 90% likely to be different (4,29 and 19). The software is character based, running on PC. Sometimes, it may be necessary to go into two separate areas of the system to look up information. It is not possible to have more than one window open at once. It is also possible to perform the same task in more than one way, depending on working practices, so there is not just one route which may be followed and it is necessary for the user to know which method is used within his organisation. This direction must be provided externally to the system. It could be that the procedures for working which have been set by the CM are felt to be awkward and unnecessary or even that they are not sufficiently clear, and that this is being reflected in the responses. One of the comments made during the interviews before the issue of the SUMI questionnaire had been that it would have been preferable had the training concentrated on the job to be performed rather than on the software.

• **Learnability** This subscale, as the title implies, measures the ease with which the user can become familiar with the software. The median score is 35. Again there are eight items which fall short of the standardised database. Three (45,5 and 15) are 99% or more likely to be different from the norm and five (30,50,25,40 and 20) are 95% likely to be different. Everyone who would use the software was given training on it and the feedback (on evaluation forms issued after the training as part of company procedures) suggested that people were happy with it (although some mentioned a dislike of the premises). This feeling is supported by the interview results. However, this does not appear to have made people feel that the software is any easier to use. A large number of people preferred to stick to the facilities they knew best (Item 20). However, even though a large number of people (73%) agreed that it is easy to forget how to do things with the software (Item 45), still only 33% said that they generally had to look for assistance when using it (Item 50). This is perhaps because they preferred to stick to the facilities they knew best (Item 20). This lack of a desire to explore the facilities of the software could be because it was indeed problematic to use or because it was not viewed as a tool but as an additional workload imposed for the benefit of others.

• **Affect** This subscale measures how the user feels as a result of interacting with the software. Perhaps surprisingly, the median for this subscale is one of the highest at 41 (37 on the second run but still one of the higher scores). However, it is still below the standard 50 and the spread between the upper and lower confidence levels is still quite wide at 12. This disparity is reflected in the individual user scores where User 12 has a score of 19 compared to User 13 who scores 58. Of the 8 items which fall outside of the norm from this subscale (items 47,42,2,27,32,7,37 and 22), it is perhaps Item 2 which is most interesting. Despite finding the software awkward to use (Item 47), lacking an attractive presentation (Item 42), frustrating to use (Item 27) and so on, still 60% of the users were undecided whether they would recommend it to colleagues (Item 2). User 3, who returned a very low score on the Efficiency scale actually ticked Agree on this question. Perhaps he feels that if he has to use the software, so should everyone else! This generally ambivalent attitude perhaps again suggests that the negative feelings being expressed are not so much with the software as with its use.

• **Helpfulness** This subscale, which measures the user's perception of how helpful the software is in making navigation easy or giving meaningful error messages etc., scored 41 on the first run and 40 on the second, making it the highest scorer along with Affect. Just four items are flagged on this scale (13,48,28,8). Only on Item 13 ("the way the information is presented is clear and understandable") is it the high number of people who are undecided which has caused the item to be flagged. On the other items the responses are more definite. However, this could be showing a distinction between the helpfulness of the software per se and its helpfulness in explaining how a task should be performed. In other words, there is a mis-fit between the way the user wishes to work and the how the system operates. The fact that the system meets the requirements specified by senior management perhaps reflects the mis-fit between the way they perceive the job and the perceptions of the people who perform it. This would seem to indicate that the global score is reflecting problems which are not to do with the software itself.

5.5.4 Summary of the MM situation

It is evident that the people using the software do not feel that it is easy to use, despite having had training which they felt was sufficient. Nor do they feel that the software is useful to them. This would seem to arise, at least in part, because of a mismatch between the way they perceive the job and the way in which the CM team perceive it. The slightly higher score on the Affect scale suggests that people are not complaining specifically about the software but that it was not their chosen tool. If the tool is designed to perform one job and people are trying to use it to perform a different job, then the tool would appear difficult to use. It could also be that the software is indeed difficult to use. In order to assess whether it was the software or the context which was the cause of the problem an attempt was made to find a comparable organisation and an independent evaluation was carried out. These are described later.

The SUMI results lend support to the results of the interviews and show that people neither perceive the software to be easy to use nor to be useful. The informal interviews also suggested that they did not have any sense of involvement in the process of selecting the software. Being asked about how they currently worked and what they would like to see in a new package, evidently did not class as involvement. The only difference between the process of selecting the MM software and the IM software is that, in the case of the latter, everyone was invited to the overviews. Even so the users of the IM system also had the software imposed on them in that it was not their first choice. However, senior management explained their reasons for preferring a different package and that package was demonstrated a second time, with closer attention to how it could meet specific user requirements. People were then able to accept the decision of senior management. The users of the MM software felt that they had not had any involvement in the choice of tool to perform the job. It is possible that this lack of involvement is partly why dissatisfaction with the software is being expressed.

Although people were working at different locations and had their own methods of performing tasks which were similar across locations, they were united in their dissatisfaction with information from the CM team. Their line management adopted a non-supportive stance with regard to the software and neither was there support from the CM team. Use of the system was dictated rather than encouraged. This contrasts with the IM system where there was one person in overall charge who took an interest in people's views even if they were not accepted.

5.5.5 Follow up interviews post SUMI questionnaire

Informal interviews were undertaken at two of the locations just over one year from issue of the SUMI questionnaires for the MM system. Two people from each were interviewed. Of these four people, two had been interviewed pre the SUMI and two had not. One was user 13 who had given a particularly high score on the Helpfulness scales (68) and one was user 14 who had returned a high score on Efficiency (58). The aim of the interviews was:

a) to confirm the interpretation of the SUMI results

b) to ascertain whether their feelings towards the software had altered in the intervening year

c) to establish whether the feeling of lack of involvement had indeed also been felt by the two people not interviewed previously and whether they felt that they understood the purpose of the software (initial interviews with others suggested that there was a feeling that the software itself was unnecessary).

d) to clarify with users 13 and 14, who had been very positive in some responses and hence out of step with the norm, how the questions had been understood and, indeed, whether their feelings towards the software were of a more positive nature.

The ages of the four people interviewed ranged from late twenties to mid fifties. Three were male and one was female.

a) **Interpretation of SUMI** The result of the discussions with all four people was that the interpretation of the SUMI in suggesting that the problem lay not just with the software, was correct. One person did point out one area that he felt was not easy to navigate but overall thought that the software was not difficult to use. This view was also held by the others. One person felt that the original purpose for which the software had been purchased had changed and that it was being used to control work that was not originally intended. One of the others also said that there had been a feeling of indecision and lack of direction in the early days which had led to a feeling of dissatisfaction with the software. This corresponds to the original interviews in which it was said that training was given on the software but not on how it was to be used. The two people who had been interviewed prior to the SUMI still felt that there had been conflicting messages from the CM team and that these were still not clear. One also said that, although the CM team could now obtain the information they needed via the software and were supposed to be in control, they would still 'phone up with questions.

b) **Change of feeling** Over the year since the questionnaire had been completed there had not been much change in view (this was confirmed by picking out items from the questionnaire and rechecking the response). However, the comment was made that things were now more settled and people were more used to using the software in the day-to-day work. It was again suggested that the problem lay not so much with the software as the way in which it was being used.

c) **Sense of Involvement** The two people who had not been previously interviewed both said that they had not been involved in the process leading to the change of software (even though both had been involved in the requirements gathering process). Both felt that they had "got what others had decided was needed" and that they had not been involved in the decision. This is a similar response to that given by people in the original interviews. The reason for introducing the software, they felt, was to provide the CM team with information so that they would have more control. From the point of view of one, it was extra work to little gain.

d) **Clarification with users 13 & 14** User 14 said that there had been a feeling of lack of direction in the early days of the introduction of the software and that this had led to a feeling of dissatisfaction. However, he had been given responsibility for the system at his location and felt that the software was very good, it was just a case of using it in the "right way".

User 13 confirmed an understanding of the questions in line with the others and felt that one reason for having more positive results on the SUMI was a feeling of being in control when using it. The user was an experienced keyboard user and in many parts of the system found that it was no longer necessary to read the screen prompts as familiarity meant that the next keystroke required was known. In areas which were less familiar it was felt that the instructions on the screen could be followed exactly in order to achieve the required results. It was felt that other users might question the instructions given by the system or try to shortcut them, rather than follow exactly the steps laid out.

There was no information given in the discussions (which were wide ranging due to the informality!) which suggested a different interpretation should be given from that of lack of perceived involvement and alienation from the CM team. However, it did seem that the software was not thought to be difficult of itself, just that it was not helpful in the job it was being used for.

Overall, it would seem that SUMI was correct in reporting dissatisfaction. However, there were reasons that were not just to do with the software.

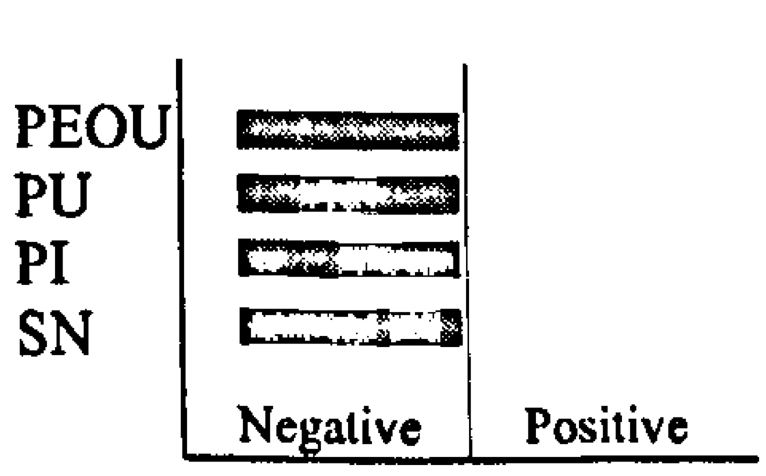
5.5.6 Comparison with the User Engagement Model

- **PEOU** Although the Helpfulness scale scored 40, placing it just within the Satisfactory category, the Learnability scale scored only 35. Generally, people were saying that they did not find the software intuitive to use. Although this must be balanced with their perception that it did not fit their way of working, it could be said that perceived ease of use was not very great.

- **PU** Both the Efficiency and Control scales achieve very low scores. Added to this are the comments that the software does not help people perform their tasks and is not there for their benefit. The implication is that, if the software were removed, the work would not suffer.

- **PI** Despite having been included in the requirements gathering process, no-one felt that they had been involved in any way. In the interviews there was a sense of people having been given a tool for something that they did not understand and could only assume it must be to give greater control to others. Had they been asked, they could have said that the software did not suit their way of working.

- **SN** Local management support was completely lacking. There was no encouragement to use the software, not even from the CM team who were seen as dictators without direction. Nor were there any individuals who viewed the software in a positive light and encouraged colleagues.



None of the factors proposed as contributing to satisfaction was strongly present and the general opinion of this software was that it was not satisfactory. The result of the SUMI indicated this lack of satisfaction which was confirmed by the informal interviews.

The distribution of the SUMI and the results for the PM software are discussed in the next section.

5.6 PM Software

5.6.1 The PM function supported by the software

The work of the Engineering division is divided across five multi-skilled teams who are each responsible, under the direction of a team leader, for planning, budgeting and scheduling their work. The resources of a team of Technical Installers, the Drawing Office and a team of Quality Assessors are shared between the Project teams. Each Project team liaises with the Drawing Office and the Technical Publications section to ensure up-to-date documentation; with the Administration section for the placing and receipt of orders; and with the Maintenance Division (described in the previous section) towards project completion in preparation for handover. A sixth group consists of three people, of whom one works full-time on project control, and two are in a senior position and responsible for resolving resource conflicts between the teams and for progress chasing.

- **Background to software purchase**

The software to be evaluated was originally purchased in the DOS version but was changed to the Windows platform two years later. It had been introduced by senior management because there was no consistent method of project planning across the five teams and it was difficult to readily see where bottlenecks were likely to occur. One of the senior people decided to see what was available by way of project management software and went to various demonstrations and spoke to a sister organisation. By the time the IT Department was involved, the decision had been made and a package selected.

- **Method of Software selection**

There was no formal method behind the choice of software, the decision on what was required having been made by one person and backed by others at his level of seniority. The IT Department was not involved until there was a requirement to purchase. At this point a brief survey by the IT Department of what was available (with no reference to how the teams worked or what was needed other than project management software) was made and it was agreed that the software could be purchased on the basis that it was well-known and amongst the market leaders of its type.

The change to the Windows version of the platform had been made because of the additional functionality felt to be available. There was no investigation by the IT Department into the benefits to be gained by the change which was again driven by the senior management.

Training was given to all members of the project teams in both the DOS and, later, the Windows versions of the software. This was organised by the Engineering management and carried out by a training company at the Engineering location just prior to the software being installed.

The section below describes the work undertaken as part of this research. At the time that the SUMI questionnaire was issued the Windows version of the software had been in use for over two years.

5.6.2 Informal interviews and the administration of SUMI questionnaires

• Informal Interviews

In this case, there was not sufficient time for any form of discussion other than that needed to explain the purpose of the SUMI questionnaires.

•Distribution of SUMI questionnaires

The SUMI questionnaire was given to each of the users responsible for ensuring that the project data was up-to-date plus one person who deputised for one of the project leaders; to the project co-ordinator; to the two senior managers; and to the person who monitored projects for budget spend (using the system for enquiry only). This numbered ten in all, of whom 9 were male and one female with ages ranging from mid twenties to early fifties.

Following the guidelines given by the producers of SUMI, each questionnaire was accompanied by a typed note giving a brief explanation of what it was and how to complete it. Where possible the questionnaire and note were handed personally to the user and a verbal explanation given as to the reason for it. Not everyone was available to be handed the questionnaire individually and some questionnaires were left with one of the senior managers for handing on. All ten questionnaires were returned.

Once all the questionnaires had been returned the data was keyed in as a table and the program supplied with the SUMI questionnaire was used to analyse it and to produce various reports.

5.6.3 Results of SUMI questionnaires

The SUMI program was used to produce the following tables:

a) Profile Analysis

b) User Scales

c) Item Consensual Analysis

When the above tables had been generated, one Outlier was flagged, the outlying score being on the Efficiency subscale. As that user was the person originally responsible for the introduction of the software and his scoring was generally positive, the data was removed and the analysis re-run as suggested, even though it meant that the number fell below ten. The re-run of the analysis after removing user 4's data did not produce any further Outliers. As this user has a natural bias in favour of the software (having been responsible for its introduction) and his responses affect the results overall, the Profile Analysis included in the text is the result of the second run with just nine people. The User Profiles for all ten users are shown and the discussion refers to both sets of results. Only those items from the Item Consensual Analysis which were asterisked (i.e. between 95% and 99.99% likely to be different from the standard database) are discussed.

a) **Profile Analysis**

The results from this are shown in table 5.7.

Profile Analysis - second run (9 users)

Scale	UF	Ucl	Medn	Lcl	LF
Global	62	44	35	27	5
Efficiency	64	45	37	29	10
Affect	72	38	28	18	-6
Helpfulness	74	48	40	32	2
Control	64	42	34	26	9
Learnability	72	50	40	30	-5

Table 5.7

Looking at the Profile Analysis scales, it may be seen that the Upper and Lower Confidence levels have a spread of between 16 and 20 around the Median, which is large even for the low number of users. The widest spread is on Affect and Learnability. This suggests a wide range of views amongst the users. The Helpfulness and Learnability subscales are the highest scores but these still fall below 50. When the Outlier (User 4) is included, the median of the Global scale is increased to 38.

b) User Scales

The results for all ten users are presented in table 5.8.

Individual User Scores - first run

User	Global	Efficiency	Affect	Helpfulness	Control	Learnability
1	21	25	20	21	23	20
2	25	37	14	26	34	40
3	18	29	20	25	14	21
4	63	71	58	54	63	49 (E)
5	24	19	21	32	27	17
6	56	58	46	50	57	62
7	41	46	28	40	35	59
8	43	48	45	45	47	47
9	46	40	53	51	45	46
10	35	28	57	53	33	26

Table 5.8

On the individual **User Scale** tables, User 4 is flagged as an Outlier on the Efficiency scale (71 as against the Upper Fence of 69 on the Profile Analysis for all 10 users). The Efficiency scale measures whether the user feels that the software enables the work to be done without imposing an additional mental workload. Looking at the results from the other users on this scale it can indeed be seen that the score seems rather high. The scores in other scales for this user are also higher than the others except on the scale of Learnability, where the score is lower. User 4 was mainly responsible for the selection and introduction of the software and so it is perhaps not surprising that he would consider it a useful tool. It is interesting to note that the aspects which could be said to relate particularly to the software itself (Helpfulness and Learnability) are the lowest scored by this user and yet are the two highest scorers overall. This suggests that the view taken by user 4 is that the software is a useful tool but one which could be improved upon. (This view was confirmed by a later interview with the user). Having a personal stake in the success of software, as user 4 does, has been proposed as a factor which influences satisfaction (Hartwick & Barki, 1994). The lower satisfaction expressed on the Helpfulness and Learnability scales could be because greater experience of software in the field has led to a higher expectation (e.g.Ramamurthy et al.,1992).

c) Item Consensual Analysis

In all, twenty-seven items are flagged over both sets of results. Twenty-six items are flagged in the run with nine people and twenty-one from the run with ten people. One item (Item 32) is flagged on the run with ten people but not on the second run with nine people. This is an example of the difficulty of having a small sample of people. The reason for Item 32 being flagged was the high number of people who either agreed or were undecided whether "There have been times in using the software when I have felt quite tense". User 4 had ticked undecided to this statement, which resulted in the 95% likely to be different flag being set on this item.

As there are so many items flagged, they have been grouped under the subscale to which they belong for discussion, rather than considered individually. Except where specifically indicated otherwise, the results are discussed from the run with just 9 people. Table 5.9 shows the items flagged from each run in the order in which they appear on the ICA. These have been shortened in some cases from how they appear in the SUMI.

Item Consensual Analysis

Item	10 users	9 users	Scale
47 This software is really very awkward	✓	✓	A
40 I will never learn to use all that is offered	✓	✓	L
7 I enjoy my sessions with this software	✓	✓	A
28The s/w has helped me overcome any problems I have had in using it	✓	✓	H
50I have to look for assistance most times when I use this software	✓	✓	L
9 If this s/w stops it is not easy to restart it	✓	✓	C
12 Working with this s/w is satisfying	✓	✓	A
38Error prevention messages are not adequate	✓	✓	H
14 I feel safer if I only use a few familiar commands	✓	✓	C
29 The speed of this s/w is fast enough	✓	✓	C
39It is easy to make the s/w do exactly what you want	✓	✓	C
5 Learning to operate the s/w is full of problems	✓	✓	L
48It is easy to see at a glance what the options are	✓	✓	H
37I think this s/w has made me have a headache	✓	✓	A
46The s/w occasionally behaves in a way which can't be understood	✓	✓	E
26Tasks can be performed in a straightforward manner	✓	✓	E
17Working with this s/w is mentally stimulating	✓	✓	A
34 The s/w allows one to be economic of keystrokes	✓	✓	C
8 The help information is not very useful	✓	✓	H
33The organisation of menus or information lists seems quite logical	✓	✓	H
32There have been times using the s/w when I have felt quite tense	✓	-	A
3 The instructions and prompts are helpful	-	✓	H
2 I would recommend this s/w to my colleagues	-	✓	A
22 Iwould not like to use this s/w every day	-	✓	A
41The s/w hasn't always done what I was expecting	-	✓	E
19 I feel in command of this s/w when using it	-	✓	C
15 The s/w documentation is very informative	-	✓	L

Table 5.9

•**Affect** This scale achieves the lowest score (28), with a spread of 20 from the Lcl to the Ucl showing a wide variation in opinion although always a low score. Seven items are flagged as likely to be different from the standardised database, three of which do not appear on the run with ten people. The seven items are 47,7,12,37,17,2 and 22. On the run including user 4, Item 32 is also flagged on this scale. No-one enjoys using the software (Item 7) although four people are undecided, so there is not a unanimous dislike of it. Although four people agree that the software is very awkward (Item 47), three people disagree. There is greater consensus of opinion against Item 37, where six people agree with the statement that the software has given them a headache on occasions and against Item 12 where five people disagree that working with the software is satisfying. However, although five people say they would not like to use the software every day (Item 32), four disagree, and a greater number of people are undecided whether they would recommend the software to colleagues (Item 2) than say they would not recommend it (4 and 3 respectively). This variation of views with an almost even split very often between those who agree and those who disagree with a statement is interesting. If the problem lay purely with the software a greater consensus view could be expected. The items are flagged as different from the standardised database because of the number of people who have ticked one of the options compared with the number who would normally be expected to tick that option. However, the fact that almost the same number have ticked the opposing option is worthy of note.

•**Control** With a median of 34, this scale has the second lowest score with a spread of 16 between the Lcl and the Ucl. Six items are flagged on the ICA (9,14,29,39,34 and 19). Item 19 does not appear on the run with 10 users. People are undecided whether the software is fast enough or whether they feel safer sticking to commands they are familiar with but are agreed that it is difficult to restart the software if it stops. None agrees that it is easy to ask the software to do exactly what is wanted, apart from user 4. Again, it would appear that the software is problematic. However, the response from user 4 contradicts this as does the independent evaluation as will be seen later. The wide spread between the Ucl and Lcl around the median of both Control and Efficiency suggests a wide range of views on these scales. Each team has a different set of projects which could be successfully completed by use of different methods. The fact that they share resources, however, demands a standard way of working which may fit the preferred approach of some better than that of others. It is possibly this that is being reflected in the scores.

•**Efficiency** At 37 this scale is closer to the top than the bottom of the scores. Three items are flagged as likely to be different from the standardised dataset, (41,46 and 26). Item 41 does not appear on the first run, which included user 4. The statements against these three items are very similar in that people are agreeing that the software hasn't always done what they were expecting and occasionally behaves in ways that can't be understood, and disagreeing that tasks can be performed in a straightforward manner using the software. User 4 gives the opposite response to each item. These results could be due to a lack of training on the software. However, that is unlikely as full training was given prior to the introduction of the DOS version and again with the introduction of the Windows version. It could also be that the software is inherently difficult to use but this is contradicted by the results of the independent evaluation as will be seen later and by the position of the software in the market at the time. This suggests that the software has been written to follow a method of working which is, at times, at variance with how the people using it would like to work. The responses from user 4, however, suggest that the method of working required by the software is approved of by those who desired its introduction (i.e. senior management).

• **Helpfulness** At 40 this is one of the highest rated scales, although there is still a spread of eight either side of the median. Six items (28,38,48,8, 33 and 3) are flagged against this scale. Item 3 is only flagged on the run with 9 people. There is certainly a consensus view that the software itself does not prevent people from making mistakes (Item 38) or help in problems with using it (Item 28). However, this could be because there is more than one way to achieve a goal in the software and the "mistake" could be that someone has not followed the standard format rather than that they have performed a "wrong" action. More help information is also required (Items 48 and 8). Two thick user manuals are provided with the software and the software is quite complex, partly because of the flexibility it offers. Again, the help required could be in how to perform the steps required for the standard way of working demanded of the teams. At the time the SUMI was distributed, a step by step guide on which codes had been set up in the software and where they were to be used, was being produced.

• **Learnability** also achieves a relatively high score (although at 40 only just on the minimum level for acceptable software) but the spread between the Ucl and Lcl is twenty showing a wider range of views than on the Helpfulness scale. Four items (40,50,5 and 15) are flagged as likely to differ from the standardised set. Item 15 is the only one that does not also appear on the run with ten users. There is unanimous agreement that people will never learn to use all that is offered in the software (Item 40). Even user 4 ticked undecided on this item on the run with ten people. The majority also agree that learning to operate the software is initially full of problems (Item 5). User 4 also agrees with this item although, surprisingly, three people are undecided. Eight people are undecided whether the software documentation is very informative (Item 15). This may be because there is so much information that it is not easy to find what is needed. These responses are perhaps an indication of the complexity of the software. Although each project individually is fairly straightforward, the need to merge them in order to share resources, meant that a simple package would not fulfil the purpose. Thus tasks which are not very complex of themselves need complex software to achieve them because of the need for the next stage in order to share resources.

5.6.4 Summary of the PM situation

The Helpfulness and Learnability scales achieve the highest overall scores. The items on both scales which are flagged as likely to differ from the standardised database all suggest that the software has a large number of functions, many of which are not needed by the users but which increase the complexity of use. The overall score and items ticked implies that, once familiarity is gained with the steps for the task in hand, the software is not difficult. This may be seen in the response to Item 23 on the Helpfulness scale ("I can understand and act on the information provided by this software"), and in the response to Item 20 on the Learnability scale ("I prefer to stick to the facilities I know best") where seven people agree with the statements in both cases. Thus the software is not perceived as easy to use but the reason is that there is far more to the software than is required to perform the job as perceived by the users. The standard method of working means that they have to use the correct codes from the coding structure to indicate which resources they are using, even though for themselves this information is not necessary. A parallel may be drawn here with the report by Le Quesne (1988) which indicates that the individual needs should be considered before attempting to co-ordinate a group effort by means of technology. However, the management reason for introducing the software was to achieve better co-ordination.

After Helpfulness and Learnability, Efficiency and then Control achieve the next highest scores. These two scales together relate to the perceived usefulness of the software. Although only three items on the Efficiency scale are sufficiently likely to be different from the standardised database to be flagged, six items are flagged on the Control scale. It may be that the need to use codes in order to achieve consistency across projects means that it takes longer to use. These codes do not assist in the managing of individual projects. Extra work is required which is not of immediate benefit to the project in hand. The result would seem to be that the software is not perceived to be useful to individual users.

The Affect scale produced the lowest score. This scale measures how people feel as a result of using the software and such a low score would suggest something is fundamentally wrong with it. However, as will be seen later, not only was the package a market leader at the time, it also scores well on the independent evaluation.

5.6.5 Follow up interviews post SUMI questionnaires

As there had been no time to carry out interviews before the issuing of the SUMI and the results indicated such a poor opinion of the software, it was decided to carry out informal, semi-structured interviews to confirm the interpretation of the SUMI. The aim of these was to try to determine whether people had felt involved and for them to express an opinion on the usefulness and ease of use of the software. Unfortunately, unforseen events meant that these did not take place until about 14 months after the SUMI had first been distributed. During this period, the project co-ordinator left the organisation. This caused a slight change round of staff before the follow-up interviews took place.

The interviews were undertaken with four people who had responded to the SUMI questionnaire. Of these, two were in the same post and two had moved posts. One of these was a sideways move and the other was a temporary promotion. However, they were all still involved in project management and using the software. All were team leaders. All were male. The age ranges were from mid twenties to mid fifties.

The answers to five formal questions were sought in the interviews with the surrounding discussion being around the general feeling about the software and its use. The five questions were to establish:

a) Whether there was still a feeling of dissatisfaction

b) What the reasons for introducing the software were perceived to be

c) Whether the software was easy to use

d) Whether the software was felt to be useful

e) Whether there had been any involvement in choosing the software

a) **Feeling of dissatisfaction** In all cases, the feeling of dissatisfaction was said to remain. However, no-one could point to any specific aspect of the software which was unsatisfactory apart from to say that the software "looked" too complex.

b) **Reasons for introducing the software** It was reported by all that the software was originally introduced in order to enforce a standard method of working onto disparate groups. The need to have the data in order to share resources was given as a secondary reason. Whilst there was agreement that this latter was needed, it was reluctantly given. There was a feeling that there had not been much wrong with the previous way of working when no software had been used.

c) **Ease of use** It was said that the software was far too complex for the needs of the job itself and was really there to benefit the management group. However, apart from saying that it was too complex, there were no specific areas that anyone pointed to with a view to how they could be improved. Conflicting comments were made. For example, one person said both that the screens were too busy and that he wanted more information displayed at the same time. Everyone agreed that full training had been given before the software was introduced but also felt that there had been too much to learn and that much of it wasn't relevant.

d) **Usefulness** It was felt that the job was made more difficult and that more effort was required as a result of the software and the enforced standards. It was used because it was a requirement of the job to use it but not because it was felt to have any benefits for the user.

e) **Involvement** Everyone felt that the software had been introduced without reference to them. However, there did not seem to be any expectation that they would be involved. The view seemed to be that management would do as they felt fit without regard to the views of others. It was not the aim of management to introduce a tool that would be useful to the people doing the work. (Yet one of the reasons given for the original need to purchase the software had, in fact, been that it would help people in the performance of their tasks!).

Each of the people interviewed was from a different team. Although they all expressed similar views, there was the impression that the teams worked independently of each other and each favoured different approaches to the tasks. It did not seem that the views of one influenced those of another. All four people were team leaders and, therefore, line managers themselves. It is possible that their views affected those of the people who worked to them but this would only affect three people.

The results of the discussions suggest that the general view of the software was that it was too complex for the needs of the individual user and that it was of use to others but not to the people who had to use it. There was no sense of ownership of the system. It was used because there was a requirement to use it. Evidence from the literature would suggest that involvement in the process of software selection would help people to appreciate the reasons behind buying it and thus feel more favourably towards it (e.g. Damodaran, 1991).

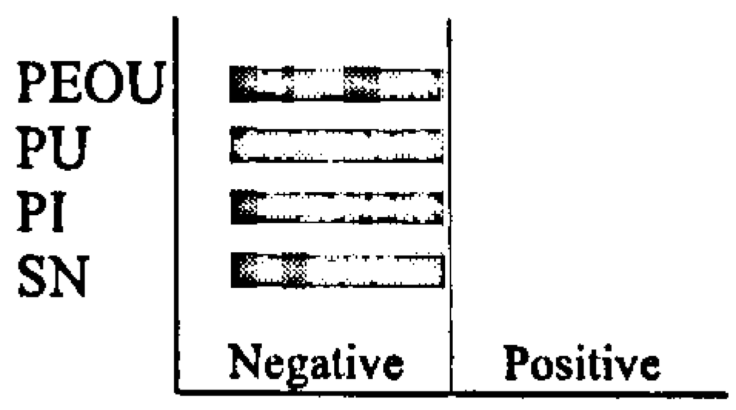
5.6.6 Comparison with the User Engagement Model

- **PEOU** Both the Helpfulness and Learnability scales were just within the acceptability limits of the SUMI, although the indication would be for corrective action. As with the users of the MM software, people said that they thought that the software was complex but could not point to any specific areas of difficulty. Even so, the perception was that the software was not very easy to use.

- **PU** The Efficiency and Control scales are both low scorers. Unlike the users of the MM software, the people using the PM software could see some useful purpose to the system but they did not perceive it as being useful to them in their job. From their point of view, they had performed the job just as well without the software.

- **PI** There had been no involvement at any time either in deciding that a software package was needed or in the choice of that package. Nor was there perceived to have been any. However, it must be said that neither did there seem to have been any expectation of involvement.

- **SN** There was no perceived management support for use of the software in the sense of encouragement or a desire for feedback to make improvements. Apart from the person who had a stake in the software, it had no champions.



Overall it would seem that none of the factors in the proposed model was present in a positive way with regard to the PM software.

The next section draws together the results from each of the foregoing studies and compares them both with each other and with the hypotheses proposed in the User Engagement Model.

5.7 Comparison of the IM, MM and PM results

The three packages just described were introduced in different sections of the same organisation with different user groups. The packages had all been in use for some time prior to the issue of the SUMI questionnaire and it is, therefore, not possible to say what the user perceptions were with regard to the software prior to its introduction. However, the SUMI results indicate the current level of satisfaction. In the absence of other measures the Helpfulness and Learnability scales of the SUMI have been used to indicate a level of ease of use and the Efficiency and Control scales have been used to indicate a level of usefulness. The informal interviews were used to determine users' sense of involvement in the choice of software and whether they felt it was useful to them. Information on the perception of management support in using the software and on the perception of how co-workers viewed the software was also obtained.

This section first discusses the results of the SUMI questionnaires and the interviews in relation to each other and to the factors generally described in the literature as affecting satisfaction. Section 5.7.2 then discusses the results by reference to the hypotheses proposed in the model of user engagement.

5.7.1 Discussion of SUMI and interview results between the packages

Table 5.10 summarises the SUMI and interview results for each package.

	IM	MM	PM
Proportion male to female users	1:1	2:1	9:1
Replaces previous software of which respondents had experience	Y	Y	N
Character based	Y	Y	N
GUI Based	N	N	Y
Tailoring to the package	Y	Y	N
Perceived management support	Y	N	N
Perceived co-worker positive attitude	Y	N	N
Perceived co-worker negative attitude	N	Y	Y
Involvement in Requirements gathering	Y	Y	N
Involvement in selection of software	Y	N	N
Perceived Involvement	Y	N	N
Average Efficiency and Control score from SUMI	55	28.5	35.5
Average Helpfulness and Learnability score	58	38	40
Training given	Y	Y	Y
Perceived ease of use	Y	N	N
Perceived usefulness	Y	N	N
SUMI Global score minus Outlier	53	31	35

Table 5.10

The literature suggests that, whilst the gender of people using software may have an effect on how they perceive it and thence on their satisfaction, there is not a direct relationship (e.g. Ramamurthy et al.,1992). The results from the three packages described lends support to this view. The proportion of male to female users for the highest scoring package was one to one. The MM software had a proportion of two male users for every female and returned a low score. If the low score were largely due to the higher proportion of male users, it could be expected that the PM score would be even lower as the number of male users was nine to one. However, although the overall score was low, it was higher than the MM software. Individually, there is no discernable difference between responses from male or female users. The scores are mixed from each gender group.

Martinko et al. (1996) suggest that previous experience of software implementations will affect user perceptions. They refer to the Learned Helplessness theory which suggests that people who have experienced implementations of software they were dissatisfied with, will have a low expectation of success with another package and are more likely to be dissatisfied with it. The users of both the IM and the MM software had experience of a previous package with which they were not happy. However, the SUMI results for the two packages are quite different. The experience of a previous software implementation with which they were not satisfied does not appear to have adversely affected the perceptions of the users of the IM software. On the other hand, the MM users were not satisfied with the earlier package and remained dissatisfied when it was replaced.

Ramamurthy et al. (1992) suggest that expertise in the skill required for the task to be performed with the software will positively influence the user's perceptions of the software as they are more likely to use it to its best advantage. However, all of the people using the PM software had been trained in how to plan and manage different projects and could be said to have a "planning" expertise. That was their specialist area. They also had good domain experience, yet they gave the PM software a low score on the SUMI. The only exception to this being user 4, who also had a personal interest in the success of the package. The fact that the software was a market leader of its type and fully met the requirements specification shows that the software was suitable for the task to be performed. A software package which helps to fulfil a specific function should be thought to be useful by the people who have to fulfil that function. Yet the SUMI score on the Efficiency and Control scales falls some way below the minimum that should be expected. The package is not felt to be useful and the reason would not seem to be that it is just too difficult (the Helpfulness and Learnability scales achieve a higher score than the Efficiency and Control scales). The feedback from the people using the software was that it did not help them to perform their task. From their point of view, they had been doing that without the help of any software. The perception was that the package had been introduced to give greater control to others.

The same feelings were reported for the MM software users. They too scored the package higher on the Helpfulness and Learnability scales than on the Efficiency and Control scales. The users of the MM software could not be said to have a particular domain expertise. The work does not demand specialist training and domain experience is more important. All users had good domain experience and they did not feel that the software helped them in the application of that experience. The perception was very much that the reason for introducing a package was to give greater control to the CM team.

A parallel may be drawn here with the study by Markus (1983) where the politics of the organisation (real or perceived) seem to have greater import than the software. In neither the case of the MM software nor of the PM software do the users see the software as primarily a tool for them.

The opposite is true for the users of the IM software, who also all had good domain experience. Although some of the comments from the interviews suggest a perception that the particular software was chosen because it was preferred by others, there is still a strong feeling that the software is a useful tool.

These results suggest that other factors are more important than domain expertise or experience for these three groups, in determining satisfaction.

5.7.2 Results related to proposed hypotheses

•H2 and H3 proposed that the presence of perceived ease of use and perceived usefulness leads to a positive disposition towards a software package.

As has been discussed, the indication from the users of the MM and PM packages was that they did not perceive the software to be very useful to them. The Efficiency and Control scores were low for both packages which reflects what was said in the interviews. The users of the IM software said that they did find the software useful and this again is reflected by the SUMI scores. In the case of both the MM and PM groups, it was perceived that the software was introduced in order to achieve a standard way of working. Neither group felt that this was particularly beneficial to them as individuals. From their point of view they had been doing the job required without the aid of the software. The software was not there to help them but for the benefit of others.

The users of all three packages were given training in the use of the software at the time it was introduced and further training in more advanced topics after its introduction. The Helpfulness and Learnability scales, used to indicate ease of use, score well in comparison to the other scales for both the MM and PM packages but are still low. In the interviews people said that the software was complex or awkward to use but were unable to point to specific difficulties. The IM software was reported to be easy to use which is reflected in the SUMI results.

In order to be easy to use, software must either be very intuitive or else easy to learn by rote. A software package supporting a fairly complex task will not lend itself to being learnt by rote as the user will need to make decisions about different actions based on the feedback from the system. An individual will find software intuitive if the actions required to complete a task match the way he would expect to work. If software is perceived to be an imposition and not a useful tool, the perception of how easy it is to use might also be reduced. These factors could be at play with regard to both the PM and MM software.

The SUMI results were poor for both the MM and PM software and indicated that neither package was perceived to be either easy to use or useful. The SUMI results for the IM software indicated that it was perceived to be both easy to use and useful. All three sets of results were supported by the interviews and general feedback and suggest support for H2 and H3.

•H4 proposed that perceived involvement in the decision process leads to a positive disposition towards a software package.

The users of the IM software were involved in the requirements gathering process but so were the users of the MM software. The users of the PM software were not consulted in any way. None of the users of any of the packages had any choice whether to use the software i.e. usage was a requirement of the job. The MM software achieved a lower overall score than the PM software which suggests that involvement in the requirements gathering process did not make people any more receptive to the software. The users of the PM software also seemed more ready to accept that the software did serve some useful purpose albeit not directly for them, than were the users of the MM software.

Only the users of the IM software were involved in the demonstrations of potential packages and given the opportunity to say which they preferred and only the users of the IM software reported having felt involved in the process of selecting the software. Only the IM software achieved a score above the mid-point on the SUMI. This supports the need for active involvement in the decision process as suggested in the literature (e.g. O'Neill, 1998; Damodaran, 1996) and supports H4.

•H5 proposed that the combined influence of management and co-workers (the subjective norm of the user) will affect the disposition of an individual towards a software package. The results of the informal discussions with users show that, in the case of the IM software, there was a perception of strong management support for using the software and that co-workers were also in favour of it. However, there is no evidence of such support in the case of the MM or PM software. Indeed, in the case of the former, the perception is that use of the software has been imposed for the benefit of the CM group and serves no other useful purpose. Although the users of the PM software agree that it has a purpose, they perceive it as an extra, imposed workload, not as something which they are encouraged to use to benefit themselves. This is similar to the case study by Hirschheim and Newman (1988) who found that a positive management attitude had a positive influence on users' perceptions of the software and is supported by Martinko et al. (1996) who argue that the attitude of co-workers and management may have a positive or negative influence. These findings support the hypothesis that the perceived attitude of management and of co-workers towards using a software package, will affect the disposition of the individual towards the package.

•H1 proposed that the characteristics of a system are not sufficient to lead to satisfaction. Research into Human Factors suggests that graphical user interfaces (GUI) with direct manipulation will be more acceptable to users (Shneiderman,1988). This would suggest that the GUI package (PM) should have an advantage over the others. All three packages met the requirements specification, although some minor modifications were made post purchase to both the IM and MM software. The changes to both packages were user driven. The IM software with a character based interface, achieved the highest overall score on the SUMI. The MM software, also with a character based interface, achieved the lowest overall score. The PM software, which had the GUI, achieved a score closer to the MM software than to the IM software. Although the Helpfulness and Learnability scales achieved a higher score than any of the other scales for the PM software, they were still only 40. This score is barely acceptable in comparison with what could be expected from state of the art software which should achieve a score of 50 or above and is below the scores achieved on these scales for the IM software. This suggests that satisfaction does not occur solely on the basis of the type of interface.

However, it is still possible that the level of perceived ease of use and of usefulness that has been inferred from the SUMI results is due to the level of usability of each package i.e. to the measurable ease of use, learnability and efficiency of the software. Just because a GUI has not been used does not mean that the software is therefore difficult. It is also possible that the differences noted in the perception of involvement and of subjective norm between the users of the IM package and those of the MM and PM packages did not affect individuals' attitudes towards the software. Although it must be said that the evidence of the interviews and general feedback points against this possibility, a further test was required in order to ascertain whether, in the case of the three packages, it was the characteristics of the software itself which had resulted in satisfaction with the IM software and dissatisfaction with the PM and MM software.

In order to compare the characteristics of the individual software packages without the potential confounding effect caused by context and so further test H1 an independent evaluation of each was undertaken. This is described in the next section.

5.8 Independent evaluation of the IM, MM and PM software

In order to remove the possible effects on the perception of the software packages caused by the context in which they were being used, an evaluation of the software itself was required. This section first describes the approach taken to obtaining an independent evaluation of each package and then the attempt made to determine the level of satisfaction for each package in different organisations. The results are discussed and reasons for the apparent discrepancy between the SUMI results and the independent evaluation are suggested with reference to the proposed model.

5.8.1 Evaluation of Software Usability

The ISO 9241-11 standard defines usability as "..the extent to which the intended goals of use are achieved, the resources that have to be expended to achieve the intended goals, and the extent to which the user finds the use of the product acceptable." (ISO/DIS 9241-11 p.V) The SUMI questionnaire measures the latter of these three i.e. the satisfaction felt by the users.

The standard also emphasises that usability can not be measured separately from the context of use. SUMI is therefore intended to be used within the work context. The results of the interviews pre and post the SUMI questionnaire and, indeed, the wide span between the upper and lower confidence limits in the SUMI itself in the MM and PM software packages, suggest that user dissatisfaction was not necessarily all due to poor software design. In order to test whether the results were more likely to be due to the characteristics of the software itself or to other factors an independent evaluation of the software was undertaken.

As has been described, the functionality of the MM and IM software packages was checked against a requirements list before purchase. In the case of the PM package, the comment often made during interviews was that it could do more than was required. It is, therefore, felt reasonable to assume that all the packages met the necessary functionality criteria. However, there had been no measurement of the software interface i.e. how that functionality was delivered. It is generally accepted that the way in which the software is presented to the user is a key factor in assessing the usability of the software itself, regardless of context of use (Hubona & Blanton,1996).

Rather than assess the software against guidelines such as those suggested by Mosier and Smith

(1986), it was felt that a more formal method should be used. The Ravden and Johnson checklist (1989) was chosen because it would enable a comparison of scores between the software packages and the results of the SUMI questionnaire. This checklist was devised from several sources in the literature on HCI (Ravden & Johnson,1989,p.20). It is intended to be used as an aid in evaluating software and detailed instructions are given on how it should be completed. There is no "satisfactory" score against which software may be compared because the evaluation method, of which the checklist forms a part, is intended as an aid to evaluating a specific software application and is not primarily intended as a means of comparing different products. It is intended that the responses to the individual checklist questions be analysed in conjunction with the overall response for each section. As explained by Ravden and Johnson (1989), there are 9 criterion-based sections, a section on usability problems encountered by the evaluator and a section of general questions on system usability. A well-designed user interface should aim to meet each of the 9 criteria which are shown in the table below. Four options are offered in reply to each question under each of the sections shown, as illustrated in table 5.11.

	Always	Most of the time	Some of the time	Never
Visual Clarity				
Consistency				
Compatibility				
Informative Feedback				
Explicitness				
Appropriate Functionality				
Flexibility and Control				
Error Prevention and Correction				
User Guidance and Support				

Table 5.11

As well as the tick box response some room is given for a general comment against the section as a whole. The questions are phrased such that a tick under "Always" is the most positive response and a tick under "Never" is the most negative response. At the end of each section there is a summary, five column, rating scale ranging from "Very Satisfactory" to "Very Unsatisfactory". Section 10 offers three possible responses to usability problems from "No problems" to "Major problems" and Section 11 is a free format response to general questions on usability. An example of part of the checklist is given in Appendix 3.

In order to carry out the evaluation, tasks must be specified which are normally performed in the daily use of the software. These tasks should be completed by the evaluator before the evaluation takes place. This means that the evaluator should be familiar with the work to be performed. In order to be able to compare the results of the evaluation of each software package with the results of the SUMI questionnaire, it was necessary that the evaluator should not be a part of the user group. At the same time, someone who was familiar with the software and the way in which it was used was needed. The only person who fulfilled these criteria was a member of the IT department. As he had been involved in the introduction of two of the software packages, it could be argued that the evaluation would be biased by a need to show those packages in a favourable light. However, the results were not to be distributed in the organisation, so would remain confidential, and it had already been agreed that the two packages which achieved low SUMI scores would be reviewed by the organisation for possible replacement. It was, therefore, felt that as objective an evaluation as is possible would be carried out.

As the questions in section 10 of the questionnaire are very subjective in format (e.g "Too many colours on the screen") and the questions in section 11 needed a free format response, neither was completed for any of packages.

In order to obtain a score for the questionnaire, each column in the options was given a scale rating. The most positive ("Always") was rated as 1 and the most negative ("Never") was rated as 4. The number of rows in each section was multiplied by the rating assigned to each column to give a possible total score per column per section. In the instance where the evaluator had written N/A against a question, that row was not counted. If the evaluator had not ticked any column nor written N/A, the row was still included as part of the possible score. This occurred in only one instance.

The number of ticks in each column was then added and multiplied by the rating for that column to give a total actual score per column per section. As the first column, "Very Satisfactory" had been given a rating of 1 and the final column, "Very Unsatisfactory", given a rating of 4, a low overall actual score would indicate satisfactory rather than unsatisfactory. However, a numeric score on its own could be misleading as one section might score very well and thus obscure poor scores from other sections. The results were, therefore, also checked against the number of ticks under each of the headings in the summary sections. Generally, there was consistency between the total scores and the summary evaluation. However, sometimes, the middle i.e. Neutral column, had been ticked. This was the case in three instances for the IM package, once for the MM package and not at all for the PM package. Where "Neutral" had been ticked, there was a fairly even number of ticks under both "Moderately Satisfactory" and "Moderately Unsatisfactory" for that section.

Once a possible and actual total had been calculated for each section, these were again summed to obtain a total possible and total actual score for "Very Satisfactory", "Moderately Satisfactory" etc. across all nine sections.

Table 5.12 shows the total possible and total actual scores for each of the possible options across all sections for each package. The total possible scores vary between the packages because of the areas which were marked Not Applicable. These are areas such as reference to a menu driven system when menus are not used. The packages were each based on a different operating system using different interfaces so a like for like comparison is not possible. However, it is possible to compare how well each package scores in relation to the lowest possible score it could obtain. A low score being indicative of a satisfactory evaluation. The row labelled Ticks shows the number of ticks under each heading from the summary sections.

The total score for the IM package was 257. As the total possible score for the "Moderately Satisfactory" column was 222, and the total possible for the "Moderately Unsatisfactory" column was 333, the overall evaluation of the package would suggest that it is slightly less than moderately satisfactory. This is supported by the number of ticks in the summary evaluations for each section. There were only three ticks under Moderately Satisfactory, three ticks under Neutral and three ticks under Moderately Unsatisfactory.

The total score for the MM package was 229. The total possible score for "Moderately Satisfactory" was 244 and for "Very Satisfactory" it was 122. This suggests that the overall evaluation of the package places it slightly above moderately satisfactory. Again the ticks under the summary evaluations for each section support this outcome. There was one tick under Very Satisfactory, seven ticks under Moderately Satisfactory and one tick under neutral.

The total score for the PM package 222. The total possible score for "Moderately Satisfactory" was 252. The total for "Very Satisfactory" was 126. Again, this suggests the overall evaluation of the package is somewhat above moderately satisfactory. This result would seem to agree with the software market at the time, where it was rated as one of the best packages. The Very Satisfactory column achieved five ticks and the Moderately Satisfactory column achieved four ticks on the summary evaluations.

IM Software	Very Satisfactory	Moderately Satisfactory	Neutral	Moderately Unsatisfactory	Very Unsatisfactory
Possible	111	222		333	444
Actual	14	104		111	28
Ticks		3	3	3	
Actual Grand Total	257				

MM Software	Very Satisfactory	Moderately Satisfactory	Neutral	Moderately Unsatisfactory	Very Unsatisfactory
Possible	122	244		366	488
Actual	27	150		36	16
Ticks	1	7	1		
Actual Grand Total	229				

PM Software	Very Satisfactory	Moderately Satisfactory	Neutral	Moderately Unsatisfactory	Very Unsatisfactory
Possible	126	252		378	504
Actual	66	119		33	4
Ticks	5	4			
Actual Grand Total	222				

Table 5.12

According to the independent evaluation, therefore, the PM software achieves the most satisfactory rating, followed by the MM software and finally by the IM software.

5.8.2 Comparison of the evaluation results and the SUMI

Table 5.13 shows the comparative scores between the results from the independent evaluation and those produced by the SUMI.

The results from the SUMI are such that anything falling below 50 also falls below the expected satisfaction level. Anything with a Global score below 40 should be regarded as poor software.

	Independent Evaluation	SUMI
IM software	Low Moderately Satisfactory	Above Average
MM software	Moderately Satisfactory	Well below average
PM software	High Moderately Satisfactory	Below average

Table 5.13

It is possible for a package to achieve a Global score above 50 whilst still falling below 50 on one or two scales. In such cases the overall package is acceptable but the areas falling below 50 should be considered as worthy of further investigation and remedial action. In the case of both the MM software and the PM software, they both fell below 50 in all sections of the SUMI and only achieved a Global score of 31 and 35 respectively, excluding Outliers, and of 33 and 38 including the Outliers. Yet in the independent evaluation they both score Moderately Satisfactory or above. The IM software, on the other hand, achieves a Global rating in SUMI of 54, yet the independent evaluation places it at the low end of Moderately Satisfactory (though not actually Unsatisfactory).

Both the MM and the PM software were PC based. The PM software had a graphical user interface. The IM software was character based and ran on a mini-computer. The results of the independent evaluation support the view that GUI software tends to be more usable in the sense of being easy to use.

The results of the independent evaluation compared with those of the SUMI analysis suggest that factors other than the software itself are influencing the responses by users and their reported feelings towards the software and, therefore, support H1.

5.8.3 Evaluation by reference to other organisations

The results obtained from the independent evaluation were for three different packages but within a single organisation. In order to further test whether the context of use was influencing the satisfaction felt by the users it was necessary to obtain the satisfaction scores for the same packages but from different companies. If the scores were all similar, the reason could still mainly lie with the software.

An attempt was made to contact other companies using the different packages, who would be willing to co-operate. However, in the case of the PM software, there was an unwillingness by the supplier to release any customer contact names. The supplier was in a take over process at the time which might have had an effect. In the case of the IM software, it transpired that there were no longer any other users of the software. It was said that people were moving away from mini-based systems and having to change software as a result. Several contacts were obtained for the MM software but only one was willing to co-operate and there were only five users.

Five SUMI questionnaires were sent to this organisation with a message to accompany each one to explain what the SUMI was and how it should be completed. All questionnaires were returned. It was not possible to gain any information regarding the management structure of the organisation (although it was a Public Sector company), nor was it possible to establish what the user involvement had been in deciding upon and introducing the software.

The global score for the software was 50 from the SUMI. The Efficiency scale achieved a score of 56 and was the second highest scorer. This compares with the score from the in-house group, which was the lowest scoring scale at 28. However, there were only five respondents and there was a wide span between the Upper and Lower Confidence intervals. With such a small group the results must be considered with caution. They do, however, support the findings of the independent evaluation for this package.

5.8.4 The influence of context

The difficulty described in finding organisations using the same packages in order to test for satisfaction, means that there is no wider support for the findings of the independent evaluation. However, the positions of the packages in the software market - particularly the PM software - lend support to the hypothesis (H1) that the characteristics of the software alone are not sufficient to ensure satisfaction.

The ISO 9241-11 standard and the Common Industry Format (CIF) stress that software usability is also dependent on context. Context is defined as: user; task; equipment; and environment (ISO/DIS 9241-11 section 5.1.3). These are now each considered in turn:

user - the people using the three packages were of a similar spread in age and capability.

(Capability being measured by the grade of the post they held). The proportion of male to female users was equal for the IM software. There were more male than female users of the PM software which achieved a low score. However, if gender were a key factor in the low satisfaction rating, it could be expected that the score for the PM software would be even lower as the users were predominantly male, and this was not the case. Information on previous experience with computer systems was only gathered from a few people and these were users of the MM and IM software. It is not, therefore, possible to compare the level of experience across the groups. However, all the people in all three groups who completed the SUMI did have experience of the type of work before the introduction of the packages, be that with a computerised or manual system. Nor was data gathered on the educational qualifications of the people involved. However, there was a minimum requirement for employment for everyone. Members of the PM workgroup were in a higher grade post than those of the IM or MM workgroups but, as with gender, if this made a difference it could be expected that the PM software would be at either end of the scale and not in the middle.

The **tasks** were necessarily different but the packages each met the requirements specification and had been used by other organisations.

The **equipment** in use was of a similar specification. Although the IM software was run on a mini-frame, PCs using terminal emulation software were used to access it. These PCs were of the same specification as the PCs used for the MM and PM software as other PC based packages were used as well as the IM software by the IM users. The response time for each of the packages was not measured. The IM software used what is known as block mode, so data on a full screen is processed rather than record by record. The MM software validated record by record as did the PM software. A direct comparison between the three is not, therefore, possible. The results of the SUMI indicate that some users of the MM software thought the response time too slow. It is not known whether this is factual or perceptual, but it should be noted that the evaluator using the Ravden and Johnson checklist did not remark it as a problem with regard to any of the packages.

With regard to **environment**, the organisation operates a strict health and safety policy which has regard for seating and lighting etc. in offices. The physical context was, therefore, similar for the use of each package.

Environment may be used to mean the organisational type and culture as well as the physical environment. As all three packages were being used in the same organisation and, indeed, the IM and MM packages were being used in the same parts of that same organisation (although not by the same people), the organisational type and culture do not seem to have been a key factor. Indeed, the MM software achieved a higher SUMI score in another organisation where the culture was likely to have been similar (it having been in the Public Sector). However, although the overall organisational context was the same for the users of each package, the perceived management support for each was not. Nor was the perceived involvement in the decision regarding the obtaining of a software package the same. These latter are two key areas of difference between the package deemed to be satisfactory and the two which weren't.

The next section summarises the results of the work described in this chapter.

5.9 Summary of initial model testing

In the case of the IM software, there was a perception of strong management support. One manager was referred to and he sought user opinion and explained management decisions. There was a strong sense of direction in how the software was to be used. The users of the software felt involved in the whole process of introducing it, including the choice of package.

This strong management support did not exist in the case of the MM users. Their line managers took a back seat role with regard to the software and continued to operate as if it did not exist. The need for the software was determined by senior management and the CM group had responsibility for producing information from it. There had also been some disagreement within the CM group as to how the system should be operated which led to a sense of uncertainty amongst the users of the software. There was no sense of leadership in the use of the software. Despite having been asked what they would like to see in a new package, there was no sense of involvement on the part of the users. The decision on which of the potential packages to purchase was made by others without reference to them.

In the case of the PM software, there was an understanding that the software was needed. Each project team had targets to meet but some of the resources had to be shared and conflicts had to be resolved. However, there was a sense of coercion rather than support in its use. People were being forced into a standard way of working by the software and there was a view that this was causing unnecessary extra work. No-one had spoken to them about what they would like to see in the software or what would be useful to them. There was no sense of involvement at any stage.

It may be seen that there is a slight difference between the users of the MM and the PM software. Although both groups seem to resent the imposition of software they did not ask for, in the case of the MM software, the view was expressed that the software was of no use to the people operating it whereas, in the case of the PM software, there was a grudging acceptance that a tool was needed in order to assist in the sharing of resources. However, there was a sense that a different PM tool could have been selected, and that the one in use was also there to impose a way of working.

The results of the SUMI questionnaire on the Helpfulness and Learnability scales suggest that users found the IM software to be easy to use but that the users of the MM and PM packages did not consider the software easy to use. The results on the Control and Efficiency scales also suggest that the IM users thought the software useful in the tasks they had to perform, whereas the MM and PM users did not rate the software as useful.

These results all support the hypotheses proposed in Chapter four with regard to software which is already in use, by tracing back from the current level of satisfaction to the reasons for that satisfaction. The independent evaluation of the software suggests that the level of satisfaction can not be explained by reference to the software alone. Whilst factors other than PEOU, PU, PI and SN may be at play in influencing how people view software, it is these factors which are key.

The work reported in this chapter indicates support for the User Engagement Model with respect to software already in use. The next stage is to determine the influence of the factors in the model prior to the introduction of a software package. In order to do this some means of assessing user engagement and expressing it as a formal metric is required. It was decided that a questionnaire would be the means by which this could be achieved. The next chapter describes the development and initial validation of a questionnaire in order to measure people's perceptions of and, thereby, engagement with a software package prior to its introduction.

Chapter 6

Development of the User Engagement Questionnaire

The previous chapter demonstrated that there was evidence to support the model proposed in Chapter 4 by comparing the current satisfaction level with three different packages, as measured by SUMI, with the results of an independent evaluation of the software, as measured by the Ravden and Johnson checklist (1989) and by tracing back the process and surrounding context which led to the introduction of the software.

The results suggest support for the hypotheses that the perceived ease of use (PEOU) and perceived usefulness (PU) of the software, combined with perceived management and co-worker support (SN) for its use and perceived involvement (PI) in its selection, contribute towards later satisfaction with software. Perceptions of ease of use and usefulness were assessed by considering the scales in the SUMI questionnaire which correspond, in part, to ease of use and usefulness. These findings are similar to those of Davis (1989) who demonstrated that software usage could be predicted according to whether people judged the software to be easy to use and useful. Further studies also support this theory (Davis, 1993; Davis et al.1989; Wiedenbeck & Davis S.,1997; Hubona & Blanton, 1996) The evidence in support of the need for perceived management and co-worker support and perceived involvement is based on informal discussions and general feedback. There is also support in the literature for these constructs (e.g Martinko et al.,1996; Damodaran, 1991; Kappelman & McLean,1991).

There is evidence in the literature to suggest that the strength of user engagement with the introduction of new software will have an effect on the satisfaction the user will feel with the software itself (Kappelman and McLean, 1991). By "user engagement" a favourable disposition towards the software is meant, in that people see its introduction as something worthwhile and of benefit to them. According to the model, it is PEOU, PU, PI and SN which together make up this favourable disposition or User Engagement.

The combined strength or weakness of PEOU, PU, PI and SN were observed to correlate with the level of satisfaction as measured by SUMI. However, no formal metric was used to show the strength of that correlation.

The question also remains as to whether the presence of these factors may be measured *before* the introduction of a software package and whether the level of user engagement they show may be used as an indicator of later satisfaction with the package. In order to more formally assess the relationship between satisfaction with software and the combined effect of PEOU, PU, PI and SN, a measure of that effect (termed "user engagement") was necessary. This measure could then be correlated with the level of satisfaction with the software, as measured by SUMI, after it had been in use for some time. As discussed in Chapter 5, one of the most used methods of obtaining a formal metric, is the use of a questionnaire. The results of the questionnaire should not then be considered in isolation from the people who completed them and any interpretation must be made with caution, but they do present a means of providing measurements which may be compared.

6.1 Development of the user engagement questionnaire

A single questionnaire to measure each of the factors in the model did not already exist as the measurement of User Engagement with packaged software prior to its introduction had not previously been undertaken. Different questionnaires were referred to for examples of items measuring the different constructs required.

- **Perceived Ease of Use and Perceived Usefulness Constructs**

TAM (Davis et al., 1989) is one of the best known models which predicts use of software by measuring perceived ease of use and usefulness. The questionnaire used to demonstrate the predictive power of this model has been through several iterations and has been tested by people other than its creator (Mathieson, 1991; Taylor & Todd,1995; Adams et al.,1992; Hendrickson et al.,1993). Although TAM itself predicts future use not satisfaction, the constructs of perceived ease of use and usefulness are the same as those in the proposed model of User Engagement. The questionnaire consists of two sets of questions. The first set relates to perceived ease of use and the second to perceived usefulness. Davis (1989) uses six questions to measure each construct. However, in Davis et al. (1989) this is revised to four items per construct. The original questions were constructed such that specific software could be named within the statements without altering the format of the statements themselves. For the purpose of the questionnaire being developed, all eight questions could be adapted by substituting the name of the software package in question in place of that used by Davis et al. (1989). The statements were grouped by ease of use and usefulness in the same way as had been tested by Davis (Davis, 1989). Although it has been suggested that grouping statements in this way may result in people just ticking down a line, Davis and Venkatesh (1996) demonstrate that item grouping has no significant effect on the level of reliability or validity of the TAM scales.

• Subjective Norm Construct

TAM does not take account of SN, the argument being that it is difficult to distinguish to what degree it is present indirectly in Attitude (Davis et al.,1989, p.985). According to TAM, Attitude is one of the predictors of Behavioural Intention (BI) which then predicts usage. Variables such as political influences and organisational structure may be classed as "external variables" which may influence perceptions of ease of use and usefulness (Davis & Venkatesh, 1996). The model itself predicts usage. Although amount of usage may be used as an indicator of degree of satisfaction when such usage is voluntary, this is not the case when usage is a requirement. When usage is required, the influence of referent others is more likely to be important. This is recognised by Davis et al. (1989) who say that the lack of evidence for the SN factor in the tests with TAM may be at least partly due to the fact that the tests were undertaken for a single-user system and with people who were using the software voluntarily.

One of the principal factors of TRA (Fishbein & Ajzen, 1975) is the SN construct. It is also a component of the Theory of Planned Behavior (TPB, Ajzen,1985). In their comparison between the predictive strengths of TAM and TPB, Taylor and Todd (1995) break down SN into beliefs concerning peers and superiors in the context of a university rather than use the more general SN questions used by Mathieson (1991), which refer to "people who are important to me" and "people whose opinions I value" (p.190). This follows Fishbein and Ajzen (1975) who state that different referents will be important in different situations. In the work context, it is the people who work with an individual and the person to whom the individual reports (the line manager) who are the important referents (Martinko et al.,1996, Hirschheim and Newman, 1988). The questions relating to SN were, therefore, adapted from those used by Taylor and Todd (1995) substituting "line manager" and "co-worker" for "professor" and "classmate" respectively.

The subjective norm "is the person's perception that most people who are important to him think he should or should not perform the behavior in question" and " the person's motivation to comply with those expectations" (Fishbein and Ajzen, 1975 p.302). Two questions were, therefore, formulated for each referent group. The first question is to determine the individual's belief about the wishes of the referent (co-worker or line manager) and the second question is to determine the importance of those wishes to the individual. It is important to specify what the wishes of the referent are believed to be (Fishbein and Ajzen, 1975). In this case, use of the software in question is specified. Subjective norm is determined by multiplying the first question by the second and summing the totals of all referent groups used.

The questions to determine SN are concerned with predicting behavioural intention. The questionnaire being designed was intended to be used as an indicator of likely satisfaction and, therefore, of an attitude rather than an intent. However, it is the combined effect of the perceptions regarding ease of use, usefulness, involvement and the beliefs of referent others that leads to the disposition towards satisfaction or user engagement with the software. The SN construct is important as an indicator of what the individual believes about referent others. If they are thought to believe that the software should be used and if their opinions matter to the individual, this would act as a positive influence on the individual's engagement with the software. Where in TRA this positive influence would result directly in an intention to behave in a certain way, in the proposed model it would result in an individual's willingness to engage with the software i.e. to have a positive disposition towards it.

The statements relating to SN are necessary because they help to show what the individual believes about the opinion of the referent other with regard to the software. There will be no choice whether the software is used or not. Therefore, if the referent other is thought to believe that it should be used, there is some perceived support for its use. The second question in each group (co-worker or line manager) gives an indication of whether the individual will be swayed by the beliefs of the referent other.

- **Perceived Involvement Construct**

Constructing the statements to assess user involvement posed more difficulties than those of ease of use, usefulness and subjective norm. As has previously been discussed, there are two aspects of user involvement:

a) that of the user having the opportunity to have a say in the development of a system (which may take place in varying degrees) or, in the case of packaged software, a say in the system to be purchased.

b) the extent to which the user wishes be involved in decisions regarding the system and whether that wish is met (Doll & Torkzadeh, 1989).

Perceived Involvement has been used to refer to the combined effect of these two.

Questionnaires which ask about involvement often do so by use of a single item construct (e.g. "Users' feelings of participation", Baroudi & Orlikowski,1988, p.57). Doll & Torkzadeh (1989) use an eight item construct because they are finding out whether there was any actual involvement and whether the user feels that it was sufficient. In the case of the research reported in this thesis the amount of actual involvement i.e. whether the user was referred to in the establishment of the requirements list; whether the comments expressed were noted; and whether the user was present when the software was chosen and had a voice in that decision would all be recorded. It was, therefore, only necessary to establish whether the user *felt* that he had been involved. This led to the creation of the item "I felt sufficiently involved in the decision process". Even so, it was felt that this item should be cross checked by asking whether the user would actually have liked to have been more involved in the process. Finally, a third statement "I had no involvement in the decision process" was used to see whether the individual's understanding of what is meant by involvement included physical presence at meetings and the opportunity to express an opinion. Such a statement could only be used, of course, in the case where the actual physical involvement of the user was known but it was felt that advantage could be taken of that fact in this case. This third statement was not to be used in the calculation of the overall result of the questionnaire. An example of the full questionnaire is shown in Appendix 4.

6.1.1 Scoring the questionnaire

In all, there are 15 questions in the questionnaire. A 7 point Lickert scale with possible responses ranging from Strongly Agree to Strongly Disagree is used. A Lickert scale approach was chosen because this is considered a natural way of eliciting opinions about a software product (Kirakowski, 1995b). Responses which Strongly Agree with the question denote a positive response in all cases with the exception of the PI construct, which will be returned to later. Strongly Agree is scored as 1 and Strongly Disagree is scored as 7. This means that a low overall score denotes a positive feeling.

The phrasing of the second item of the PI construct - "I would like to have been more involved.."- means that the results have to be reversed to achieve the true score (i.e. a Strongly Disagree response to the question actually denotes a positive response).

For each of PEOU, PU and the first statement of PI, a tick in any of the first three boxes on the scale would represent a positive response. The fourth box indicates uncertainty so may not indicate either positive or negative feelings. A tick in the final three boxes indicates a negative response. In the case of the second PI statement the ticks and scores are reversed as explained above. Therefore, for each construct of four items for PEOU and PU, a score below 16 is a positive response and a score above 16 is a negative response. A score of exactly 16 indicates uncertainty or non-committal. In the case of PI, consisting of just two items, a score below 8 indicates a positive response and above 8, a negative response (assuming the score reversal mentioned above).

Subjective norm is determined by beliefs about the wishes of referent others multiplied by the motivation to comply with them. This means that, whilst there are four questions regarding attitude towards co-workers and management, only one score is obtained. The scales for SN are slightly more complex than the others because they depend on the responses being multiplied out. The reason for this is that the construct of SN depends on the strength of the belief about the referent other multiplied by the strength of the motivation to comply. This means that there is an overlap between the possible negative and positive response. For example, a tick in any of the first three boxes to the statements "My co-workers would think I should use xxxx", and "Generally speaking, I want to do what my co-workers think I should do", would attract a score of anything between 1 and 9. Such responses would be deemed favourable in regard to the software. However, it would also be possible to tick boxes 5 to 7 in response to the first question and boxes 1 to 3 in response to the second question. Such responses would be unfavourable towards the software but the scores would be from 5 to 21. Scores between 5 and 9, therefore, could be positive or negative as illustrated by figure 6.1.

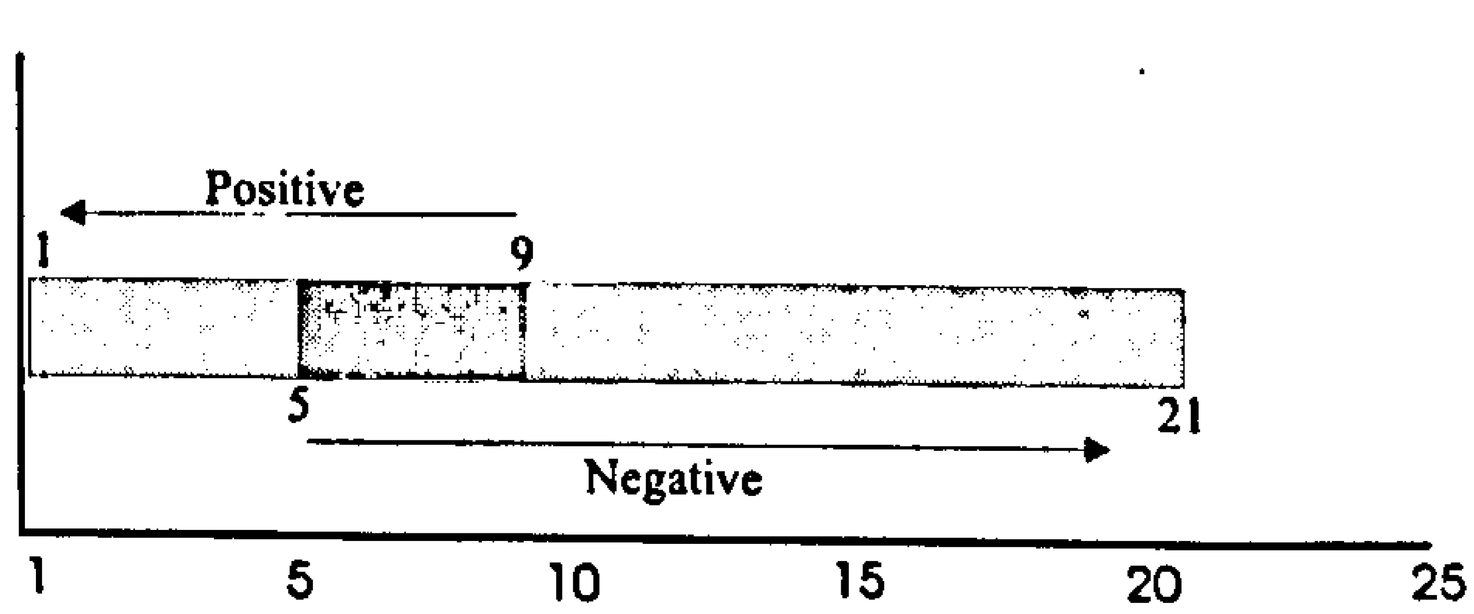


Figure 6.1

This means that for the combined score (("my co-workers would want.." x "..I want to..") + ("my line manager would want.." x "I want..")), anything from 10 to 18 inclusive would represent a response which could be either positive or negative. Above 18 would always be negative and below 10 always positive. Although someone may believe that a referent other does not think the software should be used but may equally not wish to comply with the perceived wishes of that referent other, that does not mean that the opposite action would be taken. Scores achieved as a result of these markings indicate a lack of positive support for, or engagement with, the software.

The lowest (most positive) possible score that could be achieved on the questionnaire, by always ticking Strongly Agree, is 12 and the highest (most negative) that could be achieved by always ticking Strongly Disagree, is 168. Taking account of the grey area which occurs as a result of the SN construct, a total score of 39 or lower indicates a positive feeling towards using the software. A score falling between 40 and 60 inclusive is grey and could be positive or negative (although, the higher up the scale, the more likely to be negative). A score of 61 and above is most likely to be negative.

The questionnaire was given the title ITU in order to make it easy to refer to. This is really a misnoma as it stems from Intention to Use. People who completed it were never told what the letters stood for.

The ITU itself could only be regarded as a guide and would serve to re-enforce the results of interview and observation. However, to serve as such a tool, it must be demonstrated to have validity.

6.2 Validity of the User Engagement questionnaire

All the statements used in the final questionnaire, apart from those relating to PI, had been previously tested for content and construct validity and for reliability. However, the questionnaire as a whole was new and the PI construct had not been previously tested.

Content validity refers to the "representativeness or sampling adequacy of the content" (Ives et al.,1983 p.788). This was assured for the statements relating to PEOU, PU and SN as they were the result of previous validations (e.g. Davis, 1989; Taylor & Todd, 1995). However, the statements relating to Involvement had not previously been tested for content validity. Content validity may be tested by the simple means of asking people what they believe the statements to mean (Doherty & King, 1998). A further test is to perform an interitem correlation within each construct. In order to perform this latter, of course, the questionnaire must have been completed.

In order to test that the statements were clear and were measuring the construct intended, a number of people (two of whom had previously had experience of questionnaire development) were asked what they understood the meaning of each item to be. Their responses agreed with what the items were intended to measure. The only exception was the first item on user involvement. It was suggested that rather than "I felt sufficiently involved in the decision process", the item should be re-phrased to read "I was involved in the decision process". Making this change would mean that the first item was the exact opposite of the third item. However, this change was rejected. The intention of the first item is to establish whether the person **felt** involved. The intention of the third item is to check the response against the known physical involvement.

Two measures are possible to assess **construct validity**. The first is to correlate each scale with the total score minus that scale (Ives et al.,1983, Baroudi & Orlikowski,1988). This is considered the weaker measurement of construct validity as it assumes that the total score is valid (Ives et al.,1983). The second method is to perform a factor analysis which reveals the underlying structure of the measure (Ramamurthy et al.,1992, Davis et al.,1989). Both of these require the questionnaire to have been completed in order to provide the data.

By "**reliability**" it is meant that a questionnaire is free from measurement error or is internally consistent. Two methods of testing for reliability are 1) to replicate the administration of the questionnaire under different conditions and 2) to administer it to the same people (test-retest) on multiple occasions (Hendrickson et al.,1993). This would demonstrate that the measure is able to provide consistent results. A third method is to use Cronbach's alpha to measure for internal consistency. In the first instance of testing for an instrument's reliability, Cronbach's alpha is a commonly used measure (e.g. Davis, 1989; Ramamurthy et al., 1992; Baroudi & Orlikowski, 1988). If the items have no error and measure the true score then the alpha will equal one. If the items are unrelated, then the alpha will equal zero. Generally a score above 0.7 is acceptable (Davis et al., 1989). Again, data must be available in order to perform this test.

•Construct validity of the questionnaire

Four constructs are measured within the questionnaire - PEOU, PU, PI and SN. The total of the questionnaire itself measures the construct termed User Engagement, by which is meant a favourable disposition towards software which is not yet being used. Ramamurthy et al.(1992) describe one aspect of construct validity as being the measurement of the "degree to which multiple attempts to measure the same concept through maximally different methods are in agreement" (p.487). Hence the usefulness of a factor analysis, which demonstrates the similarity of results of different means of measuring the same construct within a questionnaire. However, in order to check the validity of the User Engagement construct which is the result of the combination of all the factors, another measure of the same construct was needed. This means that another measure of the same concept must already exist. In testing the validity of the measurement of User Information Satisfaction (UIS), Ives et al.(1983) make use of a four-item general measure of satisfaction which they correlate with the results of the Pearson questionnaire. Baroudi and Orlikowski (1988) use interviews to measure the level of satisfaction and correlate these results with the short-form measure of UIS. In the case of the ITU, a well validated measure of satisfaction with software existed in the form of the SUMI. A correlation between the SUMI and the ITU would suggest that the ITU was measuring a similar construct to that of SUMI i.e. satisfaction. SUMI measures satisfaction with the use of existing software whereas the intention of the ITU is to measure the disposition of the user to feel positively towards the software, and hence the likelihood of feeling satisfied with it. A correlation between the two would show that the ITU was measuring the same general area (Anastasi, 1976). The correlation should be "moderately high, but not too high" (Anastasi,1976,p.153)

In order to test for validity of the statements for each construct in the ITU, to test for the reliability of those constructs and to test the validity of the questionnaire as a whole, it was necessary to gather some data. Doing this was problematic. As it was intended for use prior to the introduction of a software package, it had to be tested in this context. Assistance from external organisations was sought.

6.2.1 Contacting other organisations

It was decided that a face to face approach would be better than unsolicited e-mail. A conference attended by a number of IT professionals was chosen and representatives from five large organisations were spoken to. A large government body was also contacted. However, of these six contacts, only two were both at the point of introducing packaged software and willing to take part in the research. Following the initial face to face contact, all correspondence was via e-mail and post. A software supplier was also contacted. This company also provided training in the software it sold. For ease of reference, these organisations will be referred to as A, B and C. What happened with each is described below.

6.2.2 Method used at each organisation

- **Organisation A**

This is a large multi-national and was in the process of selecting software for introduction into one of its UK subsidiaries. Some elements of DSDM were being used in the process in order to involve the users. The response to the request to issue the ITU was favourable. However, despite the size of the organisation, the user group was small consisting of just ten people.

As the name of the replacement software was not yet known, the ITU was modified to read "the new software" in place of a specific name. Ten questionnaires were sent to the contact with notes to be given to the respondents which explained what the questionnaire was for and to thank them for their assistance. These were to be issued after all the users had exposure to the proposed software.

It was not possible to interview the respondents so the only information available is from the contact at the organisation. From this it would seem that all the users had experience of the software which was being replaced and that there was no choice as to whether it would be replaced. User involvement followed the DSDM format of selecting one or two people to be "ambassador" users. These people were involved in meetings and discussions and spoke on behalf of the user population. The full DSDM methodology was not used as the software was not being developed in-house.

It was agreed with the contact that the SUMI questionnaires would be sent about six months after the ITUs, by which time it was estimated the software would be fully implemented .

- **Organisation B**

This organisation is also a part of a multi-national corporation, although fairly independent of the parent body. All the information detailed below was provided by the contact as a result of a series of question/answer e-mails. He was carrying out a research project himself and was a sympathetic correspondent.

The organisation is hierarchical in structure having four tiers from top to bottom. The staple work is in the engineering field and there are seven sites in the UK. Over half the staff are male and about 85% are in the mid twenty to mid forty age range. Senior management had decided that a computerised system would be beneficial in managing their documents as part of the Quality Management System. The actual software to be used had already been selected. When approached, the organisation was about to introduce the software to 650 staff and it was estimated that a response could be expected from about sixty people.

The role of the contact in the project was that of a key user working with a consultant to introduce the software.

The software selected was a large package and designed to be flexible to meet the needs of different organisations so various design decisions had to be made pre implementation. It was intended that all users would be fully involved at every stage of determining the software interface and deciding how it would be used. As there were so many staff, representatives were selected by each department. In all, 28 people were selected from the user level to be involved. At the start of the project there was one senior manager who was reported to be enthusiastic and who ensured that there was management support through the organisation.

Twenty-eight meetings, the main purpose of which was to agree how the software should be implemented in the organisation, were held over a period of eight months before the software was brought into the organisation. The user representatives were not forced to attend the meetings with the result that several did not. Only the first two workshops had an attendance of all 28 representatives. After that the largest attendance was 18 people at a workshop which took place about half way through the project.This lack of attendance was largely due to the fact that people had not been given specific time in which to do so i.e. their workload was not reduced to allow for the meetings and they were still required to meet timescales. As the contact reported, the result was inevitably that neither they nor the people they were representing were fully informed of what was happening nor contributed to the decisions.

The contact was asked to distribute the ITU after people had seen the first overview of the completed system and to distribute the SUMI about three months after the software had been in use.

- **Organisation C**

This organisation was contacted as a means of reaching a larger number of people who would all be at the initial stage of using the software. Only people who were on introductory courses were asked to complete the questionnaires and it was unlikely that they had had much prior experience of the software. (This was checked with the training provider as well, although it can't be certain that someone might falsely claim to being a novice user).

Although the main purpose of the exercise was questionnaire validation, it was felt that some background information would be useful. Unfortunately, this too had to take the form of written questions and, therefore, was kept brief so as to avoid the problem of people feeling that it would take too much time to respond to all the questions and, as a result, respond to none.

The questions asked were:

a) Is the company you work for in the public or private sector?
b) Do you have to use the software as part of your job?
c) Approximately how many people are there in the company you work for?
d) Approximately how many other people use the software in the company you work for?
e) Would you recommend the software to someone else needing to plan projects?
f) Do you feel that there is management support and encouragement for your use of the software?
g) Have you experience of other software to perform the same functions?

The second question was necessary in order to filter out any people who would be using the software voluntarily and, therefore, not belong in the category under study. The fourth and sixth questions (d and f) were to establish whether there was likely to be co-worker influence and would also indicate whether the respondent felt that there was management support. The fifth question (e) was used to indicate if the person seemed to already have a favourable disposition towards the software. The first, third and final questions (a,c,g) were not relevant to the ITU questionnaire itself but, it was felt, would help to place the respondent in a general context.

Organisation C was provided with thirty packs, each of which consisted of:

i) a note to the respondent
ii) the seven background questions
iii) the ITU
iv) the SUMI
v) a stamped addressed envelope

Each of these was numbered as a pack so that they could be collated even if the respondents chose to remain anonymous. The note thanked the respondent for his co-operation; gave an explanation of the purpose of the questionnaires and the origin of the SUMI; requested that the ITU be completed either during or immediately after the training course and that the SUMI be returned as soon as some familiarisation with it at work had been gained, preferably no longer than four weeks after returning. This rather short timescale was given in order to avoid the problem of people forgetting about it.

The arrangement with organisation C was that they would ask for willing participants and hand out the packs at each training session. The trainer would collect all completed ITUs and additional questions at the end of each course. This would be over a period of two months. Once the number of packs ran low, more would be requested from the author.

At the end of the period no extra packs had been requested.

6.2.3 Results from each organisation

- **Organisation A**

Eight out of the ten ITU questionnaires issued were returned. The people who completed the ITU all wrote their name on it from which it is possible to say that there were definitely four females and two males. The remaining two used initials only so could be either male or female.

Although the ITU was administered and returned as agreed and the contact in the organisation said he would be happy to administer the SUMI questionnaire about 6 months later, these were never returned. There was an initial communication from the contact to apologise for the delay but nothing then happened and two e-mail requests have gone unanswered. In such a situation, not knowing if the person has moved companies or is suffering from illness, it is not possible to push for a response.

- **Organisation B**

It was not possible for the contact to distribute the ITU questionnaire. The consultancy firm which assisted with the preparation for going live with the software had itself issued a questionnaire. It was felt that to issue another one during the same time frame would cause annoyance to people and possibly cause ill-feeling.

The SUMI questionnaire was distributed after the software had been in use for about three months. By this time, the senior manager who had been enthusiastic about the introduction of the software and had ensured that those responding to him had also shown some support, had moved on.

Of the 60 questionnaires (each of which had been distributed with a written explanation of its purpose and thanks to people for taking part) only 27 were completed and returned. Ten of the completed questionnaires were from people who had been selected as user representatives.

Once it was confirmed that there would be no more completed SUMI questionnaires, the data was keyed in and the analysis program was run. The results of this were sent to the contact as had been agreed. They will not be discussed here but are referred to in Chapter 7.

- **Organisation C**

Twenty-four of the ITUs and additional questionnaires were returned. However, only ten SUMI questionnaires had been returned after 3 months. Analysis of the additional questions is shown in table 6.1.

	Yes	No	Unknown
Private sector organisation	18	5	
Mandatory use of software	18	5	
Organisation >600 people	13	8	2
Number of co-users >5	12	6	5
Recommend the software	18	0	6
Management support	19	2	2
Prior experience	20	3	1

Table 6.1

Some of the totals in the above table only add up to twenty-three because one of the respondents was unemployed and the questions were not applicable to him. The organisational size actually ranged from 120 to thousands and the number of co-users ranged from zero (in just one case) to 200. The fact that some people did not know the size of their organisation or the number of co-users (they put a "?" by these questions) suggests that they were new to the organisation or, at least, to that section of it.

6.2.4 Individual construct validation

Although the software being introduced into organisation A was quite different from that used at organisation C, the two sets of ITUs were combined in order to perform the statistical tests. The five ITUs from the people who were voluntary users and the one from the person who was unemployed were removed as the research is concerned with mandatory use of software. This gave a total of 26 questionnaires on which to perform the statistical validation for content, construct and reliability. Of the people who were voluntary users, three had returned the SUMI questionnaires. That meant that the results of only seven ITU and SUMI questionnaires could be correlated.

The Unistat package was used to perform all the calculations mentioned below. All correlations were performed using Spearman's Rank. This measure was used rather than Pearson's Product Moment correlation because the data are nonparametric i.e. it is the relative positions of the data points that matter not their absolute values.

6.2.4.1 Content Validity

An interitem correlation was performed in order to test if each item within each construct of the three constructs of perceived ease of use (PEOU), perceived usefulness (PU) and perceived involvement (PI) was essentially measuring the same thing. As the subjective norm (SN) construct is a unified item, this measure is not possible for that construct. The results are shown in tables 6.2 to 6.4.

As may be seen the lowest correlation in the PEOU construct is that between the second item in the construct and the third (0.34 rounded, significant at $p<0.05$). The highest correlation is between the first and third items (0.79 rounded, significant at $p<0.001$). This perhaps indicates a slightly different understanding between ease of use of the software as software and ease of use when trying to perform a specific task.

Perceived Ease of Use

	EOU1			EOU2		
	Corr	No	Prob	Corr	No	Prob
EOU1	*	*	*	0.4806	26	0.0065
EOU2	0.4806	26	0.0065	*	*	*
EOU3	0.7894	26	0.0000	0.3374	26	0.0459
EOU4	0.5717	26	0.0011	0.6372	26	0.0002

	EOU3			EOU4		
	Corr	No	Prob	Corr	No	Prob
EOU1	0.7894	26	0.0000	0.5717	26	0.0011
EOU2	0.3374	26	0.0459	0.6372	26	0.0002
EOU3	*	*	*	0.6241	26	0.0003
EOU4	0.6241	26	0.0003	*	*	*

Table 6.2

The correlations between the items in the perceived usefulness construct are all above 0.70, significant at $p<0.001$. This suggests that all the items within the construct are indeed measurimg the same thing.

Perceived Usefulness

	PU1			PU2		
	Corr	No	Prob	Corr	No	Prob
PU1	*	*	*	0.8199	26	0.0000
PU2	0.8199	26	0.0000	*	*	*
PU3	0.8418	26	0.0000	0.9487	26	0.0000
PU4	0.8163	26	0.0000	0.7383	26	0.0000

	PU3			PU4		
	Corr	No	Prob	Corr	No	Prob
PU1	0.8418	26	0.0000	0.8163	26	0.0000
PU2	0.9487	26	0.0000	0.7383	26	0.0000
PU3	*	*	*	0.7411	26	0.0000
PU4	0.7411	26	0.0000	*	*	*

Table 6.3

As may be seen, the two items within the perceived involvement construct have a low correlation, significant at $p<0.05$.

Perceived Involvement

	PI1			PI2		
	Corr	No	Prob	Corr	No	Prob
PI2	0.3869	26	0.0254	*	*	*

Table 6.4

6.2.4.2 Construct Validity

In order to test the construct validity of the factors within the questionnaire, Principal Component analysis with varimax rotation was used to confirm the number of factors being measured by the questionnaire. The results of the statements relating to co-workers and line managers were not included in this latter because of the way in which they are calculated to obtain a single result for SN. The results are shown in table 6.5 below.

** Rotated Factor Matrix **

	Factor 1	Factor 2	Factor 3
EOU1	-0.1870	0.7285	-0.4917
EOU2	0.1196	0.8878	0.1201
EOU3	-0.2142	0.6733	-0.5949
EOU4	0.0109	0.8526	-0.2415
PU1	0.9488	0.0690	0.0337
PU2	0.9362	-0.0488	0.1532
PU3	0.9476	0.0071	0.1175
PU4	0.8445	-0.0942	0.0312
PI1	0.0985	-0.0569	0.8399
PI2	0.0364	-0.1752	0.7520

Table 6.5

Three factors were found each of which corresponded to the items within the three constructs. The third item in the PEOU construct was the weakest, loading at 0.67. All of the items in the PU construct loaded above 0.80. The two PI items both load above 0.70.

6.2.4.3 Construct Reliability Cronbach's alpha was used to measure reliability. Only the PEOU and PU constructs could be tested in this way as three items are required to perform the calculation. The PU construct showed a high reliability at 0.94. The PEOU was also high at 0.87.

Neither of these are as high as those reported by Davis et al.(1989) who report 0.97 and 0.91 respectively. However, the number of questionnaires analysed is only twenty-six compared to over a hundred by Davis et al. (1989).

6.2.4.4 Validation of the User Engagement construct

According to the proposed model it is the combination of PEOU, PU, PI and SN that leads to user engagement with software. User Engagement is defined as akin to, although not exactly the same as, satisfaction. Satisfaction occurs following experience with and knowledge of the software. At the time of the ITU people do not have sufficient experience of the software to be able to express satisfaction with it. However, they will have formed an opinion of it which may dispose them favourably or otherwise towards the software. It is this disposition that the ITU should measure. A correlation between the SUMI and the ITU results would indicate that the ITU is indeed measuring a similar construct to the SUMI (Anastasi, 1976).

Although ten SUMI questionnaires were returned from organisation C, three of these were from voluntary users of the software. The instructions for use of the SUMI questionnaire state that the minimum number of questionnaires that should be analysed together is ten if accurate results are to be obtained, although a lower figure has sometimes been used. In this instance, it is the individual global SUMI scores which were to be correlated with the individual ITU scores. An accurate overall score for each of the scales was not required. However, the sample was very small on which to perform a correlation which may help to explain the results shown in table 6.6.

User Engagement

	ITUTOT			SUMI- TOT		
	Corr	No	Prob	Corr	No	Prob
ITUTOT	*	*	*	0.3063	7	0.2520

Table 6.6

As may be seen, the correlation between the ITU and the SUMI results is not significant. The correlation itself is positive.

6.2.5 Discussion of validation of results

• **Content validity** The low correlation between the second and third items of the Perceived Ease of use construct (0.34, significant at $p<0.05$) is interesting as it is not reported elsewhere when these items have been used with TAM. However, TAM has been tested with people who are using the software voluntarily. Perhaps if this is not the case, people are less certain about whether they will find it easy to get the software to do what they want it to do (item 2 of the PEOU construct) because that includes working in a particular way, whereas they find it easier to judge whether they will readily be able to become skillful at using the software (item 3 of the PEOU construct). This suggests a distinction between use of the software and use of the software to perform a particular task.

The Perceived Usefulness construct would appear to be the strongest of the the three constructs measured using inter-item correlation. The lowest correlation is .74 and all are significant at $p<0.001$.

The Perceived Involvement construct is weak with a low correlation of 0.39 significant at $p<0.05$.

Overall, however, the results did not indicate that any changes should be made to any of the constructs at this stage.

•Construct validity within the questionnaire

The factor analysis, used to test for the number of distinct constructs, showed that each of the items within a construct loaded on that construct. The PEOU items were slightly lower than those reported by Davis et al. (1989) but the PU items were slightly higher. The three constructs were shown to be distinct from each other. It is not possible to test the SN construct in this way. However, the results were encouraging.

• Construct Reliability

It is generally accepted that an alpha above 0.7 is acceptable on this test (Davis et al., 1989). Both PEOU and PU were above 0.8. As at least three items are required to perform this test neither PI nor SN qualified.

• Construct validity of the questionnaire

It is important not only to demonstrate convergent and discriminant validity of the items within the questionnaire but also that they are measuring what was intended. In this instance it is the total of all four constructs that gives the User Engagement measurement. Hence the need for the correlation between the total of the SUMI and the ITU questionnaires. However, the correlation shown was positive although not significant. The scoring of the ITU means that the lower the score, the more favourable the response, whereas the scoring of the SUMI means that the higher the score, the more favourable response so a negative correlation should be expected. This would appear to suggest that the ITU was not measuring a similar construct to the SUMI.

The SUMI results themselves are very wide spread, the lowest being 35 and the highest 66, thus ranging from below what would normally be expected to an above average score. This is a very wide span for just seven people. The median of the scores was 46 and the mean was 50, indicating that the response tended to be positive rather than negative.

When the individual scores on the ITUs were checked, it was found that they ranged from 40 to 58, This means that they were all within the grey area described in section 6.1.1, which lies between 40 and 60 and which could indicate a positive or negative response. This large span of uncertainty exists due to the method of measuring subjective norm. The median of the scores was 44 and the mean 47, which indicated that the responses tended towards the positive end of the scale. It was also found that one of the respondents had said that there were no co-users of the software. When his responses were removed, the correlation became negative although still not significant.

Given the small number of respondents, the indications were that the ITU was measuring a similar construct to the SUMI.

These results demonstrate one of the problems of working with small samples and performing statistical analyses. However, with the administrative size of businesses being constantly reduced, partly as a result of new technology, even large companies will often have only a small number of people operating multi-user software. Issuing the questionnaires without a good knowledge of the software being used meant that it was not possible to say whether the results of the SUMI were due largely to the characteristics of the software or to the context in which it was being used. However, the wide variation of scores from just seven people suggests that the software was not being judged alone.

All of the people who responded with the SUMI were from organisations in the Private Sector. All but one had had prior experience of similar software. The one who had no prior experience returned one of the higher scores on the SUMI but not the highest. The number of co-users varied from zero to fifteen with one unknown. All but one felt there was management support for use of the software and all said that they would recommend the software to others. This latter would suggest a positive reaction to the software but it is also possible that the fact that the ITU and additional questions were completed whilst on the trainer's premises meant that people felt they should at least appear to be positive. The response to the question about management support could also be a reflection of the fact that the organisation was paying for someone to attend the training session.

The lack of real knowledge about the reasons for the responses that people give is a weakness of the questionnaire method. Whilst it provides figures which may be compared and speculated upon, it is unable to provide any depth. Even the questions which were intended to help place the respondents into a context only generated further questions. The strength of the case study is in helping to shed light upon what lies behind the responses that people give.

As the results of the questionnaire validation were generally encouraging, it was decided that there was not sufficient basis on which to make a change to the construct items themselves at this stage.

The next chapter describes the replacement process for both the MM and PM packages. The use of the ITU, interview and discussion forms a supporting triangle of data with which both to test the model itself and the measurement instrument which is the ITU.

Chapter 7

Validating the User Engagement Model

The principle aim of this research is to determine whether it is possible to assess the likelihood of satisfaction with a package before committing to purchase. In other words, does the presence or absence of certain factors strengthen or weaken the likelihood of user satisfaction in the future. It is argued that how people perceive software will influence their feelings towards it. The model developed shows the factors which will affect an individual's disposition towards a piece of software, either favourable or unfavourable (i.e. their engagement with it). It is suggested that if people have a negative disposition towards the software from the start, they are likely to be dissatisfied later. If they have a positive disposition, they are likely to be satisfied. The level of engagement is formed as a result of perceptions regarding ease of use, usefulness, involvement and perceptions about the beliefs of referent others.

As described in section 4.6, one of the problems of working in a real life environment is that it is not possible to test hypotheses scientifically. In their investigation into the link between user characteristics and DSS effectiveness, Ramamurthy et al. (1992) were able to remove organisational influences by using university students as their subjects. In the research being described, the organisational influence of superiors and co-workers is one of the factors hypothesised to be important. It is not possible to dictate how people should behave in order to assess the effect of that behaviour on others. It is only possible to observe and explain. Neither are perceptions of ease of use and usefulness factors which may be manipulated. However, the position of the author in the organisation meant that it was possible to change the amount of involvement people had in the selection of software. This change could only be in comparison with the previous method of software selection and only in the amount of involvement in the sense of physical presence not in the degree of ownership felt by the users.

Partly as a result of the low scores achieved on the SUMI questionnaires, it was agreed that both the MM and PM software should be reviewed and, possibly, changed. This took place over a period of three years.

The key difference in the process of selection between the IM software and the other two packages was that all the people who would be using the IM software in the future were invited to the overview of the packages on the shortlist. They were then asked which one was preferred. It was decided that the process used for the IM software should be followed for the introduction of the next software package. The ITU would be issued after the software selection but before it had been installed. The SUMI would then be issued once people had had time to become familiar with its use. The Ravden and Johnson (1989) checklist would again be used to see if there was any difference between the original package and the replacement package according to the checklist.

This would allow a comparison to be made between the level of satisfaction of the original package and the replacement package, according to the SUMI, and a comparison between the usability of each according to the Ravden and Johnson (1989) checklist. The ITU results could be correlated with those of the SUMI and cross-validated by discussion with the users.

The MM software was the first package selected for review. The process is described in section 7.1 The same process was then repeated for the PM software replacement and is described in section 7.2. The results of the replacement process for each package are then compared with the results for the original package, with each other and with the model of User Engagement, in section 7.3.

7.1 Review of the MM software

Following the poor results from the SUMI questionnaire with regard to the MM software in use, the organisation decided that the software requirements should be reviewed. The results from both the pre and post questionnaire interviews had suggested that the software itself was not the main cause of dissatisfaction. This conclusion is also borne out by the results of the independent evaluation of the software. The dissatisfaction seemed to stem from a combination of people feeling that they had no say in the choice of software and a feeling that a way of working was being imposed without the reasoning behind it being either properly explained or understood. However, the central group (CM) was adamant that there must be conformity at the depots and that they were clear on what they wanted to achieve. It was, however, agreed that more effort should be made to ensure that people felt that they were involved in the choice of software and that their views would be taken into account. This would be done by discussing their requirements with them; inviting them to the overviews of the shortlisted packages; and asking them to vote for the one they preferred.

The requirements review was scheduled to take place about 18 months after the SUMI questionnaires had first been handed out.

7.1.1 Involving the users

• Assessing the requirements

It had been intended to speak to all users of the current MM system. However, it was not possible to arrange for all concerned to be available at the same time and the schedule for purchasing new software meant that it was not possible to visit the Maintenance locations (depots) on more than one occasion for the same reason. However, people were spoken to on the telephone and were asked again how they felt about the existing software and whether they felt a replacement should be sought. By this stage, a GUI (i.e. Windows) version of the software was available and all the users had heard of it. The view of all those spoken to over the telephone was that a change should be made to the newer version of the software as it would be easier to use because it was Windows based. Two expressed an interest in knowing what other software was available but the general feeling was that the new version of the current software should be considered first.

• Interviews and discussions

In total 9 people were interviewed face to face. Eight were from each of the depots and one from the CM. The ages ranged from late twenties to late fifties. Eight were male and one was female. Three were the line managers in overall charge of the work at each depot and one was from the CM group. All but one had previously completed the SUMI on the existing system. The one who had not previously completed the SUMI was a new user to the system who had the task added to his job as the CM wanted to capture more information about ongoing work at the depots, bringing in tasks not previously recorded.

The interviews were unstructured and took the form more of general discussions but with the aim of discovering what was lacking in the present system and what would be an improvement.

•Results of interviews and discussions

Two of the line managers declared that the system was there for the benefit of the people in the CM and that any decisions would no doubt be made by them. Of these, one was relatively content with the current situation. He did not use the software himself and continued to run the depot as he had in the past. The other did not use the software himself and did not feel that it was of much benefit to him in running the depot. He had occasionally had a dispute with the CM as he wished to continue to run the depot as he had in the past which meant that some staff were being given priorities which did not agree with what the CM thought should be prioritised according to the information held on the system. In order to be of any help to him, this second line manager felt that any new software should be able to record information to a much more detailed level. (In fact the current software had this capability but it had been felt by the CM to be too much effort for the benefit obtained).

The third line manager felt that the existing software could be improved by better reporting facilities, a more intuitive interface (by which, it transpired, a Windows front end was meant) and more structure to the way in which the software was being put to use. This was the only line manager who actually indicated an interest in the software and felt that there was a possibility of it being used to his benefit. He was also the only line manager who occasionally had some hands-on use of the software in order to extract information. Both of the others depended on their staff to tell them what was going on rather than look-up information on the system itself.

The person from the CM felt that the software could be improved by having a Windows front-end (the reasoning being that people were dissatisfied with the existing software because they found it difficult to use) and improved reporting facilities. Several facilities were mentioned which, in fact, were already present in the existing software. When this was pointed out, the comment was made that they were not easy to use. However, it was not possible to determine what could be done to make the facility easier to use. There seemed to be some confusion over how to use the function to achieve the desired result, which suggested a difficulty with the application of the software rather than with the software itself.

The four remaining people who had been using the existing software at the time of the SUMI questionnaire all either said that they wanted facilities already present in the software but not used or that they wanted some of the functions generally available with software conforming to the MS Windows standard (e.g. cut and paste). None could point to any specific areas of difficulty with the existing software but claimed that it was generally awkward to use and "out-of-date". With regard to the functions which already existed, they were aware of them but had been instructed by the CM that they were not to be used at present. Some of the functions were those previously mentioned by the person from the CM. This suggested that there was, in fact, agreement over what was needed in some areas, but a lack of joint understanding over how to achieve it. What was being requested, therefore, was not an enhancement to the existing software but a different way of working.

The fifth person, who was a more recent user of the software, was discontented with it. Again, the discontent arose because the user could not see how it helped in the job being performed and merely added to the workload for the benefit of others. Neither was it felt that there was really much benefit to those others. The feeling was that the system was being imposed for the sake of conformity but without any other useful purpose. The only improvement suggested was a change to working practices which did not include the software, not to the software itself. The line manager for this person was the one who had been in dispute with the CM. Interestingly, it was this manager who decided that the extra information required by the CM should be entered by this person, who did not have prior PC experience, rather than by an existing user of the software.

In brief, the result of both the telephone calls and the face to face discussions was that there was no requirement for different functionality. It was felt that the reporting module of the system was lacking and not flexible enough. However, there was a feeling that use of the software could be improved by the introduction of features generally found in Windows based software.

7.1.2 Reviewing alternative software

Another, competing, package which had been written for the Windows operating system was reviewed by IT and the person from the CM. This was to assess whether the fact that it had been written for Windows rather than converted to it, gave it a more intuitive interface and whether the reporting module gave greater flexibility in producing reports. It was not possible to involve the potential users at this time as the organisation considered it an unnecessary expense.

In the event, it was found that the package was unable to provide all the functions required and already available in the current software. The mode of operation was very similar to the current software, apart from the Windows features. This being the case, it was agreed between the IT department and the CM that the first option should be to assess the benefits of moving to the Windows based version of the present software rather than seek to introduce a completely new package.

7.1.3 Matching the software against the requirements

Checking the requirements list and information re the transfer of existing data was undertaken by the IT department. As no new functionality had been mentioned as being required during the discussions with people, the original requirements list was used to assess whether the new version of the software met the functional requirements. It was established that it did. It was also established that all existing data could be transferred to the new version without the need for re-keying. The reporting module had been completely changed and provided much greater flexibility of reporting. Existing reports would have to be re-written but this would be undertaken by the software supplier.

7.1.4 Assessing the software

It was agreed with the CM that as many of the users of the current software as possible should be invited to demonstrations of the proposed system. The aim would be to show:

- how it provided the functions they currently used;
- how the Windows features fitted with that;
- what additional functionality was available;
- an overview of the reporting module.

People would be able to pose questions and ask to be shown any part of the software they wished. It would then be for them to say whether they felt that they wanted to see what was available from other software providers or whether they were happy to move to the Windows version of the existing software.

Three sessions were arranged over three consecutive weeks. These were away from the workplace at a location convenient for access from all three depots. Overnight accommodation was provided plus evening meal and breakfast. There was a general invitation to everyone who was currently using the software to attend. Only the managers in charge at the locations were specifically asked to attend. They were also requested to allow the time for any of their staff who wished to. There was no opposition to this.

In total twenty-four people attended. This included the three people from the CM and each of the managers in overall control at the three depots. Some of the attendees were new users to the existing software and had not completed the original SUMI questionnaires. It is worth noting that the person in charge of the whole maintenance operation saw nothing to be gained by being present at any of the sessions. The line manager in charge of the CM was present at one of the sessions only. It was felt that a presence at all the sessions might be interpreted as an attempt to influence people.

The morning was given to the software providers who were reasonably familiar with how the software was used in the company. (The format of the sessions had been determined by discussion with the IT department as organisers of the sessions). They first gave a general demonstration of all the software features and then made specific comparisons about how tasks were currently carried out and how they would be performed under the new software. They also indicated how potential improvements might be made to existing tasks by use of the Windows facilities. After a buffet lunch, during which the author spoke to individuals on their reactions so far, the software provider gave an overview of the reporting module.

The session was then opened for general questions and discussion. Most of the questions centred on how specific tasks could be performed. Ideas were proposed for where the cut and paste functions might be used and how existing tasks might be better performed under the Windows version of the software. There was some concern over the reporting module which was seen to be more complex than the one currently used. However, it was also agreed that it was capable of providing much more information and that it would be worth spending the time to learn it. There was also some discussion on what other software was available. Accommodation for a second night had been booked so that there was no time limit on discussion.

7.1.5 Choosing the software

At the conclusion of each session, when there were no further questions or points to be raised, the author requested that everyone consider what they had seen and what had been discussed and to inform anyone else who had been unable to attend of what had happened. At the end of the final session one person from each depot was asked to carry out a poll of who was in favour of moving to the new version of the software and who wanted to explore other packages. They were then to inform a contact in the CM of the results. No names were to be mentioned, just the number in favour of moving to the Windows version of the present software and the number wishing to see alternative offerings. It was assumed that there was no wish to keep the present software as it was. A close date of one week on from the last session was given. If there were further questions in the interim these should be addressed to either the CM or the IT department and the question and answer would be generally distributed.

In the event there were no questions raised either of IT or of the CM and there was a unanimous decision to move to the Windows version of the current software. This concurred with the impression gained by the author during general discussion with people at each of the sessions. During those discussions several people had expressed reservations about improvements being possible without a change to working practices. However, such reservations could not be shown through a yes/no vote on changing the software.

7.1.6 Issuing the ITU

Following the vote in favour of moving to the later version of the software, the ITU questionnaire was distributed to all who had attended the discussion sessions. This was four weeks after the people who attended the first session had seen the software. They did not have the software available at that time but did have the presentation notes and screen shots from the supplier.

A note was sent to each person individually asking them to complete and return the questionnaire. It was explained that the purpose was to ascertain people's reactions to the software on the basis of the demonstration and stressed that the responses would be kept confidential. Twenty-four questionnaires were distributed and twenty-three were completed and returned. All had identified themselves by writing their name on the ITU.

Although the length of time between the first overview and the issuing of the ITU was rather long from the point of view of the people on that first session, it was not felt that the questionnaire should have been issued earlier. To have done so before everyone had been given the chance to vote, could have been interpreted as an attempt to influence people. In responding to the ITU people were indicating their general feeling towards the software not giving judgement on specific aspects of it. Therefore, whilst it would be preferable that people complete the questionnaire whilst the details of the software are still fresh in their memory, it was not felt that their overall feelings would be too much altered in the intervening time.

7.1.7 Results of the ITU

The Unistat package was used for all the analyses mentioned below and a Spearman's Rank correlation was used in all the correlations mentioned. As the use of the questionnaire with this package was also an exercise to validate the questionnaire, the content and construct validity and the reliability of it were checked first.

• **Content Validity** Inter-item correlation was performed on each of the three constructs of PEOU, PU and PI. The results are shown in tables 7.1 to 7.3 below.

Perceived Ease of use items

	PEOU1			PEOU2		
	Corr	No	Prob	Corr	No	Prob
PEOU1	*	*	*	0.6987	23	0.0001
PEOU2	0.6987	23	0.0001	*	*	*
PEOU3	0.8171	23	0.0000	0.7741	23	0.0000
PEOU4	0.6842	23	0.0002	0.8086	23	0.0000

	PEOU3			PEOU4		
	Corr	No	Prob	Corr	No	Prob
PEOU1	0.8171	23	0.0000	0.6842	23	0.0002
PEOU2	0.7741	23	0.0000	0.8086	23	0.0000
PEOU3	*	*	*	0.7552	23	0.0000

Table 7.1

As may be seen, each of the perceived ease of use items show a correlation above 0.6, significant at $p < .001$.

Perceived Usefulness items

	PU1			PU2		
	Corr	No	Prob	Corr	No	Prob
PU1	*	*	*	0.7565	23	0.0000
PU2	0.7565	23	0.0000	*	*	*
PU3	0.7133	23	0.0001	0.7767	23	0.0000
PU4	0.7515	23	0.0000	0.6643	23	0.0003

	PU3			PU4		
	Corr	No	Prob	Corr	No	Prob
PU1	0.7133	23	0.0001	0.7515	23	0.0000
PU2	0.7767	23	0.0000	0.6643	23	0.0003
PU3	*	*	*	**0.6657**	**23**	**0.0003**

Table 7.2

All the items in the perceived usefulness construct have a correlation above 0.6, significant at p<.001.

Perceived Involvement items

	PI1			PI2		
	Corr	No	Prob	Corr	No	Prob
PI1	*	*	*	**0.3714**	**23**	**0.0405**

Table 7.3

The correlation between the two items in the PI construct show a low correlation (0.37) but this is still significant at p <.05.

• **Construct Validity** Principal Component analysis with varimax rotation was used to confirm the number of factors being measured by the questionnaire. As previously explained the calculation of SN meant that it was not a suitable candidate for this procedure. The result produced three factors with each of the items loading on its expected group. The lowest result, 0.71, was produced by the fourth statement of the PU construct ("I would find xx useful in my job", where xx is the name of the software). All others were above 0.82. These results are shown in table 7.4.

• **Construct Reliability** Cronbach's alpha was used to measure internal consistency (reliability). As three items are required in order to achieve a result, it was only possible to perform the analysis for PEOU and PU constructs. Both achieved a score above 0.9.

** Rotated Factor Matrix **

	Factor 1	Factor 2	Factor 3
PEOU1	**0.8403**	0.2669	-0.0367
PEOU2	**0.8507**	0.2397	0.0117
PEOU3	**0.9430**	0.1142	0.0291
PEOU4	**0.8223**	0.2727	-0.0298
PU1	0.1054	**0.9444**	0.1713
PU2	0.2532	**0.9053**	0.0430
PU3	0.2877	**0.8658**	0.2094
PU4	0.4727	**0.7098**	0.0550
PI1	0.0702	0.1608	**0.8270**
PI2	-0.0912	0.0887	**0.8264**

Table 7.4

7.1.8 Discussion of ITU

• Validity of questionnaire

The PEOU, PU and SN constructs all fared well on the Content and Construct validation and Reliability. The apparent problem in interpretation of the second and third items of the PEOU construct, suggested by the results from the ITUs from external organisations, does not manifest itself in these results.

The PI construct in the questionnaire was evidently somewhat weak in relation to the other constructs. However, the two items to measure PI gave a low but significant correlation with each other and the factor analysis loaded the two PI items as measuring the same construct. Given the still comparatively low user sample, the results overall gave some encouragement that the questionnaire was on the right lines and that construct validation of the whole questionnaire should be carried out.

• Validity of User Engagement construct

The purpose of the ITU is as a measurement tool to give an indication of what the general feeling towards the software is i.e. the level of user engagement. The hypotheses have been proposed that if software is perceived to be easy to use and useful and if people have been involved in its selection, they are likely to have a positive disposition towards the software, provided that they perceive those around them to also have that disposition. The hypotheses are derived from the literature on the factors affecting satisfaction with software and technology in general and support was found for them from interviews with users and the results of the SUMI questionnaire from three existing packages. There needs to be some means of measuring User Engagement prior to the introduction of packaged software, in addition to the evidence offered by interviews and discussions, in order to test the hypotheses. In order to do this, the ITU must be shown to indeed be measuring a positive or negative disposition towards the software. The means of providing this validation was to correlate the results of the ITU against the results of the SUMI after the package was in use.

The scales on the ITU questionnaire are such that a low overall score (i.e. below 40) indicates a positive feeling towards using the software, whilst a high overall score (i.e. above 60) indicates a negative overall feeling. For SUMI, on the other hand, the higher the score, the more positive the feeling about the software. This means that, if there is a relationship between the outcome of the ITU and the outcome of SUMI, it should show as a *negative* correlation. The median and the mean of the user engagement scores for the 23 users who completed the ITU are shown in table 7.5. The upper and lower confidence scores are also shown. The results suggest that the general disposition towards the software is towards the negative rather than the positive end of the scale.

User Engagement scores. 95% Confidence Interval

	Value	Lower	Upper
Mean	63	54	71
Median	61	50	77

Table 7.5

The implementation of the replacement package and the distribution and results of the SUMI questionnaire are described next.

7.1.9 Implementation of replacement software

Following the decision on the purchase of the newer version of the same software package, there was a delay of about two months during which negotiations with the software supplier on price and the arranging of installation dates took place. People were kept informed of what was happening during this period.

During some of the discussions regarding the existing software, it had been suggested that hands-on use of the software prior to its live use would have been useful. Whilst the software purchase was being negotiated people were asked whether it would be useful to have the Windows version of the software available prior to the switch over. However, the response from each of the locations was that this could be confusing and that it would be preferable to go on a training course and come back to the new software already in place.

Separate training sessions were given for the transfer to the Windows version; for the use of the more advanced features; and for the reporting module, which itself was divided into two parts. The cost of training was high. In order to keep costs to the minimum, only the core users attended the sessions provided by the supplier and the remainder were trained in-house by one of the CM team. Eighteen people were due to receive formal training. Only seventeen actually received training due to sickness. Training in advanced features of the report writer was provided for six people after a period of three months.

It was arranged that the new version should be installed at each site whilst the core users were on the training course. This took place during three consecutive weeks. During the transfer period, the data from the depots was not available to the CM who themselves were transferred to the new version of the software in the first week at their request. This was so that they could have some experience of the software prior to it being introduced elsewhere.

7.1.10 Distribution of the SUMI

It had been intended to distribute the SUMI about three months after the software was in use. This would ensure that people had experience of the software but that there had been minimum staff movement since the issue of the ITU. In the event, unforseen circumstances meant that it was just over a year after the new version of the software had been installed that copies of the SUMI questionnaire were distributed.

As it was not possible to visit the different locations the SUMIs were sent to the manager in charge at each site with a request to distribute it to people using the software on a regular basis. As well as the questionnaire itself, there was a note to each person again explaining the background to the SUMI and how it should be completed (as per the instructions in the SUMI manual) and thanking them for their time. No interviews were carried out prior to distribution of the SUMI.

Twenty-four SUMI questionnaires were sent out and twenty were completed and returned for the new version of the software. Of the people who completed them, sixteen had previously completed the ITU. Six people who had previously completed the ITU had either left or changed jobs within the organisation. One person was still in the same post but did not complete the SUMI. Three of the people who completed the SUMI but had not completed the ITU were people who were not using the software when it was first introduced. One person who completed the SUMI but had not completed the ITU had been invited to attend the overview sessions but had not been able to.

Of the fifteen people who had completed the SUMI questionnaire for the original version of the MM software, eleven completed the ITU and nine completed both the ITU and the SUMI for the new version of the software. One person who completed the original SUMI had left before the new version was considered. Two people who had completed the ITU and the SUMI had left or changed jobs before the distribution of the second SUMI. Two people who were still in the same post had been unable to attend the overview sessions so had not completed an ITU. It is not known why, of these two, one did not complete the SUMI but it was later discovered that the second had been away when the SUMI questionnaires were distributed and had not been aware of it.

In summary, of the 20 people who returned the SUMI, 16 had previously completed the ITU and of them 9 had also completed the original SUMI.

7.1.11 Results of the SUMI

The SUMI program was used to produce the following tables:

a) Profile Analysis

b) User scales

c) Item Consensual Analysis

When the analysis was run, two users were flagged as Outliers. When they were removed and the analysis re-run the Global score was reduced from 31 to 29. As the global score was well below 50 even with the Outliers, the Profile Analysis presented in table 7.6 includes their scores and the discussion which follows also refers to the analysis for all twenty users.

a) Profile Analysis

The results of this analysis are shown in table 7.6. The Global score of 31 is well below the minimum 40 expected for state of the art software and is two points lower than the run for the previous version of the software with fifteen users. The Upper and Lower Confidence limits (Ucl, Lcl) span 5 either side of the median for this scale. This is a smaller span than was achieved by the SUMI results for the previous version of the software but the sample size is also larger, which suggests that there is still a range of views regarding the software.

Profile Analysis

Scale	UF	Ucl	Medn	Lcl	LF
Global	50	36	**31**	26	10
Efficiency	56	38	**32**	26	4
Affect	61	41	**36**	30	10
Helpfulness	62	41	**37**	33	12
Control	50	33	**28**	23	10
Learnability	58	33	**27**	20	2

Table 7.6

The Affect and Helpfulness subscales achieve the highest scores (36 and 37 respectively) as they did on the original SUMI but the scores are lower than was achieved for the same scales with the previous version of the software where they both scored 41. According to the scores on the Upper and Lower Confidence Limits for the Affect scale, the true score lies between a low of 30 and a high of 41. The gap narrows for the Helpfulness scale where the Lcl is 33 and Ucl is 41.

The Control scale achieves a very low 28, which is one point lower than that achieved for Control on the original version of the software. There is a span of 5 either side of the median which shows a greater difference of opinion than with the previous version of the software which spanned 4 either side of the median.

The Efficiency scale fares slightly better at 32 (28 for the original version) although the spread between the Ucl and Lcl is 6 either side of the median which is the same as for the previous version.

Overall this suggests that the new software is thought to be slightly more useful than the previous version but that it is more difficult to manage. This is confirmed by the Learnability scale which achieves the lowest score at 27. There is a wide spread of thirteen between the Ucl and Lcl on this scale though, which suggests a wide divergence of opinion about how easy it is to learn how to use the software. The software has the MSWindows look and feel but certain steps have to be followed in order to complete a task.

b) **User Scales**

As may be seen from table 7.7, the scales for which two people were flagged as Outliers were Global, Control, Efficiency and Learning. User 6, who is one of the Outliers, had also given a comparatively high score on the Control scale for the previous version of the software although not then flagged as an Outlier (this user was previously number 13). The scores for User 13 are high for all the scales. This user had recently been given more responsibility and autonomy by

Individual User Scales

User	Global	Efficiency	Affect	Helpfulness	Control	Learnability
1	37	34	26	35	27	57
2	23	20	20	28	37	36
3	31	34	35	37	24	25
4	22	14	15	39	29	20
5	25	19	46	26	22	24
6	51	55	54	48	51	47 (GC)
7	23	29	27	35	23	15
8	42	41	41	41	41	41
9	26	25	38	32	37	21
10	20	22	23	25	23	26
11	17	16	27	14	23	17
12	35	31	36	43	34	26
13	61	60	60	50	61	66 (GECL)
14	25	26	29	29	28	27
15	36	36	40	53	25	31
16	31	33	40	37	36	37
17	41	49	52	51	28	35
18	36	46	48	46	29	49
19	32	33	30	44	26	20
20	17	19	17	24	17	14

Table 7.7

his line manager with regard to the software. However, the scores did not reflect some of the comments he had made in the past about the software. As may be seen, the scores from most people are quite varied but generally low.

c) **Item Consensual Analysis**

Forty four items were flagged as likely to be different from the Item Consensual Analysis (which compares the expected responses according to the standardised database to the actual responses). Thirty-three items had been flagged for the previous version of the software. In fact twelve items were flagged which had not been flagged for the previous version of the software. One item (29) "The speed of this software is not fast enough" was flagged for the previous version of the software but not for this version.

Such a poor result was disappointing especially as there had been considerable effort to try to involve people in the demonstrations and in talking to them both formally and informally. A copy of the overall scores with an explanation of the meaning was sent to everyone who had completed the SUMI. People involved in its use were also invited to comment on the reason for the results. Only one person responded to this invitation. This was one of the line managers. He said that he was surprised that no improvement was shown between the new and old version of the software. In his opinion, the newer version was much easier to use and presented a big improvement over what had been in place before. This line manager was the one who had made some use of the previous version of the software and was line manager to user 13.

7.1.12 Measuring the ITU against the SUMI

Before investigating the reasons for the poor results any further, the ITU and SUMI results were correlated using Spearman's Rank. As only sixteen of the people who had completed the ITU also completed the SUMI, the correlation was between sixteen pairs.The Global scale of the SUMI was the one chosen for the correlation as this is the overall result.The result is shown in table 7.8.

	ITU-TOT			SUMI-TOT		
	Corr	No	Prob	Corr	No	Prob
ITUTOT	*	*	*	-0.5432	16	0.0148

Table 7.8

As may be seen, a negative correlation of .54 was returned with a .01 probability of error. This shows that as the SUMI score increases, so the ITU score decreases. The scoring on the ITU is such that a low score represents a positive disposition and a high score represents a negative disposition towards the software. This is the opposite of the SUMI for which the higher the score, the greater the satisfaction. The correlation of .54 significant at the .01 level is lower than the .55 correlation significant at the .001 level, achieved by Ives et al.(1983) when checking the UIS measure under development with a 4-item UIS measure. However, they had the benefit of a sample of over 200 people and a measure of satisfaction with information services generally, compared with this small sample of 16 people and satisfaction with specific software. It was felt that the correlation was sufficiently strong to warrant more detailed consideration of the results of the ITU and to interview people on that basis.

One further check was needed first. As it would be possible for the two scores to correlate but for the scores from the ITU to be suggesting a mainly positive disposition (i.e. scores of 40 or lower), whilst the SUMI score indicates a negative reaction (i.e. scores up to 40), the mean and median of the ITU for the sixteen people who also completed the SUMI needed to be calculated in order to see whether the scores were generally high or low. The scale scores reported by the SUMI are the median in order to reduce the confounding effect caused by Outliers.

The SUMI analysis was run with just the sixteen users who had also completed the ITU. The mean and median of the ITU for the same 16 users was then calculated. The results are shown in table 7.9.

	SUMI score	User Engagement
	Value	Value
Mean	32	63
Median	31	62

Table 7.9

A Satisfactory score on the SUMI scale is placed at a mid point of 50 with anything below requiring some remedial action and anything below 40 being in need of immediate attention. The mean of the SUMI scores was 32 and the median 31, this means that the users are reporting the software to be unsatisfactory on the SUMI scale. As may be seen, the median of the sixteen scores on the ITU was 62 and the mean 63. According to the scales discussed earlier for the ITU, a score below 40 on the ITU would indicate a positive feeling towards using the software and a score above 60 would indicate a negative feeling. Therefore, not only is the rank of the scoring correlated between the ITU and the SUMI, but the median of the scores on the ITU falls into the area which indicates a negative feeling towards the software, just as the median on the SUMI does. However, the ITU could be said to be slightly negative, whereas the SUMI is very negative.

7.1.13 Consideration of the influence of different factors

The ITU result suggests a negative disposition or lack of engagement with the software on the part of the users. Yet the vote in favour of using it had been unanimous. The only way to gain some insight into why there should be this apparent contradiction was to talk to people. In order to focus the area of discussion some further analysis of the results was undertaken.

• **Correlation of individual factors and SUMI**

The ITU is composed of four factors - PEOU, PU, PI and SN. The combination of the results gave the score which was correlated with the SUMI. In order to see whether any one factor had a greater correlation with the SUMI than the others, each was correlated separately.

There was a negative correlation between each of the factors and the SUMI scores. Those of PEOU, PU and SN are very similar, all close to .43 with a .04 probability of error. Only the correlation between the PI factor and the SUMI was not significant giving a negative result of 0.3 and an error probability of 0.12.

The fact that there is a correlation between SN and the SUMI score would seem to contradict the findings of Davis et al. (1989) and Mathieson (1991) none of whom found a significant correlation between SN and Behavioural Intention (BI) or usage. The results reported here are from one, very small sample. However, the sample is from a live work environment and from users who have no choice over whether to use the software but are reporting on satisfaction levels. The tests undertaken by both Davis et al. (1989) and by Mathieson (1991) were performed using university students and with single-user software, which the students could choose to use or not. In such circumstances, the perceived views of referent others are less likely to matter to the individual.

With the exception of PI, all the factors have a similar correlation with the SUMI. The combined score of all the factors has a higher correlation and lower probability of error than any one individually. This suggested that there was no reason to assign a greater importance to one factor than to any of the others and that the combination of factors was a better indication of people's engagement with the software than any one on its own. PI has shown itself to be consistently weak. Yet there is considerable support in the field of HCI for involving the users. This suggests that the construct in the ITU is not measuring what was intended. However, it could also be that people did feel that they were involved but that this was not sufficient to alter their overall feeling of engagement with the software. Hence the results of the PI element are out of step with the results for the other factors.

• Consideration of the ITU

The results of the ITU themselves were considered next. For this the responses from all twenty-three people were used.

Ordinary Least Squares Regression (OLS) was used to see if there was a relationship between the factors within the questionnaire. Regression is normally used to assess a predictive link between factors. Prediction does not indicate causality, it is merely saying that whenever X is observed to occur, Y is also observed to occur. This again underlines the need for multiple tests under different conditions. Regression also depends on the ability to measure the variables using universally accepted measures. Evidently in the case of the ITU, perceptions are being measured and these are individual to the person and are not universal measures. The Lickert scale means that people use the same scale to assign different weights to their responses but the strength of agreement or disagreement that they assign is based on their individual perceptions. However, regression is used by Davis (1989) to show a dependency between the different factors of perceived ease of use and perceived usefulness and usage. Usage itself is a self reported measure and, therefore, not exact. Regression analysis is also used elsewhere in the literature as a means of assessing the strength of a relationship between different factors (e.g. Davis et al., 1989; Mathieson, 1991; Taylor & Todd, 1995). It was, therefore, felt that the use of regression analysis for guidance purposes was acceptable.

Given the small number of people who completed the ITU and the fact that the measurements were not exact the results were regarded with caution. However they did help to guide the content of the discussion with people to find out why they had voted in favour of moving to the newer version of the same software rather than consider other packages and yet, as the SUMI results showed, they were not content with the result.

The results of the regression analysis are shown in table 7.10. It is interesting to note that the R^2 value is the same with either PEOU or PU as the independent variable. The coefficient is slightly higher when PEOU is the independent variable and PU the dependent. Both are significant at the .01 level. From this it may be seen that there is a link between PEOU and PU. This suggests that for some people the easier the software is perceived to be, the more useful it is perceived to be and for others, the more useful it is perceived to be the easier it is thought to be.

Dependent Variable	Independent Variable	Coefficient	R^2	Significance
PEOU	PU	0.33	.26	.01
PU	PEOU	0.75	.26	.01
PEOU	PI	0.09	.00	.80
PU	PI	0.67	.07	.21
PEOU	SN	0.26	.42	.0009
PU	SN	0.16	.48	.0002
PI	SN	0.03	.05	.29

Table 7.10

The suggestion that perception of ease of use may influence the perception of usefulness may appear to disagree with the findings of Davis (1993) who stated that however easy to use an interface was perceived to be, it wouldn't be used if it wasn't felt to be useful. However, Hubona and Blanton (1996) found that perceived ease of use did have an influence on perceived usefulness. The comparison between TAM and TPB (Davis et al., 1989), also found that PEOU was an important factor after the first exposure to the software but reduced in importance once the software was in use. They suggested that if software appears to be easy to use, people may think that they will be able to use it for certain tasks and, therefore, perceive it to be useful. Thus the perception of ease of use may influence the perception of usefulness. If it is later found that it is not possible to use the software in the desired way then ease of use alone will not ensure usage.

The link between SN and both PEOU and PU is also significant ($p<0.001$). It would appear that the individual's own perceptions particularly with regard to perceived usefulness may be predicted by the perceptions about the views of referent others. This agrees with the research by Martinko et al. (1996) and Hirschheim and Newman (1988) who suggest a link between the perceived support of others and the satisfaction of the individual. According to Fishbein and Ajzen (1975), people may be influenced by what they perceive to be the opinions of others whose views are important to them, possibly because they are in a position of authority or power.

The regression results suggest that there is no link between the perception of involvement and perceived ease of use and usefulness. This would appear to disagree with the general consensus in the HCI literature that involvement of the users will influence their attitude towards the software (e.g. Grønbæk, Grudin, Bødker & Bannon, 1993; Ives & Olson, 1984). However, the involvement described in the literature is that of people in the design of software. The involvement in the study described was limited to that of establishing requirements and of voting for the software.

Neither does the perception of the views of others (SN) appear to have any correspondence to the perception of involvement. The involvement specified is very specific and perhaps people are not inclined to refer to others when making their response.

• **Questions arising as a result of the regression analysis**

The link between PEOU and PU raised the questions as to whether people thought the software would be easy to use when they saw the overview and whether, now they were using it, that perception was thought to be correct. Similarly with regard to perceived usefulness.

The results of the regression with SN indicate a link between people's perceptions of the software and their perceptions of those around them. However, people are unlikely to admit to this being the case especially in the work environment. Given the views expressed in the interviews before and after the original SUMI for the DOS version of the software, it was felt that an attempt should be made to establish whether it was felt that the attitude of management was more supportive and encouraging than previously, or whether it was still perceived to be either neutral or coercive and what the perceptions were of the attitudes of co-workers.

Further investigation into the Perceived Involvement factor was required. The PI statements were intended to elicit whether people felt they had a choice in the software selection. It was decided that people should be asked whether they had indeed felt there was such a choice and also what they had understood by the statements on the questionnaire.

7.1.14 Independent Evaluation

Before talking to people, another evaluation of the software itself using the Ravden and Johnson (1989) checklist was needed in order to see whether that indicated any difference in usability between the two versions. The same scoring method was used and the same person performed the evaluation. This was because he was independent of the user group but was familiar with the tasks for which the software was used.

The method of scoring meant that the lower the score the greater the usability (as measured by the checklist). The result was a score of 216. This was a lower score than the 229 achieved by the package which had been replaced indicating that there were marginal improvements. There were seven ticks under Moderately Satisfactory and two under Very Satisfactory compared with seven ticks under Moderately Satisfactory, one tick under Very Satisfactory and one under Neutral on the original package.

This showed a slight improvement in the usability aspects of the software as measured by the checklist over the previous version.

7.1.15 Follow up discussions with users

It has already been pointed out that one of the problems of having only a small sample of people is that it is more difficult to apply statistical tests with accuracy. One of the advantages, however, is that it enables a more detailed review of the responses to questionnaires. This took place over a period of four weeks after the results of the SUMI were known.

Only one person had responded to the invitation to comment on the poor results of the SUMI for the new version of the software. Everyone else had to be approached for comments. This suggested that either they were concerned about making their views known or they did not think it would make much difference if they did make them known.

Sixteen people who completed the ITU had also completed the SUMI. Of the sixteen, eleven were interviewed. The views of one of the three people who had started to use the MM software after the new version was introduced were also recorded.

In all but three cases, the interviews took the form of a discussion rather than an interview and was intended to be informal in the hope that people would feel more able to express an opinion if it was not recorded at the time. In the three cases, where notes were taken during the discussion, the people concerned were quite happy that notes should be taken although they were assured of confidentiality.

The purpose of the discussions was to establish:

a) whether people felt that the software was easy to use, at the time of the overview and now;

b) whether they felt it was useful at the time of the overview and now;
c) whether they had felt involved in the process of choosing it;
d) whether they felt that management and co-workers were supportive of its use.

a) **Ease of use**: All but two people felt that the software was easier to use than the previous version. However, only three people thought that it was actually easy to use. The general view seemed to be that it must be easier because it was Windows based but that it was more complex than the previous version and that you had to know what you wanted to do with it. This was particularly true of the report writer and few people used it for anything other than the reports needed on a regular basis to give the status of certain jobs.

It was commented that, at the overview, because the suppliers had known how to use the software, they had made it look a lot easier than it really was. Thus people at the overview thought it would be easier to use than proved to be the case with hands-on usage.

In fact the mean of the PEOU score on the ITU for the 16 users was 14 with upper and lower confidence intervals of 15 and 12 respectively. This shows that people were not committing themselves very definitely (16 being the middle/uncertain score) although they were slightly favouring the easy to use side of the scale.

Regarding the apparent discrepancy between the results on the SUMI Helpfulness and Learnability scales (37 and 27 respectively), it was felt by all but two people that the flexibility of the system made it more complex. Ease of use (Helpfulness) seemed to be equated with standard Windows facilities, whereas ease of learning also meant that it should be obvious which steps to take and the flexibility of the system meant this wasn't the case. The two people who didn't agree with this felt that steps followed logically one to another. Both of these people had been flagged as Outliers on the Learnability scale.

b) **Usefulness** No-one felt that the software was more useful than the previous version. Several commented that they had thought from the overviews that it could be useful as various functions had been demonstrated. Hence their feeling at the overview and before using the software had been that there might be some useful functions. However, there had been no direction given in how or whether to make use of those functions after the software had been installed. They did not feel that it was for them to suggest that more use should be made of the software as it was not really a tool for their benefit. One person commented that the SUMI result was not really surprising. It had been good to be supplied with new hardware and a more up-to-date system but really no-one *wanted* to use it and the SUMI results would probably be the same for any system. If the system broke down everyone would be able to carry on just the same.

The mean of the PU score on the ITU for the 16 users was 15 with the upper and lower confidence intervals at 18 and 12 respectively. This is a wider spread than on the PEOU scale and is again indeterminate.

One person particularly wanted to air his views about how much work the system caused him, taking him away from what he saw as his "real" job. This same person had put on the ITU that his line manager would want him to use the software and that he generally wished to conform to his wishes. He had also ticked that his colleagues would wish him to use the software but was non-committal about wishing to comply with their wishes. However, it was known that this same person felt so strongly about being required to use the software when he thought it was not necessary that he was taking his grievance to the union. When his ITU response was removed from the regression analysis the R^2 value of subjective norm on PU increased from 0.48 to 0.59 significant at $p<0.001$ and on PEOU from .42 to .44 significant at $p<0.001$. This change was confirmed by consideration of the full responses on his ITU, where he had indicated that he would not find the software easy to use and that it would not be useful. Such a cross check was only possible because the person had identified himself on the questionnaire and his circumstances were known. This serves to underline the danger of questionnaires being used for anything other than general guidance without good knowledge of the individuals completing them.

c) **Involvement** When asked about the level of involvement, there were just two people who felt that they had not been involved. Both had been at the overviews. One was the one above who felt that system caused him extra work. According to him no-one had taken account of his views. He had not wished to move to any version of the software but that had not been accepted as a vote. The other felt that it would have been nice to see other packages and that, whilst he claimed to be quite happy with the present software (borne out by his ITU and SUMI scores), it was useful to know what else there was. In both these cases, the respondents had understood the question on the ITU as asking whether they would like to have had more of a say (and their responses, indicating that they had not felt sufficiently involved, support this).

The others interpreted involvement as being asked their opinion and being present at the overviews but not of having a choice. Several said that they appreciated the effort to involve them and that it had been a useful exercise. At least they had been asked what they thought of the previous software and had been given the opportunity to see the new version and ask questions about it. However, the decision over which they would really like to have been given some control, was whether to use a software system at all and which work method to follow. There had been no choice in this. That being the case, the view expressed by most people was that there really had been no option but to move to the newer version of the software. They had to use something and a move to another package would have meant a lot of extra work on their part without any benefit. Moving to a newer version of the software meant that there was a change to the Windows operating system and that other packages were upgraded at the same time. At least in this way they felt some benefit had been achieved.

The mean of the PI score was 6 with the upper and lower confidence intervals at 5 and 7 respectively. As this is a two item construct the middle/undecided score is 8. These results suggest a positive view with regard to involvement. The distinction that seems to have been made by people between being involved in the exercise to choose new software and being part of the decision process which affected their work helps to explain the apparent weakness of the PI construct in the ITU. Although people were responding positively to the PI questions, the level of involvement they felt was at the low end of the scale.

The third item of the PI construct served merely as a check against what people said re their involvement in the sense of having been physically present at interviews and the overviews. With the exception of the two people mentioned above and two who had not been consulted before the overviews, no marks were in the negative section although there were some who marked the middle/undecided box.

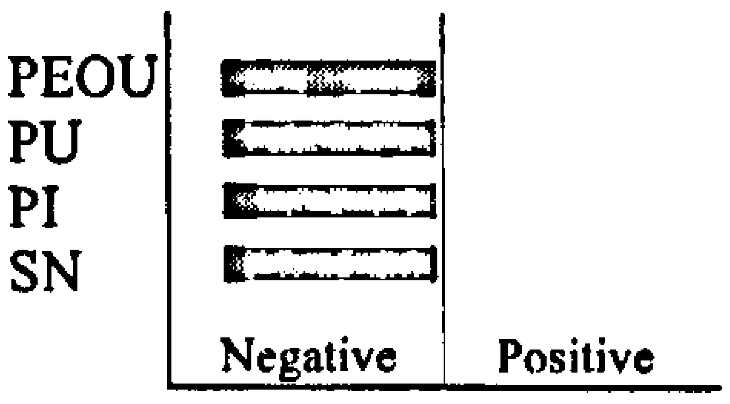
d) **Subjective Norm** All felt that those around them felt the same about the lack of usefulness of the system. They had been able to do the job before any software was introduced (this was said even by those who had not been in the organisation before the software was first used).

Comments concerning ease of use were that everyone thought it was easier to use than the previous version because it was Windows based. At the same time, however, all felt that others struggled in trying to remember the steps required for a task.

There was no perceived management support either by way of encouragement to use the software or guidance in how to best use it. This lack of support was perceived to be both at the local level and from senior management especially the senior manager in charge of the MM function.

None of the factors in the proposed User Engagement model was present in a very positive way. An indication of this had been given by the results of the ITU. The result was a low score on the SUMI which showed no improvement over that achieved by the original software.

7.1.16 Summary of MM case study



From these discussions it may be concluded that little had changed in the way people felt about how they were expected to work. The change of software had not improved the perception of its usefulness. Although no-one said that they would have preferred to keep the previous version of the software, it would seem that an underlying reason for changing it was because that provided a means of moving operating system platform and, thereby, changing other PC based packages as well. The attempt to involve them more in the process of changing the software had been appreciated. However, this form of involvement had not served to make them feel a greater need for the software. This difference helps to explain the lack of correlation between the PI construct and the other constructs in the ITU. It is the same problem expressed elsewhere in the literature between the physical presence of people at meetings and discussions and the sense that their opinions actually count. In this case, even though people were told that the decision whether to move to another version of the software was in their hands, the real choice from their point of view would have been whether to continue working in the same way as they were. This was not an option and, therefore, they did not believe that they had a choice. The same underlying feelings about the lack of usefulness of any computerised system and lack of active management support for the use of one remained. In effect, the software itself was not the source of dissatisfaction but was being used as the vehicle to voice it.

The results of the ITU indicated this lack of engagement but were only just negative. The results of the informal discussions suggest that people felt that the software demonstrators made the software appear easy to use and showed functions that could be useful but that these benefits were not apparent when they had to use the software. It is possible that when the ITU questionnaire was completed people were focussing more on the software itself. When the SUMI was completed the software was an aspect of a way of working they disliked. By this stage, the positive feeling produced as a result of the time away from the workplace to see the software had disappeared and the old frustrations had returned.

No interviews were undertaken after the ITUs were issued because these may themselves have influenced people's perceptions. It is possible to speculate that, had such interviews been undertaken and the dissatisfaction with the way of working been more strongly revealed, it might have been possible to encourage senior management to have open discussions in an effort to improve line management support and acceptance by the end-users of the system. However, to do that would require faith by senior management in the results of the ITU and the interviewers, and a desire on their part to change people's attitudes.

The selection and implementation of the replacement MM software had been used as a second iteration of the ITU validation. The ITU both serves as a formal metric by which to test the User Engagement model and as a proposed simple tool whereby organisations buying packaged software can assess User Engagement with the software. The results from the replacement MM software give support both for the model and for the ITU as a measure of User Engagement.

The next section describes the selection of a replacement for the PM software.

7.2 Review of the PM software

The replacement of the project management software was planned to happen during Year 4 of the research with the initial investigations taking place during Year 3. The results of the SUMI had been circulated in Year 2. The low satisfaction rating gave added weight to the need to seek a replacement but was not the prime reason. The company policy with regard to software replacement was that it should be considered either if the software was failing to perform its function (perhaps because the requirements had changed) or because ongoing support and maintenance was becoming costly or unavailable. In the case of the PM software, the version of the product in use had been placed on the "not supported" list in Year 3 and it was clear that a replacement should be sought if future problems with incompatibilities were to be avoided.

In the event, due to resource problems in the IT department combined with the extra workload imposed by the final preparations for the Year 2000, the plan was delayed and the introduction of the replacement software was moved to the first quarter of Year 5.

7.2.1 Involving the users

•Assessing the requirements

During the period from Year 3, resourcing in the IT department was becoming an increasing problem. It was, therefore, agreed with senior management that a full requirements analysis for software which was being replaced would not be undertaken, the focus being on the key requirements only. It was felt that, as software had already been in use, there would be a greater understanding on the part of the users as to what was feasible with software and that they would be more able to articulate their needs. At this point, there was still the option of moving to a later release of the existing software as it was not known that there would be no further releases. It was also agreed with senior management with regard to the PM software, that the people who would have hands-on use of the software would be fully consulted and involved in choosing to either move to the later release or selecting a replacement package.

• Interviews and Discussions

The first step was to speak to the person who was now responsible for managing the overall resourcing amongst the project teams (from now on referred to as the Project Planner (PP)). This person was also a project manager. The meeting took place in Year 3 and took the form of an informal discussion during which the history of the introduction of the existing software was discussed. The PP suggested that much of the feeling of dissatisfaction with the software probably arose from the system of working which it imposed rather than with the software itself. The view was also expressed that people were perhaps more accepting of its use now than they had been when the SUMI questionnaires were distributed.

Following the discussion an initial key requirements list was drawn up and suggestions of possible replacement software were noted for further investigation. Demonstration copies of the later release of the existing package and of alternative packages which the PP had expressed a wish to see and which appeared to meet most of the requirements mentioned, were obtained and a laptop was provided on which to run them so that the PP could have a trial use and the opportunity to properly compare their capabilities.

At this stage, illness prevented further headway with the project which was resumed in the following year, when the implementation had been planned to take place.

Early in the following year each person in the project teams who used the software and each project manager was interviewed in order to get his views of the key requirements of any replacement software. In some cases, only the project manager used the software. The PP was also included. By this time, it had become clear that retaining the existing software, even with a move to the latest version, was not an option as the product was no longer being developed.

The interviews were semi-structured and were carried out by two people from the IT department. In all, six people were interviewed. All were male (by this time there were no females in the project teams)..Ages ranged from late twenties to mid fifties.

People were asked to explain how work was currently performed and where the problems with the existing software were felt to be. They were also asked what their key requirements of any replacement software would be and also whether they already had a preferred software package. This latter question was asked in order to ensure that if anyone had a preference for a package, that should be considered as an option. Notes from the interviews were written up and sent to the interviewees asking them to make any corrections necessary. Finally, a combined key requirements list was drawn up and circulated for confirmation by all the participants.

• **Results of interviews and discussions**

It was clear from the requirements list that there was a need to retain all the functionality of the existing software. Everyone spoken to based their requirements on the current way of working. When asked whether they thought that the current way of working was the best method, only two people said that they themselves could not see any other way of managing their work. The others said that they had to work with the current method because that was what was needed in order to enable the PP to do his job. They did not need it for their own job. They were unwilling to suggest how else the shared resourcing could be managed.

A principal request from all those spoken to was that any new software should be intuitive to use. However, apart from being Windows based, it was unclear what this meant. There was a complaint that, as the existing software was perceived to be complex, people who could make use of it, had not done so. The result of this lack of use was that each PM team struggled in the absence of the one person who was able to use the software.

During the interviews it was apparent that there was less resentment about the way of working and more acceptance of the need for everyone to work to the same methods than there had been at the time that the SUMI questionnaires were issued.

One package was named as a favourite for the replacement by four of the six people.

7.2.2 Matching the software against the requirements

Due to other demands, the PP had not managed to properly evaluate the demonstration software which had been obtained on the basis of the initial list of requirements drawn up with him the previous year. One of these packages had been that mentioned by four of the six people interviewed.

These and several other packages were considered by the IT department in conjunction with the PP by comparing their claimed functionality against that of the final requirements list. This meant relying on the marketing information and speaking to sales people plus using comparison tables from specialist magazines. Use was also made of a comparative report purchased from a specialist firm.

Some of the packages which would fulfill all the requirements were in a high price bracket and used by international conglomerates. Of the ones which were in the medium price bracket, all claimed to perform the key functions. One of these was a package for which a demonstration copy had already been obtained. It was produced by a large, specialist supplier with an international reputation. It was agreed with the PP that, if it proved to perform as claimed, there was little need to consider similar packages. The other package selected was the one named by the people in the project teams. However, it was clear that, as this package was primarily designed for a single user managing one or many projects but not for several people managing different projects and sharing the resources for them, another piece of software would be needed in order to perform the resource management function required. Again, there was an investigation into what was available. The choice was more limited and the software chosen was selected because it had a reasonable customer base.

7.2.3 Assessing the software

Senior management had expressed a wish to extend use of the software not only within the teams but also to other areas. It was, therefore, agreed that all who were interested in the replacement software and who might use it in the future, both from the project teams and from other areas who interacted with them, would be invited to attend the demonstrations of the two possible replacement packages. It was agreed that the demonstration of the software for linking the data from the apparent favourite of these packages (i.e. the one cited by several people) would be demonstrated to just the PP as he was the person who had need of it in order to properly fulfill his function and its use was invisible to the others. The software package frequently mentioned by people in the project teams will be referred to as A and the other software as B in the following description.

The following information was circulated with an invitation to attend the demonstrations to all the project teams and to people who interacted with them:

a) The reasons for selecting the software that was chosen for the demonstrations

b) Some background to each software package

c) Background information to the companies which would be demonstrating the packages;

d) Details of the format that the demonstrations would take.

People were asked to bring the list of key requirements with them, so that they could tick them off as they were shown and make comments or add any extra items that arose. The author was present at all the demonstrations and meetings as a member of the IT department.

The demonstrations took place over a period of two days. Package B was demonstrated on the first day and package A was demonstrated on the second day. The order of demonstrations was determined by the availability of the demonstrators.

Each demonstration started with an introduction to the demonstrator and where he was from and a brief summary of the software and why it was being demonstrated. The people carrying out the demonstrations had been sent details of the way in which the teams worked and of the key requirements that it was important to show. In the event, neither followed the format requested and instead retained their standard presentation.

People had been asked to note questions for the end of the demonstration as each should have worked through the key requirements. However, as the requested format was not followed, questions were soon asked as the demonstrations progressed.

Eleven people attended each of the sessions. Of these, eight were project team managers or members . This included the PP who was also a project team manager. Of the other three, one represented senior management and two were responsible for some of the resources which were shared amongst the teams. All were male. The age range was from early thirties to mid fifties.

On the morning of the third day, the software to link the data from package A was demonstrated to the PP. He was able to pose questions as to ways of working and how the link software worked with the PM software. Details concerning support and maintenance were also covered.

On the afternoon of the third day, there was a wash-up meeting. Twelve people attended. These were the same as those who had attended the demonstrations plus another who was from senior management.

There was an open discussion regarding the software. It was not questioned whether the current way of working should be changed. At the end of the discussion, seven people had a preference for package A and five for package B. Of these, there was one senior manager who was in favour of package A and one who was in favour of package B. Although the PP had voted in favour of package A, he added that he was not sure that it was the right package. There was a general consensus that the PP should have final say as he had the greatest workload and greatest use of the software in his resource sharing role. It was agreed that the PP should talk to users and visit companies using each of the software packages and that there should then be a report back on the findings.

Reference sites were sought from the demonstrator of package B and from the providers of the link software for package A so that both package A and the link software could be considered as a single package. Three references were given for the former and one for the latter. Following telephone conversations to establish similar ways of operating, arrangements were made for the PP and members of the IT department to visit two of the companies:- One using package A and the link software and one using package B.

It was towards the end of the year before both companies had been visited. Between times the PP had generally informed people of what was happening but there had been no formal updates. By this stage, the PP had formed a definite preference for package B based on the ease with which his task could be performed. This was a change from his original view.

7.2.4 Choosing the software

A meeting with the project managers and interested parties was arranged to take place during December during which a decision on which package to purchase would be made.

As a result of the interviews following the findings with the SUMI questionnaire for the MM software, it was decided that a change in the way the software was voted for should be made. Involving people was the only variable over which the author had some control and that was limited to the final selection process. It was decided that rather than ask people to vote remotely as had been the case for the MM software, which resulted in one person claiming his opinion was not asked, they would be asked to vote whilst all together. This would be possible for the PM software as the group was very small. It would also avoid the need to go through a contact - with hindsight, asking for the votes for the MM package to be reported back to the CM group may have resulted in people being more cautious in their views. It was also decided that one of the options for the PM system that would be suggested at the meeting should be to change the working method. This latter was not raised with senior management before the meeting.

Twelve people attended the meeting. Of these seven were project managers or team members of whom six had been present at the previous demonstrations and meeting. Three were from senior management, two of whom had been present at the previous sessions. Two were in charge of the resource teams, both of whom had been present on the previous occasions. Those who had attended the previous sessions but did not attend this one were either on annual leave or sick leave.

The meeting started with a review of the key points that were required of a replacement software package. It was pointed out that there was a potential conflict in both wanting something that was very easy to use but yet had a fairly complex functionality. The previous demonstrations and the results of the wash-up meeting were recalled. A brief overview of the questions asked and responses given at the reference sites was then given. These pointed to the fact that one of the packages was a better fit to the overall requirements and those of the PP than the other but that either would do the job required.

The meeting was then opened to general discussion. Various questions regarding the suppliers, the software packages and the companies in which they were used were asked and responded to. However, no-one expressed a preference for either package.

As it was clear that package B met the current working requirements better than package A, the question was asked whether, before going any further in considering software, it should be questioned whether the current way of working was to be retained as this was obviously influencing the requirements of the software. The question was met with some surprise. The representatives of senior management said that they did not see how the working practices could sensibly be changed. However, a general discussion then took place at the conclusion of which all but two of the people present agreed that there wasn't any other way in which the fair sharing of resources could be met. The option to change working practices was, therefore, rejected.

One of the project team members asked which package the PP preferred. Once he had said which and given his reasons, another person said that the decision was made then and people seemed to generally agree. There was further discussion as to how easy it would be to change packages again in the future if one were selected now and found not to be suitable after all. Despite the fact that it was pointed out that both packages could be made to work, but one would require more effort at the linking stage and the other would require more effort at the learning stage, there were murmurs that there really wasn't a choice if they wanted to work in the same way as at present.

A vote was then called for. There were two abstentions. One abstention was on the basis that he had not been present at any of the demonstrations and did not wish to form an opinion from what had been discussed. The other was because he preferred package A but felt that package B was needed. However, he was not in favour of package B. All the others voted for package B. However, a feeling that there was not really an option was expressed by some.

The next stage was for a reseller of package B to be selected and for training and implementation to be arranged.

7.2.5 Issuing the ITU

It had been some months since people had seen the package overviews when the decision meeting took place. The process of negotiation on best overall price for the supply of the software and training took another two months. The ITU was, therefore, distributed immediately following the introductory training session.

This was a difficult time at the location as it was known that there was going to be reorganisation and potential loss of jobs. Only those people who were given training were asked to complete the ITUs. Eleven questionnaires were issued. Two of the people who completed the questionnaire were from senior management. They each wrote a note to say that they had the training because they would be using the software for enquiry and that they wished to understand the tool people would be using. They had both been at the final meeting. Two of the people were from the project teams but would be new users and had not been in either the original demonstration sessions nor in the group which had made the final choice. The remainder had all been at both the demonstrations and the final meeting. The PP was not given a questionnaire as it was felt that he had a different perspective on the process having been more closely involved in the requirements assessment and having attended the visits to other sites.

7.2.6 Results of the ITU

All the questionnaires were returned and none was returned anonymously. As previously, the Unistat package was used to perform the calculations and all correlations used Spearman's Rho. As the two senior managers who completed the ITU were voluntary users of the software, their questionnaires were removed. The statistical analyses described below was performed on the remaining nine questionnaires.

• **Content Validity** Inter-item correlation was performed on each of the three constructs of PEOU, PU and PI. The results are shown below in tables 7.11 to 7.13.

Perceived Ease of Use items

	PEOU1			PEOU2		
	Corr	No	Prob	Corr	No	Prob
PEOU1	*	*	*	0.5992	9	0.0441
PEOU2	0.5992	9	0.0441	*	*	*
PEOU3	0.5670	9	0.0557	0.3198	9	0.2007
PEOU4	0.6899	9	0.0199	0.6786	9	0.0222
	PEOU3			PEOU4		
	Corr	No	Prob	Corr	No	Prob
PEOU1	0.5670	9	0.0557	0.6899	9	0.0199
PEOU2	0.3198	9	0.2007	0.6786	9	0.0222
PEOU3	*	*	*	0.8616	9	0.0014

Table 7.11

As may be seen the highest correlation is between the third and fourth items (0.86 significant at $p<0.001$) of the PEOU construct. The correlation between the second and third items is not significant. The correlation between these two items ("I would find it easy to get the software to do what I want.." and "It would be easy for me to become skillful at using..") was also poor on the original ITU validation when they were returned from external organisations. This suggests that people may indeed be making a distinction between ease of use of the software and ease of applying it to a task. However, this was not evident in the returns for the MM software where the correlation between these items was high.

Perceived Usefulness items

	PU1			PU2		
	Corr	No	Prob	Corr	No	Prob
PU1	*	*	*	0.8603	9	0.0015
PU2	0.8603	9	0.0015	*	*	*
PU3	0.9604	9	0.0000	0.8816	9	0.0008
PU4	0.9188	9	0.0002	0.8967	9	0.0005

	PU3			PU4		
	Corr	No	Prob	Corr	No	Prob
PU1	0.9604	9	0.0000	0.9188	9	0.0002
PU2	0.8816	9	0.0008	0.8967	9	0.0005
PU3	*	*	*	0.8954	9	0.0005

Table 7.12

All the items in the perceived usefulness construct have a correlation above 0.80, significant at $p<0.001$. This corresponds with the results from the ITUs for both the external organisations and the MM software, where the PU construct was stronger than the PEOU construct. It also seems to correspond with the results obtained by Davis (1989), although he does not present the actual figures for comparison.

Perceived Involvement items

	PI1			PI2		
	Corr	No	Prob	Corr	No	Prob
PI1	*	*	*	0.5866	9	0.0484

Table 7.13

The perceived involvement construct has a slightly higher correlation for this small sample than in either of the other two samples. However, the results are still only significant at the $p<0.05$ level.

• **Construct validation** Principal Component analysis with varimax rotation produced the results shown in table 7.14.

** Rotated Factor Matrix **

	Factor 1	Factor 2	Factor 3
PEOU1	0.3612	0.8165	0.1370
PEOU2	0.6603	0.3769	0.3439
PEOU3	-0.3151	0.9239	-0.1398
PEOU4	0.1014	0.9362	-0.1160
PU1	0.9742	0.0195	-0.1258
PU2	0.9608	-0.0539	-0.0702
PU3	0.9048	0.0279	-0.3630
PU4	0.9772	-0.0203	0.1329
PI1	-0.2067	0.5982	0.7501
PI2	-0.0203	-0.3222	0.9237

Table 7.14

The second item of the PEOU construct fails to load on that factor. It shows a greater correspondence with the PU construct. Again, this suggests that the wording of the item - "I would find it easy to get xx to do what I want it to do" may be interpreted as being able to use the software to fulfil a task i.e. easy to use in context, whereas the other PEOU items may be interpreted as referring to the software alone, not within a particular context. It may be that this potential for different interpretation is masked in larger samples.

• **Reliability** Internal consistency as measured by Cronbach's alpha was 0.79 for PEOU and 0.97 for PU. A score above 0.70 is generally deemed acceptable (Davis et al., 1989). The lower score for the PEOU construct reflects the findings of the other tests.

7.2.7 Discussion of the ITU

It is interesting that the apparent problem between the second and third items of the PEOU construct should reappear in this sample. Although the sample size used is very small, the same potential problem with these two items was flagged following the analysis of the 26 ITUs from the external organisations. However, in the case of the MM software the correlations between all the items in the contruct were moderately high and significant at $p<.001$. Although the constructs of PEOU and PU are reported to have been tested on several occasions (Mathieson,1991; Adams et al.,1992; Davis, 1993; Hendrickson et al.,1993), the number and wording of the actual items in the constructs has varied (Hubona & Blanton, 1996), which could explain why the potential problem with the wording of item two of the PEOU construct has not been flagged before.

The PU items return a consistently good correlation across all three studies. It may be that people find it easier to assess what would or would not be useful to them after a brief exposure than what is easy to use, preferring not to commit themselves to a view without having had the time to get to know the software. This interpretation might help to explain the apparently contradictory findings of other studies discussed by Hubona and Blanton (1996).

The two items in the PI construct are still relatively weak but comparable across the studies with the PEOU items.

The nature of the SN construct, which results in a unified item total, means that it is not a suitable candidate for the above forms of statistical analysis.

7.2.8 Interpretation of the ITU results

First of all it must be re-iterated that these results are from just nine people. Even if the questionnaire had been through several iterations with a much larger user base, results from such a small number of people should still be treated with caution. However, they are perhaps worth some consideration.

The mean of each construct is useful in that they give a rough indication of what may be considered as the approximate consensus score. However, it may be misleading as it may be affected by one or two extreme scores. The median, on the other hand, gives the mid point score for each construct. If the median and mean are close, it suggests that people are generally in close agreement.

The mean and median of each of the three constructs of PEOU, PU and PI and of the User Engagement construct (i.e. the total of the ITU) are shown in table 7.15 below. The 95% upper and lower confidence levels are also shown. Figures have been rounded.

On the PEOU and PU constructs a tick in box 4 of each of the items indicates uncertainty or non-commital. A total of 16 will result if the middle box is ticked all the way through. The PI construct consists of only two items and a total of 8 on this scale indicates uncertainty/non-commital. The total score of the ITU includes the SN score and gives an indication of the overall disposition towards the software. A score below 40 is positive and above 60 is negative. The area between 40 and 60 could be either positive or negative.

Construct	Value	Lower	Upper
PEOU Mean	13	11	15
PEOU Median	14	9	16
PU Mean	13	9	17
PU Median	13	5	18
PI Mean	6	5	7
PI Median	7	4	8
Total (UE) Mean	49	42	57
Total (UE) Median	47	39	64

Table 7.15

Perceived ease of use has a mean of 13 and a median of 14. These are both on the positive side of the scale, with even the upper confidence levels no higher than the 16 of the mid point. However, they are quite close to that mid/undecided score.

Looking at the upper and lower confidence levels of the median of the PU scale (18 and 5 respectively), there is greater disagreement about the perception of usefulness. However, the mean and median are both on the positive side of the scale which suggests an overall positive rather than negative view of the usefulness of the software.

Similarly, the perception of involvement is positive rather than negative.

All the scores, even the confidence levels, generally cluster towards the mid point. This suggests that people prefer not to give too definite a response at this point. It is perhaps not surprising that the overall UE score falls into the grey area which lies between 40 and 60. However, overall, the scores suggest a positive disposition towards the software.

• **Regression Analysis**

The results of the regression analysis performed on the ITU data is shown in table 7.16. As may be seen, none of the results has a significance of $p<.01$ or lower. These results could be due to the small number of people in the sample. It is also possible that the people in the PM group are rather independently minded. However, the R^2 value when perceived ease of use is the variable

Dependent Variable	Independent Variable	Coefficient	R^2	Significance
PEOU	PU	0.14	.06	.51
PU	PEOU	0.45	.06	.51
PEOU	PI	0.50	.05	.53
PU	PI	-0.89	.05	.54
PEOU	SN	0.27	.33	.10
PU	SN	-0.12	.02	.71
PI	SN	0.07	.09	.41

Table 7.16

dependent on subjective norm is interesting. One of the principle requirements of the group was that the software should be easy to use and the preference expressed for package A at the overview was because it was perceived to be easier. It is possible that, without hands-on use of the software, people are influenced by what they feel others think with regard to ease of use. For each of the other factors, the R^2 value is very small indicating that there is no link between them.

7.2.9 Implementation of replacement PM software

Following the decision to purchase package B, various negotiations with different resellers of the package took place. This process took nearly three months but had been allowed for in the schedule of software purchase and implementation. The PP and one other person were given initial training so that they could determine coding structures. This was done with the full agreement of the other people who would be using the software. The aim was to be able to transfer existing work without compromising any of the functional benefits that the software offered. At this point the PP was also provided with a copy of the software on a separate laptop so that others could have a preview of it if they wished. Just after this, it was announced that there would be twenty-percent job losses at the location, that the posts affected would be announced shortly and that volunteers would be asked for. This naturally had a disruptive effect.

The installation of the new software was planned to take place at the same time as hardware and operating system upgrades at the depot. It was agreed that the software being replaced would not be re-installed as paper copies of projects could be used to cross-check the data under the new system. Training was organised so that people were away on the training course whilst the new system was installed and everything was completed by their return. Despite the proposed job cuts the instructions were that the installation plan was to go ahead as it was not known which jobs would be lost. One result of the disruption caused by the announcement was that the reconstruction of the projects under the new system by the PP was not complete by the time the software was in place.

7.2.10 Distribution of the SUMI

The posts which would be cut were announced and volunteers asked for during the installation and training period.The introduction of the replacement software was not a cause of the job cuts and nor was it seen as such. The deadline for volunteers was extended but in the end all redundancies were taken by volunteers. The effect of this was to leave some people in posts which were no longer required and other posts which were vacant. This led to a shift around of people and posts which started six months after the installation of the software and was not predicted to be completed until the following year. The result was that seven months after the installation of the software only the PP was using it to any degree. Effectively he was constructing the projects as it was not known who would be responsible for using the software in the reduced number of teams which would be the final outcome of the re-organisation. One person had "played around" with it soon after installation and another had "had a look" but none started to use it in the day to day work. This was partly due to the ongoing disruption and partly because no major work was started in the period meaning that the software was not essential.

The result was that it was not possible to use the SUMI as no-one, apart from the PP, had the necessary experience with the software itself. It was only possible to assess by discussion the general attitude towards using the software should this be needed and whether people had felt involved.

7.2.11 Follow up discussions with users

a) **Ease of Use** The regression on the ITU suggested that people might be influenced by those around them in their perception of the ease of use of the software (although the results were not significant, the R^2 value was relatively high and should not be completely ignored (Pervan & Klass,1992)). The mean and medians of the PEOU factor suggested that there was a tendency towards the positive view rather than the negative. However, this could not be confirmed without people being able to refer to actual experience.

b) **Usefulness** Those who had seen the projects contructed by the PP with the software expressed the view that it looked complex. However, they also said that this was probably because of the detail required and that it was just a case of getting used to it. It is worthy of note that they themselves were asking for the level of detail given. This fact suggests that people did perceive that the software would be of use to them in their jobs.

c) **Involvement** The feeling was that the process by which the software had been selected meant that people were involved in the decision. Even though, on the basis of the software alone, they might have preferred package A, it was felt that package B was the better package overall.

d) **Subjective Norm** Senior management had been present at the overviews and had taken part in the discussions and the final vote. However, they had very much taken a back seat in the process and the main role was taken by the PP. There was a general respect for and confidence in the PP on the part of the other people who would be using the software, evidence of which was shown by the decision that he should be the one to visit the reference sites for the different packages.

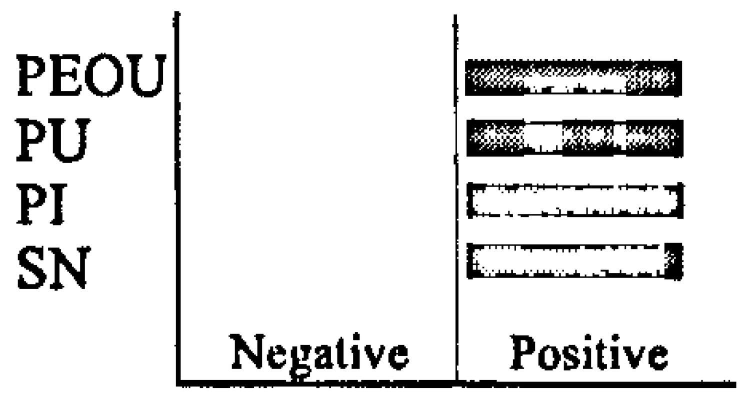
Overall, this response was in keeping with the results of the ITU. There was no great strength of feeling about the software but the overall view was positive rather than negative.

7.2.12 Independent Evaluation Although there were no SUMI results to compare it with, another evaluation of the software itself using the Ravden and Johnson (1989) checklist was undertaken.The same scoring method was used and the same person performed the evaluation as for the original PM software. This was because he was independent of the user group but was familiar with the tasks for which the software was used. It also ensured a consistency of interpretation of the checklist.

The method of scoring meant that the lower the score the greater the usability (as measured by the checklist). The result was a score of 232. This was a higher score than the 222 achieved by the package which had been replaced. There were seven ticks under Moderately Satisfactory, one under Neutral and one under Moderately Unsatisfactory (this on the Informative feedback scale). This compares with four ticks under Very Satisfactory and five under Moderately Satisfactory on the original package. This is surprising as newer products are generally thought to bring improvements. However, a verbal comment by the evaluator was that he preferred the original software and that perhaps the reason for it being no longer developed was that there was little to be done that would improve it and so no way of making money from it.

According to this evaluation, the replacement software was not an improvement on the original package.

7.2.13 Summary of PM case study



From the discussions and the results of the ITU it would seem that there was a change in the perception of the workgroup. Although the lack of use of the software means that it is not possible to say how people felt about the software itself post installation, no-one said that the process had been a waste of time. Whilst it might be thought that if people were keen to use it, they would do so regardless of the uncertainty about who would be in the posts that were required to use it, it must be remembered that the situation meant that no new projects were being planned. The ones which were still in progress were minor and were being converted by the PP. Anyone who really used the software would be doing so for its own sake and not because it currently served a useful purpose. Even so, it must be admitted that the impression given was that it would take a while for people to learn to use the software efficiently not because it was difficult per se but because there was a lot to it.

The PP said that people had initially pestered him to get the existing projects converted so that they could start to use the software. This suggests that people did indeed feel that the software would be of use to them. Had people really not wanted to use the software, they are more likely to have encouraged delay. This is a change from the use of the previous software which seems to have been resented.

The involvement of people at the overviews and discussions and ensuring they were aware of what was being considered with regard to the packages seems to have been appreciated. However, the move towards acceptance of the need for software to help with resource usage planning seems to have started before the investigation for a replacement package was started. This relates back to the first discussion with the PP where he said that he felt people's perceptions may have changed slightly since the issue of the SUMI. When the suggestion was made at the final meeting that a different way of working should be considered, people were surprised that this should be offered as an option. The surprise was that they should be asked directly whether this was something to be considered given that people seemed to think it had been easier to work before the existing software was introduced. Although there was a lively discussion, the result was that no-one could see any viable alternative. Whilst the discussion may have helped to highlight the constraints that people were having to work within and engender an acceptance of them, it seems likely that most people had accepted the need for the software before this but preferred not to openly admit it. The fact that the PP had done his best to use the existing software as a means of informing the project teams of clashes and to help them when they needed to negotiate a change of resource usage with other teams, seems to have created some confidence in the system. This meant that by the time a replacement came to be considered, people were willing to have less direct involvement in its selection and were happy to trust to the judgement of the PP.

Whilst it was not possible to validate the User Engagement construct of the ITU, the internal consistency tests were encouraging given the number of users. The overall score was in the grey area but towards the positive end of the scale and this again seemed to reflect the mood of the PM group who were generally positive about the software but not wanting to commit themselves to how easy it was to use without full experience of it.

7.3 Discussion of both case studies with reference to the model

The model of user engagement shows the perceptions of ease of use, usefulness, involvement and management and co-worker support as all contributing to the engagement of the user with software. The influence of each of these factors on user satisfaction with software was shown in Chapter 5 by use of the SUMI questionnaire and interview and discussion. The aim of the two case studies just described was to test whether those same factors affected an individual's engagement with software before having used it and whether the level of engagement at this early stage correlated with later satisfaction with the software.

The ITU questionnaire was used to measure the level of user engagement shortly after people had seen a demonstration of the software in the case of the MM2 users and following a training session in the case of the PM2 users. The results of the ITU for the MM2 software correlated negatively with the results of the SUMI questionnaire, which was issued after the software had been in use for some time.

The scoring of the ITU is such that this negative correlation indicates that it is measuring a similar construct to that of the SUMI. This suggests that PEOU, PU, PI and SN do indeed contribute to an individual's engagement with software before that software has been introduced.

Although it was not possible to repeat this exercise with regard to the PM2 software as organisational circumstances meant that the SUMI questionnaire was not issued, the results of the ITU, which fell into the positive side of the grey area on the scale, suggested a more positive than negative disposition of the users with the software pre use and this was supported by later conversation with some of that group.

However, despite the fact that the User Engagement construct correlated with later satisfaction, the influence of each of the factors in the model is not clear cut, as shown by the results of the ITU, and some exploration of each factor is required. In order to distinguish between the original and the replacement packages, they will be referred to as MM1 and PM1 for the original software and MM2 and PM2 for the replacement software in the following paragraphs.

7.3.1 Perceived Ease of Use and Perceived Usefulness

The perceptions of ease of use and of usefulness are the two principal factors in the Technology Acceptance Model (TAM, Davis et al., 1989). It is also generally acknowledged that people are more likely to use something that they find intuitive than something that they have to struggle to learn. Various studies, however, have shown that ease of use tends to be more important when people are first getting to know software and is less important as they become familiar with the software (Davis 1989; Davis et al., 1989). Perceived usefulness, on the other hand, is always important. As Davis (1993) says "..no amount of ease of use can compensate for a system that doesn't do a useful task" (p.484).

In the interviews for the MM1 and PM1 software a repeated complaint had been made that the software was too complex and not easy to use. Yet people had been trained in the use of both packages and had indeed been using them for a number of years. When use of software is mandatory it is quite likely that the importance of ease of use will be raised and it might be thought that this was the reason for the complaints. However, people were unable to say which parts of the software were difficult and made contradictory statements such as wanting both more information on a screen but for the display to be less busy.

The MM1 software was DOS based whereas the MM2 software was written for the Windows operating system. Although the MM1 software had scored quite well on the Ravden and Johnson checklist (1989), the evaluation of the MM2 software showed an improvement.

The results of the ITU for the MM2 users were just positive on the PEOU scale suggesting that they too perceived the software to be easy to use at the time of the overview. However, the SUMI scores on the Helpfulness and Learnability scales were lower than they had been for the MM1 software suggesting that it was less easy to use. By the time the SUMI was issued people had been using the software for some time and it could be expected that the initial problems caused by unfamiliarity with the operating system as well as the software would have been overcome.

It would appear from the discussions with the users that, even at the time of the SUMI, they felt that the software was easier to use in that it was Windows based and so allowed for ease of movement around the system. However, there was a lot to learn and it was too complex and this was the reason for the low score. When asked whether they had thought the software easier to use at the time of the overview, they again pointed to their perception that simply being written for the Windows operating system meant that the software was easier and that this had been enforced by the apparent ease with which the demonstrators had used it.

It must be said that the MM software was not simple. The nature of the task it was to help perform meant that data had to be held in a particular format (i.e. as a Bill of Materials, BoM) and that this had to be linked to spares. Jobs were associated with individual items in the BoM and with the whole structure. The tracking of items and time spent on work had not been carried out in a structured manner before the introduction of the MM1 software. It is understandable that the change of the way of working would mean that the software was not initially intuitive as people were relating it to a task which was carried out differently without the software. However, by the time of the introduction of the MM2 software, it might be thought that the new way of working had been assimilated and that familiarity with the new software was all that was needed. Yet the SUMI results indicate that this was not the case. In other words, the way of working had not been accepted during the period of using the MM1 software. In fact the results of the interviews and discussions for the MM1 software had indicated that people perceived the software to be awkward because they did not feel it was useful.

However, the results of the PU scale of the ITU for the MM2 software were just within the positive side of the scale suggesting that, before the software was introduced, people had a slightly positive perception of its usefulness. This perception seems to have altered by the time of the SUMI.

The regression results on the ITU suggested that there was a link between PEOU and PU. It is possible that, at the time of the overview people saw various facilities that they felt they could easily use and, as a result, felt would be useful. Alternatively, that they could see where a facility could be used and because of this understanding and being able to relate it to their work, the method of use appeared easier. The overviews took place away from the work context. People were being shown software running under an operating system they wanted to use and they were able to discuss how they might be able to use that software. It is possible to speculate that this had a positive effect on how they viewed the software and that this is what is reflected in the results of the ITU. However, this slightly positive effect was lost once the software was installed and people were left without guidance or leadership in how to make the best use of it. The discussions post the SUMI questionnaire revealed that people did not feel that they were able to make suggestions about how to use the software and that the situation remained the same as it had been with the MM1 software. Once again there was no incentive to put any effort into getting to understand the structure of the task and then relate that to the software. The battle of thinking one way and being forced by the software to act in a different way continued. In a sense, the fact that the software was not felt to be useful re-enforced the idea that it was difficult to use.

The results of the ITU on both the PEOU and the PU scales for the PM2 software were positive, although not strongly so. The regression results on these scales suggested there was no link between them. This might be explained by the small number in the sample compared with the MM users. However, it could also be that the PM workgroup had already been persuaded that a package to help with the job would be useful. The PM1 software had been resented because it was felt to be an imposition by management. However, the discussion at the time of choosing the PM2 software demonstrated that the way of working had been accepted. This was a grudging acceptance rather than an enthusiastic response but it perhaps meant that people were more inclined to judge ease of use and usefulness independently of each other.

It is clear from the SUMI results of the MM1, MM2 and PM1 software and from the interviews and discussions that followed, that people's perceptions of how easy to use and how useful software is will affect how satisfied they are with it. TAM (Davis et al., 1989) demonstrates that these factors are also influential before software is used. The results of the two case studies support the proposal that PEOU and PU will contribute to people's engagement with software before it is introduced. In the case of the MM2 software the effect was to lessen the overall negative disposition before the installation of the software because people perceived that it could be easy and of some use to them

7.3.2 Perceived Involvement

One of the reasons given in the HCI literature for the need to involve people in the introduction of new software is that doing so will help them to understand and appreciate the need for and benefit of the software (Damodaran, 1996). This implies that people's perceptions of the software may be altered by the amount of involvement they have in its development. The opportunity for involvement of people in the introduction of packaged software is rather more limited than it is when software is being developed in house. However, it was observed that the users of the IM software had been given the opportunity to see the different packages which could potentially fulfill the requirements and were able to express an opinion on them. They were the only group who said that they had been involved in the process leading to the purchase of the software and they were also the only group who were satisfied with the software (even though it achieved the lowest score on the Ravden and Johnson (1989) checklist). It was therefore proposed that if people are involved in the selection of a package they will be more positively disposed towards it.

At the outset of this research it was felt that some control could be exercised over the opportunity for involvement that was offered to people. It was hoped that at least with this variable there could be some consistency between the two case studies in the amount and degree of opportunity that was offered, so that the results could be compared. In fact, however, there were differences between the two case studies:

a) In the case of the MM software, a decision was made to consider the later version of the same software before looking at other packages. Although another package was considered, this was not seen by the workgroup members as it did not meet the functional specification. This appeared to be acceptable to the workgroup who said that they did not wish to suffer the hassle of converting the existing data. The PM workgroup saw demonstrations of two different packages one of which had been mentioned by the workgroup members themselves. The other was selected by the PP in conjunction with the IT department based on the key requirements list, market standing and cost.

b) The MM workgroup were able to vote on whether to take the later version of the existing software or to consider other packages but the vote was verbal via an intermediary. It is possible that people felt unable to express their real view in this way, although it should be pointed out that these same people had been quite vociferous in their complaints about the existing software at the demonstrations of the later version and did not appear to feel pressurised in any way. The PM software was voted for by those present at the final meeting.

c) There was no discussion about the way of working for the MM group. The PM group discussed the option of altering the way of work and even going back to the way they had operated before the software was introduced.

In both cases the users reported that they felt they had been involved in the decision over which software to purchase. In the case of the MM2 users, this sense of being involved and given the time to question the software supplier may have contributed to the initially more favourable disposition towards the software with regard to PEOU and PU (although the regression results of the ITU showed no link between these factors and PI). Hence no grave reservations about the software were expressed when people were questioned at the overviews. Yet once they were back in their normal work environment the situation returned to as it had been. The way of working was dictated and people felt that they could not make suggestions. The software became just another management imposition and not a tool for their use. This reflects the need expressed nearly twenty years ago by Eason and Sell (1981) for the people providing the data to feel identified with the system. Any sense of identification that may have been felt as a result of the opportunities given at the overviews was lost when people found that they were not encouraged to make use of the facilities they had thought beneficial at the demonstration.

The PM users were slightly different. They were asked during the final selection meeting whether a different way of working should be considered. This caused some discussion but the conclusion was that the way of working was, in fact, necessary under current staffing constraints and that, therefore, there was no choice about which package should be selected as only one would assist the PP. It is possible that the open discussion helped people to see that, despite their proclaimed dissatisfaction with the PM software, it was a necessary tool. This may in turn have influenced their perception of the usefulness of the PM software although this is not apparent from the regression results on the ITU.

This again raises the question of what it is to be involved especially when a COTS package is to be purchased. The multi-national organisation which was asked to take part in the validation of the ITU but which was only able to return the SUMI questionnaires (Organisation B), had used the Dynamic Systems Development Method (DSDM) to introduce a package which could be tailored by the users to better fit their needs. The aim of DSDM is to ensure user involvement whilst at the same time ensuring that targets are adhered to. However, the results of the SUMI were a poor global score of 35.

Some caution must be exercised in interpreting this low score as only 27 out of the 60 questionnaires were returned. It is possible that the people who were happy with the software did not return their questionnaires. However, of the 27 people who did return the SUMI, ten had been ambassador users and should, in theory, have felt more favourably towards the software by virtue of their involvement. The contact at the organisation also said that he felt that the SUMI results did indeed reflect the general feedback on the software. It should be mentioned here that it seems unlikely that the package was just poor software. Whilst it was not possible to perform a separate evaluation of it, its sales and background suggest it was at least average of its type.

Although the intention of DSDM is to involve users in the process and ambassador users were nominated in Organisation B, the contact reported that the people nominated as ambassador users were expected to fulfill the role in addition to their daily work. The result was that there was low attendance at the workshops. Decisions were therefore made about the presentation of the software without their input and they and, as a result, their colleagues were unaware of what was happening. This lack of enthusiasm might suggest that people did not wish to be involved and that they preferred to be presented with the final product. However, another explanation presents itself, which is that the lack of enthusiasm was a sign that people did not think that the software would be useful to them. They had not been involved in any discussions or dissemination of information that led to the decision to introduce the software. That decision had been made by senior management in order to fulfil a specific target. People had not been involved in any discussions about how to meet the target. They were just presented with the decision to purchase. In such circumstances, it is perhaps understandable that people might not feel that the software had any direct relevance to them and their work. This had also been true of the MM1 and PM1 software in the organisation under study. In both cases, the decision to introduce the software had been made without reference to the end-users and for reasons which were neither explained nor discussed with them. It is interesting to note that the Control and Efficiency scales are the two lowest scorers on the SUMI for the Organisation B users (35 and 32 respectively) supporting the suggestion that people did not perceive the software to be useful to them. Similar results were true of both the original and replacement MM software.

The lack of involvement for the users in Organisation B, then, was at a stage earlier than the implementation of the software. It was lack of involvement in the decision that a software tool was needed in order to fulfill a particular purpose and which tool to use. This lack of involvement by workers in board room decisions is normal for all but the smallest organisations in the UK and probably elsewhere in Europe. Although the PD philosophy at its most pure advocates that any decision that may affect the way of working should not be made without reference to the people doing the work, in practice it may be seen that the question addressed has usually been how, not whether, to introduce technology. As with any other decision that results in a change in the workplace and in which the affected parties have not been involved, the decision to introduce software may result in feelings of fear, resentment and uncertainty (Hirschheim & Newman , 1988). This means that there is likely to be some antipathy towards the software even before it is introduced.

If the software is to be developed in-house these effects may be mitigated by involving people in that process and, indeed, participation by users in the design of a new system tends to be viewed as a more acceptable approach for organisations than involvement of workers in board level decisions (and indeed forms the bulk of the reported research into participatory methods). However, such involvement is not possible in the case of COTS packages as the software has already been designed. Involvement in the selection of a new package appears to have had a positive if transitory effect on the perceptions of the users of the MM2 software. That positive perception was not built on by continuing their involvement in decisions of how best to use the software once it was introduced. In the case of the PM2 software, people seem to have come to understand the benefit of using a software package as a result of the work of the PP with regard to the PM1 software. Their involvement in the selection meeting for a new package gave them the opportunity to talk through the situation and find that they were in general agreement with each other as to the need for a package. In a sense, their involvement had commenced in advance of the search for a new package.

It is clear from the two case studies that involving people in the selection process was appreciated and it appears to have had a beneficial effect, thus supporting the proposition that people will have a positive disposition towards it. However, it is clear that this involvement needs to go beyond involvement in the selection of the software. Whilst the response on the ITU to the PI questions for the MM2 software was positive, the involvement to which it referred was very restricted and a positive response was not sufficient to ensure positive engagement with the proposed software.

7.3.3 Subjective Norm

The potential for the influence of referent others forms a key part of the Theory of Reasoned Action (TRA, Fishbein & Ajzen, 1975). The model of user engagement proposes that an individual's perception of the response of referent others (i.e. managers and co-workers) to a software package will influence the individual's own disposition towards that software.

The need for communication between management and workers about the introduction of new technology is fundamental to Soft Systems Methodology (SSM, Checkland, 1981) and the ETHICS method (Mumford, 1993) amongst other participatory methods. At the same time, there is a tacit recognition even by the advocates of participatory methods that organisations do not consult their workers about organisational changes and, indeed, as Eason (1987) points out, the introduction of technology may entail changes that have not been anticipated by the organisation.

It was reported by the contact in Organisation B that one manager had been very enthusiastic and encouraged use of the software during the tailoring phase. However, it is evident that people did not take up the opportunity for involvement at this stage. The fact that the work demands on the ambassador users were not reduced in order to enable them both to attend workshops and complete their normal job suggests that the software was not given a high priority by their managers. This lack of encouragement is also evident with regard to the MM2 software.

Although there was agreement by management that people should attend the overview of the MM2 software, there was no perceived encouragement by the local managers in the use of the software once it had been installed nor any lead from the CM team. This potential lack of perceived support was indicated by the SN score on the ITU. It would seem that the MM2 software was introduced into an environment no different from that of the MM1 software and any benefit gained by the attempt to involve people during the selection process was, therefore, lost.

In the case of the PM software, people seem to have been persuaded by the PP of the need for the software and perceived some benefit to themselves. Senior management no longer felt the need to enforce the use of the software as this was starting to happen almost voluntarily as a result of the actions of the PP. In a sense the PP had taken on what might be thought of as the management role with regard to the software even though he wasn't a line manager to the PM users. He was the person who would chase up and to whom people would refer if they had a problem. There was no-one in a comparable position in the MM workgroup. The PP parallels the manager in overall charge of the IM software. Hence by the time a replacement package came to be considered people no longer felt the same antipathy towards use of the package. A similar circumstance could explain the reason for the SUMI score for the IM package. This package was itself a replacement for one with which people had not been satisfied. The fact that the manager in charge had started to discuss with people why a package was needed and had listened to their complaints perhaps meant that they were more receptive to a new package even before one was selected. In the case of the PM workgroup, this change in user perception was not sufficient to alter their dissatisfaction with the software in use but the indication from the ITU was that the new software would be introduced into a more favourable environment.

The importance of the perception of the support of referent others is noted by Martinko et al. (1996) and is indicated in the regression results of the ITU for the MM2 users. In these results a dependency is shown between both SN and PEOU ($R^2 = 0.42$) and PU ($R^2 = 0.48$) for the 23 people who completed the ITU. When the same calculation was carried out just for the people who also completed the SUMI (minus the one who had given responses at variance with his actions, as described in section 7.6) the R^2 value increases to 0.63 $p>0.001$ for both PEOU and PU. Whilst too much should not be inferred from these results given the small user sample, it was clear from the comments of these users that their perception was that there was little support from local management for use of the software. They gave no positive encouragement. Similarly, there was a perception that none of their co-workers was enthusiastic about its use.

The situation is slightly different with regard to the PM users where no dependency of other factors on SN is indicated. The position of the PM workgroup members in the organisational structure was different from that of the MM workgroup members. Although they did have line managers, these people were higher up the hierarchy than were the line managers of the MM workgroup and the PM members themselves were higher up than the MM members. They also had slightly more autonomy in their pattern of work. It is possible that these factors influenced the strength of the SN factor as people were not so affected by those around them. Hence there was no significant regression from the ITUs. However, the R^2 value for perceived ease of use was comparatively high. PEOU was also the factor that had most importance for the PM users by this time as they were already convinced of the need of the software. It is possible that the individual perception of ease of use was indeed influenced by how they perceived others to be reacting at a time when they had little experience of the software.

The need for people to at least understand why a software package should be introduced even before the software is chosen has been suggested as one reason for the difference in satisfaction levels between both versions of the MM software and the IM and PM2 software. This understanding is not obtained as the result of a one-off exercise. It requires constant communication between a supportive management and the users. In the case of the IM software user involvement extended beyond the original purchase and installation. Concerns about the way of working and the need for certain steps continued to be discussed and explained. A similar pattern may be seen with the PM1 software. When a different person took on the role of PP, greater communication about the use of the software took place even though the software had been in use for some years by then and this appears to have improved people's attitude toward using the software if not towards the software itself. This shows that concerns about the way of working required by the software are not addressed simply by ensuring people understand how to use a software package. Whilst training is essential if people are to be able to use the software properly, it is not sufficient of itself. Ongoing support is also a major requirement if people are to be able to assimilate the new system and to turn it to advantage (Damodaran & Eason, 1981). In the studies reported, full training was always provided. However, ongoing support was limited to Help Desk support from the IT department for specific software problems. There had been no on-going review between managers and users of the MM1 and PM1 software at the time of the first SUMI and the situation did not change for the MM users when the MM2 software was installed.

Thus it would seem that user involvement and management support are bound together. The involvement of the users in the decision over which package best meets the needs of the organisation is important in order for people to feel that consideration has been given to their requirements. However, just involving people at this stage is not sufficient and management support and encouragement is needed throughout the time that people are becoming familiar with new software. In fact, with regard to ongoing satisfaction "..the implementation phase is the time which is most critical..It is a phase when user participation is crucial.."(Eason, 1988, p.186). In the case of the COTS system especially, this phase of the software introduction is very important and user-manager co-operation is essential if the user is to understand the needs of the organisation and the manager is to understand what is possible for the user. Whereas when a system is being designed in-house or technology is being introduced for the first time, it might seem fairly straightforward to involve people in discussion about how the software will look or where PCs/terminals will be placed and thus ensure some involvement, this is not necessarily true of a COTS package. The hardware is often already in place, the software has already been designed. There may be a tendency on the part of organisations to think that it is merely a case of installing the software and providing training. Yet the results of this study show that implementation, in the sense of addressing problems and concerns that arise as a result of software installation is still a critical time when user participation and management support is crucial.

Twelve years after Eason (1988) made that point it may seem strange that this need has not been generally recognised and perhaps the organisation under study is abnormal. However, a more recent report on the performance of IT in organisations draws attention to continuing lack of awareness on the part of managers of the need to involve people in the introduction of new software and to give active support during that process (Clegg at al., 1997). Having purchased a package the main emphasis is perhaps still on the software working rather than on the people working with the software.

The model of User Engagement proposes that the perception of management and co-worker support will affect an individual's engagement with software, and the evidence of the two case studies provides support for this being the case both pre and post the introduction of a software package.

7.3.4 The ITU questionnaire

Some reliance has necessarily been placed on the ITU in assessing the level of user engagement before the introduction of the software packages. Whilst the scores obtained for the four factors of PEOU, PU, PI and SN and for the overall construct of User Engagement were all supported by the results of the interviews and discussions and, indeed, the results of the SUMI, the ITU must be regarded as still being under development and is, therefore, worthy of some further discussion.

The User Engagement score for the MM2 software was only just within the negative area whilst the SUMI score was quite definitely negative. An exact correlation between the SUMI and ITU scores was not anticipated. The aim of the ITU is to give an indication of the general disposition of people towards the software without having used it. By the time of the SUMI they have had the benefit of use within the normal work environment but by this time a fixed opinion of the software may have been formed which is difficult to alter. The ITU is intended to be used as a flag for potential problems so that corrective action may be taken before or during the introduction of the software.

The PEOU construct was a little weak with regard to item 2 ("I would find it easy to get xx to do what I want it to do"). In the results both from the external organisations and the results for the PM2 software it appeared that this item might be interpreted as meaning that the software matched the task to be performed and so was useful, rather than that it was easy to use of itself. However, it might be argued that match to the task is a part of ease of use when use of the software is mandatory and this apparent weakness was not present with the MM group. Even so, the indication is that further work is needed on this construct.

The PU construct was consistently strong. It is possible that people also feel more able to judge the usefulness of a product after a short exposure to it than to judge how easy it is to use.

The PI construct focussed on a narrow view of involvement i.e. that of being involved in the decision whether to purchase a specific software package. This was because it was the only variable over which some control may be exercised by the person tasked with software purchase. The decision that software is required has already been taken by this stage and the software purchaser himself may or may not have been involved in that process. The earlier discussion has indicated that there is much more to involvement than simply voting on the final selection. This raises the question of whether the inclusion of the construct in the ITU actually adds any benefit. In the case of the MM2 software the effect was to lower the overall ITU score, thus making it more positive. However, discussions with the users indicated that a more positive disposition towards the software was indeed the short term effect of the involvement they had. The meaning of involvement and what people understand by it has been shown to be very complex and subject to multiple interpretations (e.g. Doll & Torkzadeh,1989; Hartwick & Barki, 1994). Whilst a more detailed measure of involvement would improve the accuracy of the ITU the construction of the items would require much attention and careful testing.

The Subjecive Norm construct is one of the most difficult to measure. SN is defined as "the person's *perception* that important others desire the performance or nonperformance of a specific behaviour" (Ajzen & Fishbein,1980, p.57). This perception might be determined as a result of the person's inference about the referent's perceived attitude towards the behaviour (Fishbein & Ajzen,1975). It is possible that someone may perceive that an important referent has an attitude towards a behaviour which conflicts with the individual's own attitude towards that behaviour. In order to allow for this, account is taken of the individual's motivation to comply with the wishes of the referent other. It is this "motivation to comply" which Fishbein & Ajzen (1975) describe as the least understood term in their model. This is because of the many possible influences on the motivation to comply - such as the power of the referent other over the individual and the individual's liking or respect for the referent other. Within the context of the workplace, the assumption has been made that line managers and co-workers are the people whose opinions are most likely to affect the individual and the SN items specifically refer to these people. In the case of the MM users this measure gave a reasonable indication of the influence of people's perceptions of their managers' and co-workers' attitude towards use of the software. However, in the case of the PM workgroup, the person who exerted most influence was neither a line manager nor a co-worker, in the sense of being a user of the software on the same basis as the others. Although he was also a project manager, his role as overall project planner took him beyond that of the rest of the workgroup and meant that he was also responsible for sensible use of resources. Whilst this could be seen as giving him some power over the other users, in fact his decisions could be over-ruled by their line managers. He was, however, well liked and respected and, under the definition of the construct, an important referent. The wording of the SN construct means that this influence was not recorded.

Fishbein and Ajzen (1975) recommend that time is spent in determining who the people are who are important referents as this will vary according to circumstance. However, to do this in advance of issuing the ITU could be very time consuming and it may be more practicable to assume that the individual's managers are likely to have some influence. Whilst this could result in an incorrect guideline result from the ITU, it would seem more likely that positive support from a referent other would either add to a supportive management, thus making the overall user engagement score more positive or would mitigate against an unsupportive management again making the overall user engagement score more positive.

The grey area of the scoring on the ITU is itself problematic. The span is rather wide as a result of the method of measuring SN and this needs further work.

7.4 Conclusions

The results of the two studies suggest that, whilst ease of use of the software is important, it is not the only factor which affects user engagement with software. The findings support those of Davis et al. (1989) that perceived usefulness is more important and that this is the case even when there is no choice but to use the software. Both perceptions would seem to be influenced by the perception of the views of referent others. This means that it is important for the person tasked with procuring a software package for use in an organisation to not only consider the merits of the software itself but to also ensure that there is a supportive environment in which to introduce it. It is perhaps all too easy for organisations to blame lack of benefit of new technology on the limitations of the technology itself rather than consider whether the problems might also lie closer to home. Whilst it is right that software developers should take account of the users of their products, organisations also need to take account of the users and build in the time for them to become familiar with a different way of working and to ensure that the need is understood.

To persuade an organisation to spend as much or even more money on the introduction of packaged software than for the software itself is a difficult task (Grudin, 1993). This is because the necessity of doing so is not evident. However, if a warning is flagged of a lack of user engagement before the software is introduced, the need for the extra time and effort may be accepted.

The studies reported suggest that the ITU may be used to give guidance on how people are responding to the introduction of a package. It does not give a detailed analysis of what action needs to be taken, nor is intended to do so. The aim was to develop a simple tool which would indicate if there was a potential problem. The User Engagement Model indicates the four key areas which may need investigation and the results of the ITU may be used to assess the importance of each. This needs to be performed by someone who is familiar with the organisation in order to be able to interpret the results and the questionnaire should not be used in isolation but in combination with follow-up discussions and interviews. In this respect it is similar to the SUMI questionnaire.

Although, as discussed, further work is required, the results of the two studies described suggest that it is possible to measure user engagement with a software package before its introduction and that this may be done cheaply and simply. The Model of User Engagement used in conjunction with the ITU, is a step in that direction.

Chapter 8

Summation

The previous chapters have described the process of developing and testing a model of User Engagement with packaged software before that software has been introduced into an organisation. That work is summarised in the first section of this chapter and the contribution made by the research is then detailed. Section 8.3 summarises the weaknesses present in the research reported. Finally, section 8.4 suggests future work.

8.1 Summary of work

The purpose of this thesis was to determine whether it is possible for an organisation to discover in advance of introducing a software package, whether the people who will be using it have a favourable disposition towards it. If people are favourably disposed towards the software they are more likely to have a positive attitude towards handling problems and greater efficiency of use should result. The type of software is a multi-user application such as inventory management or purchasing. The question arose as a result of the author's own experience working in an organisation which does not develop software in-house. This is a common phenomenon in small to medium sized organisations, which do not have the necessary finance to develop bespoke systems and so are not able to involve users in the design of software. Even so, a heavy investment is often made in the software purchase.

The research started with a review of the literature into how people interact with computers and thence how to improve that interaction. A brief overview was given in Chapter 2. This served to indicate the global areas for consideration with regard to user satisfaction which is the focus of Chapter 3. These factors were further considered in Chapter 4 and a simple model of User Engagement, i.e. of those factors which influence an individual's disposition towards software, was proposed. This User Engagement Model draws on the Theory of Reasoned Action (TRA, Fishbein & Ajzen, 1975), work in participatory methods and techniques (e.g. Mumford, 1991, Ehn, 1993) and the Technology Acceptance Model (a predictive model of voluntary use of software),(Davis et al., 1989). It is based on the perceptions of ease of use and usefulness of the software; of the user's feelings of having been involved in its purchase; and on the user's perception that there is management and co-worker support for the use of the software.

There were four steps to testing the model:

- Establish current satisfaction with software and compare with the User Enagagement Model
- Develop an instrument to measure the User Engagement construct
- Measure User Engagement before the introduction of a software package and the level of Satisfaction post introduction in a case study
- Repeat the process in a second case study

• Step One

The first step is described in Chapter 5. Three packages were selected. These represented a cross section of the applications in use in the organisation and ran under three different operating systems. The SUMI questionnaire was used to measure satisfaction. The package which scored well on the satisfaction scales (IM) was based on the miniframe and was accessed by people using PCs with terminal emulation. This meant that the interface presented was necessarily restricted and inflexible from the user's point of view. Neither of the other packages (MM and PM) reached the satisfactory score. Yet one of them (PM) was a market leader of the time with a graphical user interface (GUI).

An independent evaluation of each of the packages using the Ravden and Johnson (1989) instrument ranked the packages quite differently from the satisfaction levels reported but in line with what would intuitively be expected, with the GUI software scoring the highest and the miniframe software the lowest. There was also some evidence to suggest that, in other organisations, the MM software would be given a higher satisfaction rating than that achieved in the organisation under study. Semi-structured interviews and informal discussions with people confirmed that the level of satisfaction was reflecting more than dissatisfaction with the software product itself. There was a sense that the software had been imposed and was not primarily intended to be a useful tool for the individual in the case of both the MM and PM software.

When the methods of selecting and introducing each package were considered, some differences were found in the amount of involvement of the users and the perceived management and co-worker support for each package. Interviews and discussions also supported the findings of the SUMI with regard to the difference in how the different user groups viewed the ease of use and usefulness of each of the packages:

• The users of the IM software were involved in the requirements gathering process and were invited to the overview of potential packages and were able to ask questions and to say which package they preferred. The users of the MM software were involved in the requirements gathering process but the users of the PM software were not. Neither the users of the MM nor of the PM software had been allowed to attend any software overviews. The software was selected by senior management according to their own requirements.

• The actual IM package selected was not the first choice of the user group but the reasons for selecting it were explained and were accepted by the user group. There was one manager who spoke regularly to the user group both before and after the purchase of the software. Although he was not their direct line manager, he was viewed as being so with regard to the software. This show of management support was totally lacking in the purchase and implementation process for the MM and PM packages. In the case of the MM software, the direct line management were at best indifferent to the system and, in one case, against it. In the case of the PM package, the software imposed a standard way of working and its imposition without any reference to the users was resented. The views of those interviewed in the IM workgroup were that their co-workers generally liked the package whereas those interviewed in the MM and PM workgroups felt that none of their co-workers thought much of the software.

• In the case of the IM software, the users perceived it to be useful to them in their work and not difficult to use. In the case of the MM and the PM software the sense that the software had been introduced to fulfill the requirements of senior management seems to have resulted in people feeling that it was of no direct use to them. The software was also reported to be complex and problematic to use. This reflects what has been reported in the literature for a good many years e.g. "Many systems have been rejected or poorly used because users felt the system was being imposed upon them" (p.374, Damodaran & Eason, 1981).

These differences are summarised in table 8.1.

IM software	PM and MM software
a) Preview of alternatives and vote in selection	a) Package chosen by senior management alone
b) Encouraging, supportive management	b) Uninterested or coercive management
c) Perceived not difficult to use	c) Perceived problematic in use
d) Useful in the job	d) Of use to other people

Table 8.1

The first and second items (a & b) translate into the PI and SN constructs of the model and the third and fourth (c & d) into the PEOU and PU constructs. These results lent support to the hypotheses proposed and thence to the model of User Engagement.

• **Step 2**

The second step, reported in Chapter 6, was to develop an instrument by which the User Engagement construct of the model could be measured. The questionnaire approach was selected and the ITU questionnaire was developed by building on existing questionnaires, specifically those which had been used in testing and comparing TAM (Davis,1989; Davis et al.,1989). In order to validate the four measurements within the ITU questionnaire (perceived ease of use; perceived usefulness; perceived involvement and subjective norm) and the overall construct (user engagement), external organisations were asked for assistance. As the questionnaire deals with people's perceptions of software prior to having gained experience of it, it was necessary that this was the situation in the organisations approached.

The nature of the ITU questionnaire means that it must be issued after people have seen the software but before they have started to use it in their day to day work. Validation of the User Engagement construct was dependent on the SUMI questionnaire being completed by the same people for the same software once they had started using it. Although three organisations agreed to assist with the questionnaire validation, the result was only twenty-six usable returned ITU questionnaires and seven usable returned SUMI questionnaires. The statistical analyses performed to validate the questionnaire were promising although some questions were raised which could not be answered without access to the individuals themselves. The correlation between the SUMI global scale and the ITU total was positive but not significant. Statistical significance means the probability that the results obtained are likely to be due to chance alone. It is, therefore, generally encouraging to obtain a probability value of 0.05 or lower as this indicates a low probability that the results are due to chance. However, such results are difficult to obtain with small samples such as used here and should be regarded with caution even if they are obtained as there could be a population bias (Pervan & Klass, 1992). The fact that the correlation was positive rather than negative required investigation regardless of the test of significance as the scoring of the ITU and SUMI meant that a negative correlation would indicate a similar measure. An inspection of the individual results of the ITU and SUMI revealed that there was a wide variation of scores on the SUMI and that all the ITU scores fell into the grey area between 40 and 60 meaning they could be positive or negative. The removal of one user who had reported that he had no co-workers (perception of co-workers' attitude forms part of the SN construct) resulted in a small negative correlation but which was still not significant. As the sample was so small and the responses so varied and the questionnaire demonstrated internal consistency, it was decided that no change to the ITU could confidently be made at this stage.

• Steps 3 and 4

Steps 3 and 4 are reported in Chapter 7. The aim was to test for the presence and effect of the factors in the model at the time of selecting a replacement for the MM and PM packages and again after the software had been in use for some time. Data for the presence of the factors at the time of selection was gathered by use of the ITU questionnaire and followed up with interview and discussion.

The use of the ITU questionnaire at this stage may be queried given the results of Chapter 6 as the instrument could not be said to be fully validated. If a questionnaire is to be used with a reasonable level of confidence it must be shown to produce reliable results on more than one occasion.

According to Hendrickson et al. (1993) "An instrument's reliability can be examined by using tests of internal consistency, replication with different samples, and test-retest using the same sample"(p.227). Internal consistency had been demonstrated with the sample from the external organisations. There is a problem with the test-retest method in the case of the ITU as the questionnaire should be used before people are using the software in the workplace, which limits the time period in which it may be administered. Thus it was not feasible to re-administer the ITU to the same sample of people from the external organisations. Replication with different samples was achieved by its use in the two case studies and the statistical analyses were re-performed. However, this means that the instrument being used to measure the factors and overall construct of the User Engagement model, was itself still being validated. Had the ITU been the only means of testing the model this might have been problematic. However, interview and discussion were also used thus forming a supporting triangle of evidence.

During the selection process for the replacement MM and PM packages (MM2 and PM2) an effort was made to involve those people who would be using the software as part of their job. The management structure remained as it had been in the case of the MM workgroup. At the start of the replacement process there had been a slight change in the structure of the PM workgroup. Following the installation of the replacement there was a complete re-organisation of this group which meant that the study could not be completed. Even so, it was possible to get a general feel for the attitude of people towards the software. Table 8.2 below summarises the results. From this it may be seen that, with regard to the MM2 software, the only change from the MM1 software was that people saw an overview of the proposed software before purchase and were able to cast a vote regarding its purchase. The perception of ease of use and usefulness was little different from what it had been with regard to the MM1 software and there was still a perceived lack of management and co-worker support for the software.

MM software	**PM software**
a) Preview of package and vote in selection	a) Preview of alternative packages and vote in selection
b) Uninterested and perceived coercive management	b) Supportive lead user. Background senior management
c) Easier than previous but more complex and still not easy	c) Unknown
d) Of use to other people	d) Useful but due to organisational constraints

Table 8.2

In the case of the PM software there had been a change in the perception of the need for the software. Rather than its use being viewed as an imposition by senior management, it was seen as necessary because of the constraints on use of resources.

The SUMI result in the case of the MM2 study was that the replacement package scored no better than the package it was replacing. This was also indicated by the ITU results. In the case of the

PM2 study, the indication was that the score would have been higher than for the original package although the SUMI could not be issued to confirm this. The results of the ITU were indeterminate for this package with the score falling in the grey area, but the indication was positive rather than negative engagement.

The results from both studies supported the hypotheses proposed in constructing the User Engagement model. However, the discussion in Chapter 7 suggests that there is a close link between the involvement and subjective norm constructs. When new technology of any description is introduced into an organisation it is likely to bring about change. If people can not understand or accept the need for this change they will resent it. It is clear from the results of the PM2 study that people had started to accept the need for a software tool even before the replacement exercise. This was not due to the length of time that they had had to use the software and thus a reflection of resignation rather than acceptance. Most of the change in perception seemed to be largely due to the work of the PP. Rather than just demanding the data, he had shown them how it was used to their benefit. Whilst they might have preferred not to need the software, they had accepted that organisational constraints meant that it was needed.

No such change had happened in the case of the MM users. Use of the software had continued to be seen as an unnecessary chore through all the years of its use. There was a lack of discussion about where the benefits were or how things could be improved. All decisions were perceived to be taken by senior management for their own purposes. The local management did not perceive a benefit to using the software and with the lack of a respected champion of the software their view influenced that of the users.

The discussion in Chapter 7 drew a parallel between what had happened with the PM workgroup by the time of the selection of the PM2 software and the situation with the IM users. In the case of the latter, the IM software had itself replaced a package which was disliked. In the time before and after its introduction a well-liked and respected manager had discussed the need for the software and how best to use it. It would seem that, due to the input of this manager, the IM group had come to view the software as a tool which was needed to enable them to do a job rather than as a burden imposed for the benefit of others. Similarly, due to the work of the PP, the PM workgroup had come to accept the need for the software. This change in view with regard to the need was not sufficient to change the way in which the existing software was perceived, but it did mean that there was a more positive disposition towards a replacement even though it effectively did the same job. It is likely that their involvement in the selection of a replacement, and the fact that the preferred choice was not rejected out of hand but rejected by them on the basis of the extra effort required by the PP to use it, strengthened the apparent positive feeling by giving people some control over which tool to use. However, this involvement on its own is not sufficient as evidenced by the MM group.

This leads to the suggestion that people must feel that a software package would be a useful tool. Involving them in the selection of that tool is also necessary if they are to feel positively disposed towards it, but it is not sufficient. When people have been working in a manner which appears to be perfectly satisfactory, the introduction of a software package may seem to be an imposition for the benefit of others, as was the case with the MM1 and PM1 software. This would be true even in the case of the purchase of a new package to enhance existing software in order to have more functionality (as was the case with Organisation B). Any difficulties with the software will be seen as obstacles to working and the software itself will be viewed as awkward and time-consuming to use. In such circumstances, changing the package will not change this perception even if people are involved in the change, as shown with the MM2 software. The perceived support of referent others for use of the software by explanation and discussion is what is important. At the same time, a change of package may be necessary even if the perception of the need has changed, so that people may be given the opportunity to be involved in the selection of the tool they will use, as was the case for the PM workgroup.

Thus involvement in the selection of a tool and perceptions of the views of referent others (i.e.subjective norm) regarding it are closely linked. Involving people means not just giving them choices but listening to their concerns and ensuring that they understand the reasoning behind the decisions of others. This means a continuous dialogue between management and users before, during and after installation. People may report positively on their involvement in choosing a package but unless they also believe there is a requirement for it, the overall level of engagement will be low as shown by the ITU results for the MM group.

This acceptance of the need for the software and, therefore, of its usefulness also seems to have affected people's view of the ease of use of the software. Whereas, in the case of the MM workgroup, the results of the ITU indicate a link between the perception of ease of use and of usefulness, this does not appear to be the case with the PM workgroup. The suggestion is made that the link between PEOU and PU for the MM workgroup was a mix of being able to relate an aspect of the software to a task they performed and thus perceiving it as easy, or, on the other hand, perceiving something to be easy and thereby seeing a use for it. In the case of the PM workgroup, they already felt that the software was needed. Being able to relate functions in the software to the task to be performed was what they expected. Usefulness was, therefore, regarded as something distinct from ease of use.

The conclusion reached in Chapter 7 is that, although it is clear that further work is required on the PI and SN constructs of the User Engagement Model, the work described provides evidence for the validity of the model and that the factors of Perceived Ease of Use, Perceived Usefulness, Perceived Involvement and Subjective Norm combine to form the level of User Engagement with a software package before its introduction. Also that the level of user engagement with the software before its introduction may be measured by means of the ITU questionnaire and is indicative of the level of later satisfaction with the software.

8.2 Contribution made

The aim of this thesis was to develop a simple but predictive model of User Engagement with commercial off-the-shelf software (COTS). The purpose was to thereby assist purchasers of a COTS system in determining the level of user engagement with a package prior to its introduction. The financial outlay for a COTS package may be quite considerable and a positive disposition of the users towards the software is more likely to result in its being used to the best advantage.

The introduction of packaged software into an organisation is different from software which is designed in-house. In the latter case, there is the opportunity for an organisation to involve people in discussions regarding how the system may impact their work and to design the system and work pattern on a best fit basis. It is in this area that most of the work in the literature on participative design techniques has concentrated. Such an opportunity for involvement will often not exist with regard to COTS software as the organisation retains the standard package with no tailoring. However, although it is possible to draw up a functional requirements list for a COTS package it is inevitable that none will exactly fit the way the organisation currently operates. COTS systems are designed on a one size fits all basis which means that some parts will fit better than others. How well different parts fit will not become apparent until the software has been installed. It is, therefore, not possible to ensure that there will be no disruption to work patterns as a result of the software. Nor is it always economically feasible to introduce the software as a pilot in order to assess the implications in advance of a full scale roll-out. In such cases a positive disposition towards the software on the part of the users will mean that alterations to work patterns are more likely to be accepted or even suggested, whereas a negative disposition will result in dissatisfaction and minimal use of the software (Eason & Damodaran, 1981).

The work in this thesis moves the focus from in-house development to the purchase and implementation of packaged software. It considers the factors which affect satisfaction with software in use and then examines whether those same factors influence people's disposition towards software before use. In doing this it re-examines the concept of involving people when the design and development of software has already been completed and its findings underline the need for ongoing support from management both pre purchase and post installation.

Involvement may happen at many levels - from information provision to participation in decision making (Damodaran, 1996). The work in the thesis demonstrates that it is possible to involve people at different levels as much with regard to packaged software as to in-house developed software and that this may help to engender a positive disposition towards it.

The discussion in Chapter 7 proposes a link between an individual's perception of involvement and his perception of the support of referent others. Involving people in the steps to deciding which package to purchase and even in that final decision, whilst beneficial, is not sufficient as evidenced in the MM2 case study. People must also perceive that others who are important to them in that context are also positively disposed towards the software. In the example given from Organisation B, the attempt by the organisation to involve people in the tailoring of a package failed because people did not perceive that their line managers supported the project. Similarly for the MM2 users, once the software had been chosen, there was no perceived support for its use. Thus the work undertaken demonstrates that involvement is not something that occurs at a particular point in time nor in a particular format. In order for people to feel fully involved, they must feel that use of the software is seen as desirable by those who are important to them. Whilst this perceived support, or lack of it, becomes more evident after a software package has been introduced, the work undertaken suggests that it also influences people's engagement with the software before that introduction.

The problems faced by an organisation introducing packaged software have been addressed in the literature. The need for people to be motivated to use the software if it is to be successful and the difficulty people experience in being able to assess its impact before using it, have been discussed (e.g. Eason, 1982, Eason 1987, Mumford, 1993). This research considers the same problem from a slightly different angle - that of assisting an organisation to assess how well disposed people are towards the introduction of a package before it is brought in. It is hoped that, if some guidance may be offered cheaply and simply, an organisation will then be prepared to invest in corrective action, if necessary, at an early stage.

The User Engagement Model that has been developed in this thesis is a step on the way to determining those key factors which contribute to user engagement with a software package before it is introduced. The results of the two case studies using the newly developed ITU questionnaire further suggest that it is possible to measure the level of that engagement before the introduction of a software package. It is believed that this goes some way to developing a practical tool that will assist organisations in ensuring that they achieve the desired benefits from their investment in software packages.

8.3 Weaknesses identified in the research

• Research method

In order to scientifically test the proposed hypotheses that perceptions of ease of use, usefulness, involvement and support of referent others all contribute to user engagement with a software package it would be necessary to be able to control each of these variables. This is evidently not possible especially as the context of the work environment does not lend itself to simple manipulation. Even the user involvement, over which some control could be exercised, differed between the MM2 and PM2 workgoups. The MM user group saw an overview of only one software package whereas two were presented to the PM group. Nor were the MM users given the opportunity to discuss the way of working and whether it could be changed. It is difficult to assess how important these differences were in affecting the outcome of the two case studies.

The length of the study and the fact that it took place in a working environment increased the difficulty of maintaining a similarity between studies.

As well as being testable, scientific knowledge should also be generalisable. This poses a problem for the case study approach. Whilst the case study may "provide rich information and detailed understanding" nonetheless "they suffer from problems of generalizability" (Clegg et al.,1997 p.852). Each case study brings with it a collection of variables some of which may be unique to that particular situation and the question must be asked to what extent a generalisation may be made from the particular, if at all. As pointed out in Chapter 4, the case study method is similar to a craft. Knowledge gained from one study may then be applied in another but the knowledge gained is not generalisable. (Long & Dowell, 1989).

The use of a questionnaire not only has the benefit of providing numeric measurement but, given sufficient quantity of data, allows conclusions to be generalised. In this research, however, the questionnaires were used to provide a formal metric which was used in conjunction with interview and discussion from which to draw conclusions. The user samples were too small for the results to be generalisable and, indeed, the very nature of the data they were designed to gather, meant that they must be supported by knowledge of the context in which they were used. Use of the ITU and correlation of the results with those of the SUMI across a broad spectrum of organisations and users is now needed, although that should not be performed without some knowledge of the context of their use.

• **The position of the author in the organisation**

The position of the author as a member of the IT department in the organisation meant that account had to be taken of economic and timescale considerations during the research.

Economic constraints are evident in the MM software where the alternative package was viewed and assessed only by IT and the CM group member. Although the organisation agreed to the financial cost of ensuring that as many people as possible were present at the overviews of the proposed replacement software, this financial outlay did not extend to also letting them view an alternative. Although the people in the MM workgroup said that their preference was to consider a newer version of the same software as a first option, the fact that no other software was considered by them means that dissatisfaction due to problems inherent in the software can not be entirely ruled out (although the independent evaluation and evidence from another organisation reduce this possibility).

Timescales for investigating and implementing the replacement software had to be constantly juggled as other unplanned work arose. The size and nature of the IT department in the organisation means that the people within it have to be able to undertake a range of tasks. This means that it is never possible to dedicate time to one project and priorities have to be reassessed on a regular basis. The result was that the replacement of the PM software was a year later than planned and changes then happening meant that the research project could not be completed.

However, timescale, financial constraints and the availability of people for interviews and overviews would also apply to someone external to the organsisation. Such a person would not have the benefit of knowledge gained by virtue of being within the organisation for a number of years. People were also very co-operative in completing the questionnaires and agreeing to be interviewed. This might not have been the case with someone external to the organisation who could have been viewed with some suspicion especially by the time of the PM software replacement.

• **The ITU questionnaire**

In order to test the User Engagement construct that the ITU is intended to measure a correlation between it and the SUMI was needed. In order to complete the SUMI people need to be fairly familiar with the software. This means that there is a time gap between the two questionnaires. This time gap seems to have been sufficient for the initial impetus to help with the testing of the ITU by the external organisations to have been lost. Even though people were provided with stamped addressed envelopes in which to return the SUMI questionnaires, only 10 (less than 30%) were returned out of a possible 32 by the people who had also completed the ITU (of these 7 were usable). Testing the questionnaire by means of mailshot was deliberately discounted because the work context of the people responding would not be known and the ITU was to be completed before experience with the software had been gained. It was felt that the mailshot would provide quantity without the necessary quality. It had been hoped that two iterations of the use of the validated ITU, both correlated with the SUMI, would be possible in this research. Although these would both have been within the same organisation, it would have been with two distinct workgroups in two separate locations. In the event, only one full study was completed. Although the signs were encouraging for the second study, this can not be validated. On reflection, given the small return from the selected organisations, a mailshot might have been worth risking.

• The IM study

Some reliance was necessarily placed on the results of the SUMI. As this has been tested to industry standard and the results indeed reflected what was known about people's feelings towards the software in use, interviews and discussions were not undertaken with the users of the IM software after the SUMI results were known. Had such interviews been undertaken they would have provided more formal confirmation that the factors which contributed to people's satisfaction with the IM software were the same as those which were absent in the case of the MM and PM workgroups. Although informal discussions have been undertaken since, they were not carried out at the time.

8.4 Future Work

As Kirakowski (1995a) said, before an instrument which measures attitude may be said to be reliable it must be put through several iterations. The ITU is such an instrument. Although three different test groups were used during the course of this research, each group was comparatively small and in only one case was the number of SUMI questionnaires sufficient to measure the overall User Engagement construct with any degree of confidence. Further testing is evidently required although this would preferably be where follow-up interviews are possible in order to cross-check the results and the possibility that other factors, not measured in the ITU, have influenced the results of the SUMI.

The scoring of the ITU means that there is a very large grey area (between 40 and 60) where the score could be positive or negative. It would be preferable if this span could be reduced thus giving a more accurate indication of the level of user engagement. However, this should preferably not be at the expense of the shortness of the questionnaire as people seem to be more willing to complete something if it does not require much effort.

The results of the research suggest that the perceived support by referent others for a software package will influence an individual's perception of and engagement with the package. Work is needed on the Subjective Norm construct to see if it is possible to readily determine who is important in a particular context. As a precursor to this, maintaining the simplicity of the existing measure of the construct, it would be useful to determine whether the perception of management support is more or less important than the perception of co-worker support. The results of the discussions in the case studies suggested that it was management support which initially affected the level of engagement but that the perception of the lack of engagement on the part of co-workers served to reinforce the negative disposition.

Whilst it would certainly be useful for an organisation to be able to determine the level of user engagement before introducing a package, there then needs to be some guidance on how to improve that engagement, if needed, at the the time of initial purchase and how to maintain it throughout the life of the product. This question extends the current research beyond validation of the model and honing of the ITU. It parallels the work suggested by Morris and Dillon (1996) and that being undertaken in the CIF project (Bevan, 1999b) to ensure that software packages are usable, and turns the focus onto the organisation into which software is to be introduced. An organisation must ensure that the right climate is created for the introduction of software or even usable software will fail just as good seed will fail in badly prepared soil. Organisations need guidance in how to achieve this in a cost effective and efficient manner.

References

Ackerman M.S. & Starr B.(1996) Social Activity Indicators for Groupware *Computer, June 1996* p.37-42

Adams D.A., Nelson R.R. & Todd P.A. (1992) Perceived Usefulness, Ease of Use, and Usage of Information Technology: A Replication *MIS Quarterly, June 1992, (16:2),* p.227-247

Ajzen I. & Fishbein M., (1980) Understanding Attitudes and Predicting Social Behavior. Prentice Hall, Englewood Cliffs, NJ.

Ajzen I., (1985) "From Intentions to Actions: A Theory of Planned Behavior" in J. Kuhl and J. Beckmann (Eds) Action Control: From Cognition to Behavior, Springer Verlag, New York, p.11-39 ***as referenced in*** Taylor S.A , Todd P.A. (1995) Understanding Information Technology Usage: A Test of Competing Models Information Systems Research (1995), **6** (2) p.144-176

Alavi M., (1984) An Assessment of the Prototyping Approach to Information Systems Development. *Communications of the ACM, 27 (6)* p.556-563

Alexander H. (1987) Formally-Based Tools and Techniques for Human-Computer Dialogues, Ellis Horwood Ltd. Chichester, p.15

Anastasi A., (1976) Psychological Testing, fourth edition: Macmillan Publishing Co., Inc., New York. Collier Macmillan International Editions

Anderson W.L. & Crocca T. (1993) Engineering Practice and Codevelopment of Product Prototypes *Communications of the ACM, June 1993,36,*(4)* p.49-56

Anderson B., Smyth M., Knott R.P., Bergan M., Bergan J. & Alty J.L. (1994) Minimising Conceptual Baggage: Making Choices about Metaphor *People and Computers IX. Proceedings of HCI '94, Glasgow, 1994* G. Cockton, S.W. Draper and G.R.S. Weir (eds), Cambridge University Press, Cambridge, p.179-194

Axtell C., Clegg C. & Waterson P., (1996) Problems for User Involvement: A Human and Organizational Perspective: *People and Computers XI Proceedings of HCI '96*: M.A. Sasse, R.J. Cunningham and R.L. Winder (eds); Springer Verlag 1996, p.187-200

Baecker R.M. & Buxton W.A.S. (1987) Readings in Human Computer Interaction: A Multi-disciplinary Approach, R.M. Baecker and W.A.S.Buxton (eds), Palo Alto, CA: Morgan Kaufmann Publishers

Banville C. & Landry M. (1992) Can the Field of MIS be Disciplined? in: *Information systems research: issues, methods and practical guidelines,* R.Galliers (ed) Blackwell Scientific, Oxford, 1992 p.61-88

Barber R.E. & Lucas H.C. (1983) System Response Time, Operator Productivity, and Job Satisfaction *Communications of the ACM 26, (11),* p.972-986

Barki H. & Hartwick J.(1994) Measuring User Participation, User Involvement and User Attitude: *MIS Quarterly, March 1994,* (18:1), p.59-77

Barnard P. (1991) The Contributions of Applied Cognitive Psychology to the Study of Human Computer Interaction: *Human Factors for Informatics Usability,* B. Shackel, S. J. Richardson (eds): Cambridge University Press, Cambridge, 1991 p.151-182

Baroudi J.,Olson M., & Ives B. (1986) An Empirical Study of the Impact of User Involvement on System Usage and Information Satisfaction: *Communications of the ACM, 29, (3),* p.232-238

Baroudi J.J. & Orlikowski W.J. (1988) A Short-Form Measure of User Information Satisfaction: A Psychometric Evaluation and Notes on Use *Journal of Management Information Systems, 4,(4),* p.44-59

Bennett K. (1996) Software evolution: past, present and future *Information and Software Technology 38,(11),* p.673-680

Bevan N. (1999a) Industry Standard Usability Tests, *Tutorial 11, Interact '99, Seventh IFIP TC13 Conference on Human Computer Interaction.*

Bevan N. (1999b*) Common Industry Format Usability Tests Proceedings of UPA '98, Usability Professionals Association, Scottsdale, Arizona, 9 June-2 July, 1999. (copy contained in Tutorial 11, Interact '99, Seventh IFIP TC13 Conference on Human Computer Interaction, Interact '99, *Not page numbered).

Bevan N.& Macleod M. (1994) Usability measurement in context *Behaviour and Information Technology 13,(1-2),* p132-145

Beyer H.R. & Holtzblatt K. (1995) Apprenticing with the Customer *Communications of the ACM 38, (5),* p45-52

Bjørn-Andersen N. (1988) Are "Human Factors" Human? *The Computer Journal 31, (5)* (Oct 1988) p386-390

Böckle G., Hellwagner H., Lepold R., Sandweg G., Schallenberger B., Thudt R., Wallstab S., (1996) Structured Evaluation of Computer Systems *Computer, June 1996,* p.45-51

Caldwell Barrett S.& Uang Shaw-Tsyr (1995) Technology Usability and Utility Issues in a State Government Voice Mail Evaluation Survey *Human Factors, 1995, 37 (2),* p.306-320

Canfield Smith D., Irby C., Kimball R., Verplank B. & Harslem E., (1982) Designing the Star User Interface in: *Human-Computer Interaction* J. Preece and L. Keller (eds), Hemel Hempstead, 1990: Prentice Hall in association with The Open University, p237-259

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

Carmel E., Whitaker R.D., George J.F.(1993) PD and Joint Application Design: A Transatlantic Comparison *Communications of the ACM June 1993,36,*(4)* p.40-47

Carroll J.M., Mack R.L. & Kellogg W.A.(1988) Interface Metaphors and User Interface Design *Handbook of Human Computer Interaction,* Martin Helander (ed): Elsevier Science Publishers B.V., North Holland, Amsterdam, (1988,1991) p.67-85

Checkland P. (1981) Systems Thinking, Systems Practice. John Wiley & Sons, Chichester

Checkland P. & Scholes J. (1990) Soft Systems Methodology in Action. John Wiley & Sons, Chichester

Cheney P.H. (1984) Effects of Individual Characteristics, Organizational Factors and Task Characteristics on Computer Programmer Productivity and Job Satisfaction, *Information and Management, 7,* Elsevier Science Publishers B.V., North Holland, p.209-214

Chin J.P., Diehl V.A. & Norman K.L., (1988): Development of an Instrument Measuring User Satisfaction of the Human-Computer Interface. In: *CHI'88 Conference Proceedings: Human Factors in Computing Systems,* New York: Association for Computing Machinery, p.213-218

Chin W.W. & Todd P.A., (1995) On the Use, Usefulness, and Ease of Use of Structural Equation Modelling in MIS Research: A Note of Caution *MIS Quarterly June 1995 (19)* p.237-246

Christie B. (1985) Human Factors of the User-System Interface: *A Report on an ESPRIT Preparatory Study (1985),* B. Christie. (ed); Elsevier Science Publishers B.V, North-Holland, Amsterdam p.115-143 (specifically p.121 discusses the designer's model)

Christie B., Scane R. & Collyer J.(1995) Evaluation of human-computer interaction at the user interface to advanced IT systems in: *Evaluation of Human Work, A practical ergonomics methodology 2nd Edition, 1995.* J.R Wilson, E.N.Corlett (eds). Taylor and Francis p.310-356

Clarke A.A. (1986) A three-level human-computer interface model *International Journal of Man-Machine Studies, 24,* (6) p.503-517

Clegg C. (1988) Appropriate Technology for Humans and Organizations *Journal of Information Technology,3,(3), Sept. 1988,* p.133-146

Clegg C., Axtell C., Damodaran L., Farbey B., Hull R., Lloyd-Jones R., Nicholls J., Sell R. & Tomlinson C. (1997) Information technology: a study of performance and the role of human and organizational factors *Ergonomics, 40, (9), p.851-871*

Clement A., (1994) Computing at Work: Empowering Action by "Low-level Users" *Communications of the ACM 37, (1),* p.53-63

Clement A. & Van den Besselaar P.(1993) A Retrospective Look at PD Projects *Communications of the ACM June 1993,36,*(4)* p.29-37

Cohen Daniel I.A. (1986) Introduction to Computer Theory: John Wiley and Sons Inc. p.237-264

Dagwell R. & Weber R. (1983) System Designers' User Models: A Comparative Study and Methodological Critique *Communications of the ACM 26, (11),* p.987-997

Damodaran L. (1991) Towards a Human Factors Strategy for Information Technology Systems in: *Human Factors for Informatics Usability,* B. Shackel and S.J. Richardson (eds) Cambridge University Press, 1991, p291-324

Damodaran L. (1996) User involvement in the systems design process - a practical guide for users *Behaviour and Information Technology 1996, 15, (6),* p.363-377

Damodaran L. & Eason K.D. (1981) Design Procedures for User Involvement and User Support in: *Computing Skills and the User Interface*, M.J. Coombs and J.L.Alty (eds); London: Academic Press, p.373-388

Davis F.D. (1989) Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology *MIS Quarterly, September 1989,(13:4)* p319-339

Davis F.D. (1993) User acceptance of information technology: system characteristics, user perceptions and behavioral impacts in: *International Journal of Man Machine Studies,38, (3)* p.475-487

Davis F.D., Bagozzi R.P. & Warshaw P.R.(1989) User Acceptance of Computer Technology: A Comparison of Two Theoretical Models *Management Science 35,(8),* p.982-1003

Davis F.D. & Venkatesh V. (1996) A critical assessment of potential measurement biases in the technology acceptance model: three experiments *International Journal of Human-Computer Studies (1996) 45,(1),* p.19-45

Desurvire H.W., Kondziela J.M. & Atwood M.E.,(1992) What is Gained and Lost when using Evaluation methods other than Empirical Testing in: *People and Computers VII Proceedings of the HCI '92 Conference*, A. Monk, D. Diaper and M.D. Harrison(eds):Cambridge University Press, Cambridge (on behalf of the BCS) p.89-102

Diaper D. (1989) Task Analysis for Knowledge Descriptions (TAKD): the method and an example: in *Task Analysis for Human-Computer Interaction,* D. Diaper: Ellis Horwood, 1989, Chichester, p.108-159

Dillon A. & Watson C. (1996) User analysis in HCI - the historical lessons from individual differences research *International Journal of Human-Computer Studies (1996) 45,(6),* p.619-637

Dix A., Finlay J., Abowd G. & Beale R. (1993) Human-Computer Interaction: Prentice Hall International (UK) Ltd., Hemel Hempstead

Doherty N.F. & King M. (1998) The consideration of organisational issues during the systems development process: an empirical analysis *Behaviour and Information Technology,17,(1),* p.41-51

Doll W.J. & Torkzadeh G. (1989) A Discrepancy Model of End-User Computing Involvement *Management Science, 35, (10),* p.1151-1171

Dourish P, Adler A., Bellotti V.& Henderson A.(1996) Your Place or Mine? Learning from Long-Term Use of Video Communication, *Rank Xerox EuroPARC, Cambridge UK: from dourish@europarc.xerox.com* (obtained 1996) p.1-11

Downton A. (1993) Engineering the Human-Computer Interface (student edition): McGraw-Hill Book Company, 1993, p.46

Dzida W (1995) Standards for user-interfaces *Computer Standards and Interfaces,17, (1995)*: Elsevier Science B.V, p.89-97

Eason K.D. (1982) The process of introducing information technology *Behaviour and Information Technology, 1, (2), p.197-213*

Eason K.D. (1987) Methods of planning the electronic workplace *Behaviour and Information Technology,6, (3), p.329-238*

Eason K.D. (1988) Information Technology and Organisational Change: Taylor & Francis Ltd., London (1988, 1990)

Eason K. (1989) Designing Systems to Match Organisational Reality in: *People and Computers V: Proceedings of the Fifth Conference of the British Computer Society Human-Computer Interaction Specialist Group,* A Sutcliffe and L. Macaulay (eds): Cambridge University Press, Cambridge, p.57-69

Eason K.D. & Damodaran L. (1981) The Needs of the Commercial User in: *Computing Skills and the User Interface*: M.J. Coombs and J.L. Alty (eds), 1981: Academic Press, London, p.115-139

Eason K.D. & Sell R.G. (1981) Case Studies in Job Design for Information Processing Tasks in: *Stress, Work Design, and Productivity,* E.N. Corlett and J. Richardson (eds), 1981: John Wiley and Sons Ltd., p.195-208

Eason K.D. & Harker S. (1988) The Supplier's Role in the Design of Products for Organisations *The Computer Journal **31,5** (Oct. 1988)* p.426-430

Eason K. & Harker S. (1991) Human Factors Contributions to the Design Process *Human Factors for Informatics Usability:* B. Shackel, S. Richardson (eds), Cambridge University Press, Cambridge, 1991 p.73-96.

Ehn P., (1993) Scandinavian Design: On Participation and Skill in: *Participatory Design: Principles and Practice,* D. Schuler & A. Namioka (eds), Lawrence Erlbaum Associates, NJ., 1993, p.41-77

Farhoomand A.F. (1992) Scientific Progress of Management Information Systems in: *Information systems research: issues, methods and practical guidelines,* R..Galliers(ed) Blackwell Scientific, Oxford, 1992, p.93-111

Fishbein M. & Ajzen I., (1975) Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research. Addison-Wesley, Reading, MA

Fitzpatrick R. & Higgins C. (1998) Usable Software and Its Attributes: A Synthesis of Software Quality, European Community Law and Human-Computer Interaction in: *People and Computers XIII HCI'98,* H. Johnson, L. Nigay & C. Roast (eds): Springer-Verlag London Ltd., p.3-21

Franz C.R. & Robey D. (1986) Organizational Context, User Involvement, and the Usefulness of Information Systems *Decision Sciences, Summer 1986, 17 (3),* p.329-356

Furner S.M., Watkinson N. & Clark W. (1990). An ISDN multipoint teleconferencing environment (1990) presented at the IEE Colloquium on "CSCW: Computer Supported Co-Operative Work", Savoy Place, London, 24/10/90 (organised by Professional Group C5, Man-Machine Interaction) Digest No:1990/132

Gale A. & Christie B. (1987) Psychophysiology and the Electronic Workplace: The Future in: *Psychophysiology and the Electronic Workplace,* A. Gale and B. Christie (eds). Chichester/New York: John Wiley and Sons Ltd. p.315-333

Galliers R. (1992a) Information systems research: issues, methods and practical guidelines, Blackwell Scientific, Oxford, 1992

Galliers R. (1992b) Choosing Information Systems Research Approaches in: *Information systems research: issues, methods and practical guidelines,* R.Galliers (ed), Blackwell Scientific, Oxford, 1992, p.144-162

Gentner D.R. & Grudin J. (1996) Design Models for Computer-Human Interfaces *Computer June 1996* p.28-35

Gentner D., & Nielson J. (1996) The Anti-Mac Interface *Communications of the ACM 39,(8), (Aug.96)* p.70-82

Goodhue D.L. (1995) Understanding User Evaluations of Information Systems *Management Science 41, (12),* p.1827-1844

Gray W.D., John E.B. & Atwood M.E. (1993) Project Ernestine: Validating a GOMS Analysis for Predicting and Explaining Real-World Task Performance *Human-Computer Interaction, 8,(3),* p237-309

Greif S., (1991) Organisational Issues and Task Analysis *Human Factors for Informatics Usability,* B. Shackel, S. Richardson(eds): Cambridge University Press, Cambridge, 1991 p.247-266

Grønbæk K., Grudin J., Bødker S. & Bannon L. (1993) Achieving Cooperative System Design: Shifting from a Product to a Process Focus in: *Participatory Design: Principles and Practice,* D. Schuler & A. Namioka (eds), Lawrence Erlbaum Associates Inc., NJ, p.79-97

Grønbæk K., Kyng M., Mogensen P. (1996) CSCW Challenges: Co-operative Design in Engineering Projects *Communications of the ACM, June 1993, 36, (4)* p.67-77

Grudin, J. (1989) The Case Against User Interface Consistency *Communications of the ACM, 32,(10)* p.1164-1173

Grudin J. (1992) Utility and Usability: research issues and development contexts *Interacting with Computers 4 (2), August 1992* p.209-217

Grudin J., (1993) Obstacles to Participatory Design in Large Product Development Organizations in: *Participatory Design: Principles and Practice,* D. Schuler & A. Namioka (eds), Lawrence Erlbaum Associates Inc., NJ, p.99-119

Harper B. D. & Norman K.L. (1993) Improving User Satisfaction: The Questionnaire for User Interaction Satisfaction version 5.5 (supplied by University of Maryland, College Park with the QUIS questionnaire).

Hartwick J. & Barki H.(1994) Explaining the Role of User Participation in Information System Use *Management Science 40, (4) April 1994* p.440-465

Hather R.M., Burd E.L. & Boldyreff C.(1996) A method for application management maturity assessment *Information and Software Technology 38 (1996):* Elsevier Science B.V., p.701-709

Hawk S.R (1993). The effects of user involvement: some personality determinants. *International Journal of Man-Machine Studies,38,(5),* p.839-855

Heinbokel T. Sonnentag S., Frese M., Stolte W. & Brodbeck F.C. (1996) "Don't underestimate the problems of user centredness in software development projects - there are many!" *Behaviour and Information Technology, 15,(4),* p.226-236

Hendrickson A.R., Massey P.D. & Cronan T.P. (1993) On the Test-Retest Reliability of Perceived Usefulness and Perceived Ease of Use Scales *MIS Quarterly (17:2),* p.227-230

Hirschheim R.A. (1992) Information Systems Epistemology: An Historical Perspective in: *Information systems research: issues, methods and practical guidelines,* R.Galliers(ed): Blackwell Scientific, Oxford, 1992, p.28-60

Hirschheim R. & Newman M. (1988) Information Systems and User Resistance: Theory and Practice: *The Computer Journal 31,(5)* (Oct 1988) The WARWICK study p.398-408

Holtzblatt K. & Beyer H. (1993) Making Customer-Centered Design Work for Teams *Communications of the ACM,36,(10),* p.93-103

Hornby P. & Clegg C. (1992) User Participation in context: a case study in a UK bank *Behaviour and Information Technology, 1992,11,(5)* p.293-307

Hubona G.S. & Blanton J.E. (1996) Evaluating system design features *International Journal of Human-Computer Studies, 44,(1),* p.93-118

Hutchings A.E. & Knox S.T. (1995) Creating Products Customers Demand *Communications of the ACM, 38, (5),* p.72-80

Igbaria M. & Nachman S.A. (1990) Correlates of user satisfaction with end user computing. An exploratory study: *Information & Management, 19,* Elsevier Science Publishers B.V., North Holland, p.73-82

Ives B. & Olson M.H. (1984) User Involvement and MIS Success: A Review of Research *Management Science, 30, (5),* p.586-603

Ives B., Olson M.H. & Baroudi J.J. (1983) The Measurement of User Information Satisfaction *Communications of the ACM, 26, (10),* p.785-793

James E.B. (1981) The User Interface: How we may Compute in:*Computing Skills and the User Interface*: M.J. Coombs and J.L. Alty (eds) (1981); Academic Press, London p.337-371

Johnson-Laird P.N. (1983) Mental Models, Cambridge University press. p29-34

Kappelman L.A. & McLean E.R., (1991) The Respective Roles of user Participation and User Involvement in Information System Success *Proceedings, Twelfth International Conference on Information Systems, New York, NY,* p.339-349

Kaptelinin V. (1996) Computer-Mediated Activity: Functional Organs in Social and Developmental Contexts in: *Context and consciousness: Activity Theory and Human-Computer Interaction,* B.A. Nardi (ed) Cambridge, MA, The MIT Press p.45-68

Keen Peter G.W.(1981) Information Systems and Organizational Change *Communications of the ACM, 24, (1),* p.24-33

Keil M. & Carmel E. (1995) Customer Developer Links in Software Development *Communications of the ACM 38, (5),* p.33-44

Kensing F. & Munk-Madsen (1993) PD: Structure in the Toolbox *Communications of the ACM 36, (4),* p.78-85

Kensing, Simonsen & Bødker (1998) Must Method For Participatory Design *International Journal of Human-Computer Interaction (1998) 13,(2),* p.167-198

Kieras D.E. & Bovair S. (1984) The Role of a Mental Model in Learning to Operate a Device in: *Human-Computer Interaction: selected readings*, J. Preece and L. Keller (eds),1990, Hemel Hempstead: Prentice Hall in association with The Open University, p205-221

Kiger J.I. (1984) The depth/breadth trade-off in the design of menu-driven user interfaces. *International Journal of Man-Machine Studies,20,(2),* p.201-213

Kilgour A. (1987) Theory and Practice in User Interface Management Systems in:*Human-Computer Interaction:selected readings,* J. Preece and L. Keller (eds) (1990), Hemel Hempstead: Prentice Hall in association with The Open University, p.421-428

King W.R. & Lee T-H. (1991) The Effects of User Participation on System Success: Toward a Contingency Theory of User Satisfaction *Proceedings, Twelfth International Conference on Information Systems, New York, NY,* p.327-338

Kirakowski J. (1995a) The Software Usability Measurement Inventory: Background and Usage (supplied by author in 1996. Due to appear in P. Jordan, B.Thomas and B. Weerdmeester: Usability Evaluation in Industry. Taylor and Francis)

Kirakowski J.(1995b) The Use of Questionnaire methods for Usability Assessment (supplied by author in 1996. Due to appear in a collection of papers from the MUSiC project)

Kjær A. & Madsen H. K. (1995) Participatory Analysis of Flexibility *Communications of the ACM, 38, (5),* p.53-60

Klein H.K. & Hirschheim R.(1983) *Issues and Approaches to Appraising Technological Change in the Office: A Consequentialist Perspective (excerpt)* in: Readings in Human Computer Interaction: a Multidisciplinary Approach, R.M. Baecker, W.A.S Buxton (eds), 1987, Palo Alto, CA: Morgan Kaufmann Publishers, p.83-92

Kling R. & Iacono S. (1984) The Control of Information Systems Developments After Implementation *Communications of the ACM,27, (12),* p.1218-1226

Kuhn S. & Muller M.J. (1993) Participatory Design *Communications of the ACM, June 1993 36,*(4),* p.24-28

Kuutti K. (1996) Activity Theory as a Potential Framework for Human-Computer Interaction Research in *Context and consciousness: Activity Theory and Human-Computer Interaction,* B.A. Nardi (ed) Cambridge, MA, The MIT Press p.17-44

Kyng M. (1991) Designing for Co-operation: Co-operating in Design *Communications of the ACM,34, (12),* p.65-72

Land F. (1992) The Information Systems Domain in: *Information systems research: issues, methods and practical guidelines,* R..Galliers (ed), Blackwell Scientific, Oxford, 1992, p.6-13

Le Quesne P.N. (1988) Individual and Organisational Factors and the Design of IPSEs *The Computer Journal 31, (5)* (Oct 1988) p.391-397

Lim K.H., Benbasat I. & Todd P.A., (1996) An Experimental Investigation of the Interactive Effects of Interface Style, Instructions, and Task Familiarity on User Performance *ACM Transactions on Computer-Human Interaction, 3,(1), March 1996* p.1-37

Long J. & Dowell J., (1989) Conceptions of the Discipline of HCI: Craft, Applied Science, and Engineering in: *People and Computers V: Proceedings of the Fifth Conference of the British Computer Society Human-Computer Interaction Specialist Group,* A. Sutcliffe and L.Macaulay (eds), Cambridge University Press, Cambridge p.9-32

Mackay W.E. & Fayard A-L. (1997) HCI, Natural Science and Design: A Framework for Triangulation Across Disciplines in: *Proceedings of the conference on Designing interactive systems: processes, practices, methods and techniques, 1997* p. 223-234

Maddix F. (1990) Human-Computer Interaction theory and practice: Ellis Horwood Ltd., London

Manketlow K. & Jones J., (1987) Principles from the psychology of thinking and mental models in: Applying Cognitive Psychology to User-Interface Design, M.M.Gardiner & B. Christie (eds), John Wiley and Sons, Chichester p.83-117

Mantei M.M. & Teorey T.J. (1988) Cost Benefit Analysis for Incorporating Human Factors in the Software Life Cycle. *Communications of the ACM* ***31**, (4)* p.428-439

Markus M. Lynne (1983) Power, Politics and MIS Implementation *Communications of the ACM, 26, (6),* p.430-444

Martinko M.J., Henry J.W. & Zmud R.W. (1996) An attributional explanation of individual resistance to the introduction of information technologies in the workplace *Behaviour and Information Technology, 1996, 15, (5),* p.313-330

Mathieson K. (1991) Predicting User Intentions: Comparing the Technology Acceptance Model with the Theory of Planned Behaviour *Information Systems Research (1991), 2,(3)* p.173-191

Medcof J.W. (1989) The Effect of Extent of Use of Information Technology and Job of the User upon Task Characteristics *Human Relations, 42, (1),* p.23-41

Molich R. & Nielsen J. (1990) Improving a human-computer dialogue *Communications of the ACM, 33, (3)* p.338-348

Moran T.P (1981) The Command Language Grammar: a representation for the user interface of interactive computer systems. *International Journal of Man-Machine Studies, 15, (1)* p.3-50

Morris M.G. & Dillon A.P. (1996) The importance of usability in the establishment of organizational software standards for end user computing *International Journal of Human-Computer Studies, 45, (2)* p.243-258

Mosier J.N. & Smith S.L. (1986) Application of guidelines for designing user interface software. *Behaviour and Information Technology,* 5, (1), p.39-46

Muller M.J., Wildman D.M. & White E.A. (1993) Taxonomy of PD practices: a brief practitioners guide. *Communications of the ACM,(June 1993), 36, *(4), p.26-27*

Mumford E. (1991) Participation in Systems Design - What Can It Offer? *Human Factors for Informatics Usability:* B. Shackel and S. Richardson (eds), Cambridge University Press, Cambridge, 1991, p.267-290

Mumford E., (1993) The Participation of Users in Systems Design: An Account of the Origin, Evolution, and Use of the ETHICS Method in: *Participatory Design Principles and Practices*, D. Schuler and A. Namioka (eds); Lawrence Erlbaum Associates Inc., NJ, p.257-270

Mumford E., (1993) The ETHICS Approach *Communications of the ACM, June 1993,36, *(4),* p.82

Murch G.M. (1987) Colour Graphics - Blessing or Ballyhoo in: *Readings in Human Computer Interaction: A Multi-disciplinary Approach,* R.M. Baecker and W.A.S.Buxton (eds), Palo Alto, CA: Morgan Kaufmann, p.333-341

Nardi B.A. (1996a) Activity Theory and Human-Computer Interaction in: Context and consciousness: Activity Theory and Human-Computer Interaction, B.A. Nardi (ed), Cambridge MA; The MIT Press, p.7-16

Nardi B.A. (1996b) Context and consciousness: Activity Theory and Human-Computer Interaction, Cambridge MA; The MIT Press

Nielsen J. (1986), A virtual protocol model for computer-human interaction. *International Journal of Man-Machine Studies, 24,(3)* p.301-312

Nielsen J.& Levy J. (1994) Measuring Usability, Preference vs. Performance *Communications of the ACM, 37, (4),* p.66-75

Nurminen Markku I. (1988) People or Computers: Three Ways of Looking at Information Systems: Student litteratur, Lund, Sweden, Chartwell-Bratt, Kent

O'Donnell P.J., Scobie G. & Baxter I. (1991) The Use of Focus Groups as an Evaluation Technique in HCI: in *People and Computers VI: Proceedings of the HCI '91 Conference* D. Diaper and N. Hammond (eds), Cambridge University Press, Cambridge, p.211-224

Ogden W.C. (1988) Using Natural Language Interfaces *Handbook of Human Computer Interaction,* Martin Helander (ed): Elsevier Science Publishers B.V., North Holland, Amsterdam, (1988,1991) p.281-299

Olaniran B.A. (1996) A model of group satisfaction in computer-mediated communication and face-to-face meetings *Behaviour and Information Technology, 1996,15, (1),* p.24-36

Olson J.S., Card S.K., Landauer T.K., Olson G.M., Malone T. & Leggett J. (1993) Computer-supported co-operative work: research issues for the 90s *Behaviour and Information Technology, 1993, 12, (2)* p.115-129

O'Malley C., Langton S., Anderson A., Doherty-Sneddon G. & Bruce V. (1996) Comparison of face-to-face and video-mediated interaction *Interacting with Computers, 8 (2)*, p.177-192

O'Neill E.J. (1998) User-developer cooperation in software development: building common ground and usable systems. Unpublished PhD thesis. Queen Mary and Westfield College, University of London.

Parkinson S.R., Hill M.D., Sisson N. & Viera C. (1988) Effects of breadth, depth and number of responses on computer menu search performance: *International Journal of Man-Machine Studies,(1988) 28,* (6), p.683-692

Payne S.J. & Greene T.R.G. (1989), Task Action Grammar: the model and its developments: in *Task Analysis for Human-Computer Interaction,* D. Diaper (ed), Chichester, 1989: Ellis Horwood Ltd., p75-107

Pervan G.P. & Klass D.J. (1992) The Use and Misuse of Statistical Methods in Information Systems Research in: *Information systems research: issues, methods and practical guidelines,* R.Galliers (ed), Blackwell Scientific, Oxford, 1992, p.209-227

Poltrock S.E. & Grudin J.(1994) Organisational Obstacles to Interface Design and Development: Two Participant - Observer Studies: *ACM Transactions on Computer-Human Interaction,1, (1), (1994)* p.52-80

Preece J., Rogers Y., Sharp H., Benyon D., Holland S. & Carey T. (1994) Human-Computer Interaction: Addison-Wesley,Wokingham

QUIS The Questionnaire for User Interaction Satisfaction, developed by Kent Norman and Ben Shneiderman from 1987 at University of Maryland at College Park.

Ramamurthy K., King W.R., Premkumar G. (1992) User characteristics - DSS effectiveness linkage: an empirical assessment *International Journal of Man-Machine Studies (1992), 36, (3)* p.469-505

Rasmussen J. (1992) the Ecology of Work and Interface Design: *in People and Computers VII: Proceedings of the HCI '92 Conference*, A. Monk, D. Diaper and M.D.Harrison (eds), Cambridge University Press, Cambridge, p.3-18

Ravden S. & Johnson G.(1989) Evaluating Usability of Human-Computer Interfaces: A Practical Method. Chichester: Ellis Horwood

Reiterer H. (1992) EVADIS II: A New Method to Evaluate User Interfaces - *People and Computers VII. Proceedings of the HCI '92 Conference,* A. Monk, D. Diaper and M.D.Harrison (eds), Cambridge University Press, Cambridge, p.103-115

Reiterer H. & Oppermann R.(1993) Evaluation of user interfaces: EVADIS II - a comprehensive evaluation approach. *Behaviour and Information Technology 12,(3),* p137-148

Rice R.E. & Tyler J.(1995) Individual and organizational influences on voice mail use and evaluation *Behaviour and Information Technology, 1995, 14, (6),* p.329-341

Rivard S. & Huff S.L. (1988) Factors of Success for End-User Computing *Communications of the ACM, May 1988, 31, (5),* p.552-561

Roast C. (1994) Modelling Interaction using Template Abstractions *People and Computers IX. Proceedings of HCI '94, Glasgow, 1994 G. Cockton,* S.W. Draper and G.R.S. Weir (eds), Cambridge University Press, Cambridge, p.273-284

Rubin, T. (1988) So you want to be a professional? in: *User Interface Design for Computer Systems,* T. Rubin (ed.), Chichester: Ellis Horwood Ltd. p 168-177

Rushinek A. & Rushinek S.F.(1986) What makes users happy? *Communications of the ACM,29,(7),* p.594-598

Sanderson P.M. & Fisher C. (1994) Exploratory Sequential Data Analysis: Foundations *Human-Computer Interaction, 9,* p.251-317

Saunders C.S. & Scamell R.W. (1986) Organisational Power and the Information Services Department: A Reexamination *Communications of the ACM, 29, (2),* p.142-147

Schneider M.L. (1982) Ergonomic Considerations in the Design of Command Languages *Human Factors and Interactive Computer Systems: Proceedings of the NYU symposium on User Interfaces, New York, May 26-28, 1982.* Yannes Vassilou (ed): Ablex Publishing Corporation, Norwood, New Jersey, p.141-161.

Schuler D. & Namioka A. (1993) *Preface to*: Participatory Design. Principles and Practices, D. Schuler & A. Namioka (eds), Lawrence Erlbaum Associates Inc., NJ, p.ix

Selker T. (1996) New Paradigms for Using Computers *Communications of the ACM,39,(8),* p.60-69

Shackel B. (1990) The Business Case for Ergonomics/Human Factors in Information Technology: Tutorial paper presented at *INTERACT '90*

Shackel B. (1991) Usability - Context, Framework, Definition, Design and Evaluation *Human Factors for Informatics Usability,* B. Shackel and S. Richardson (eds)*: Cambridge University Press, Cambridge, 1991* p.21-37

Shapiro D. (1995) Ethnography and HCI *Interfaces,* **27,** (Winter 1995) p.8-11

Shneiderman B. (1982) The Future of Interactive Systems and the Emergence of Direct Manipulation *Human Factors and Interactive Computer Systems. Proceedings of the NYU symposium on User Interfaces, New York, May 26-28, 1982,* Yannes Vassilou (ed) : Ablex Publishing Corporation, Norwood, New Jersey, p.1-27

Shneiderman B., (1988) We can design better user interfaces: A review of human-computer interaction styles IEA Plenary Paper, *Ergonomics, 31, (5),* p.699-710

Shneiderman B.(1991) A Taxonomy and Rule Base for the Selection of Interaction Styles. *Human Factors for Informatics Usability,* B. Shackel and S. Richardson (eds) *:* Cambridge University Press, Cambridge, 1991, p.325-342

Shneiderman B. (1992) *Designing the User Interface: Strategies for Effective Human-Computer Interaction,* 2nd edn., Reading, MA:Addison-Wesley.

Staggers N. & Norcio A.F.(1993) Mental Models: concepts for human-computer interaction research *International Journal of Man-Machine Studies, 38,(4),* p.587-605

Stammers R.B. & Shepherd A. (1995) Task Analysis: *in Evaluation of Human Work, A practical ergonomics methodology (2nd Edition, 1995),* J.R. Wilson and E.N. Corlett (eds): Taylor and Francis Ltd., p144-168

SUMI Software Usability Measurement Inventory (1993) developed by Porteous M., Kirakowski J. & Corbett M., of the Human Factors Research Group, University College Cork, Ireland.

Sutcliffe A.G. & Springett M.V.(1992) From User's Problems to Design Errors: Linking Evaluation to Improving Design Practice in: *People and Computers VII: Proceedings of the HCI '92 Conference,* A. Monk , D. Diaper and M.D. Harrison (eds), Cambridge University Press, Cambridge, p.117-134

Tan C.L., Quah T.S. & Teh H.H. (1996) An Artificial Neural Network that Models Human Decision Making *Computer, March 1996,* p.64-70

Taylor M.M. (1988) Layered protocols for computer-human dialogue. 1:Principles *International Journal of Man-Machine Studies (1988), **28,*** Special Issue: Multimodal Computer-Human Interaction (no.2&3), p.175-218

Taylor S.A & Todd P.A. (1995) Understanding Information Technology Usage: A Test of Competing Models *Information Systems Research (1995),6, (2)* p.144-176

Tomlin R. (1991) Developing a management culture in which information technology will floursih: how the UK can benefit *Journal of Information Technology, 6,* p.45-55

Turoff M.(1982) Interface Design in Computerized Conferencing Systems: A Personal View. *Human Factors and Interactive Computer Systems. Proceedings of the NYU symposium on User Interfaces, New York, May 26-28, 1982* Yannes Vassilou (ed) : Ablex Publishing Corporation, Norwood, New Jersey, p.243-259

Wagner I. (1993) A Web of Fuzzy Problems: Confronting the Ethical Issues *Communications of the ACM, June 1993,36, *(4),* p.94-101

Wallace M.D. & Anderson T.J.(1993) Approaches to interface design *Interacting with Computers 5, (3),* p.259-278

Weaver, P.L. (1993) Practical SSADM v.4. A Complete Tutorial Guide, Pitman publishing, London

Webby R., & O'Connor M. (1994) The effectiveness of Decision Support Systems: the implications of task complexity and DSS sophistication *Journal of Information Technology, 9,* p.19-28

Wegner E. (1995) Quality of software packages: the forthcoming international standard *Computer Standards and Interfaces, 17,* : Elsevier Science Publishers B.V., North Holland, Amsterdam, p.115-120

Welzel D. & Hausen H-L (1995) A method for software evaluation. Contribution to the European project SCOPE to international standards, *Computer Standards and Interfaces, 17,*: Elsevier Science Publishers B.V., North Holland, Amsterdam, p.121-129

Whitefield A. & Hill B. (1994) Comparative analysis of task analysis products. *Interacting with Computers 6, (3),* p.289-309

Wiedenbeck S. & Davis S. (1997) The influence of interaction style and experience on user perceptions of software packages *International Journal of Human-Computer Studies,46, (5),* p.563-588

Willer D. & Willer J. (1973) Systematic Empiricism: critique of a pseudoscience: Prentice Hall Inc.,Englewood Cliffs, NJ.

Williams M.G. & Begg V.(1993) Translation between Software Design and Users *Communications of the ACM, June 1993,36,*(4),* p.102-103

Wong E.Y.W. & Tate G. (1994) A study of user participation in information systems development *Journal of Information Technology, 9,* p.51-60

Zmud R.W. (1979) Individual Differences and MIS Success: A Review of the Empirical Literature *Management Science, 25, (10) October 1979,* p.966-979

* The June 1993 issue of Communications of the ACM has the printed number 4 on the copy but is actually No. 6 in the volume as the month would imply.

Appendix 1

The SUMI questionnaire

The Software Usability Measurement Inventory (SUMI) was developed in 1993 by the Human Factors Research Group of Cork University. The questionnaire may be obtained from the University in Ireland.

There are 50 questions in the full questionnaire, which is printed in booklet form with instructions on the front page as to how it should be completed. It has been used in this research to assess the level of satisfaction with software in use. The Global scores from the SUMI were correlated with the total score of the ITU in order to validate the User Engagement construct of the latter.

The second page of the SUMI (the first page of the questions) has been reproduced on the next page.

An example of the note which accompanied the initial issue of the SUMI and the follow up note with the results follow on the next two pages. The wording of these notes varied slightly according to the groups to which they were being sent. The name of the application has been blocked out in order to maintain confidentiality. People were also provided with a printed explanation of the meaning of the scales similar to that given in the body of this thesis.

		Agree	Undecided	Disagree
1	This software responds too slowly to inputs.	☐	☐	☐
2	I would recommend this software to my colleagues.	☐	☐	☐
3	The instructions and prompts are helpful.	☐	☐	☐
4	The software has at some time stopped unexpectedly.	☐	☐	☐
5	Learning to operate this software initially is full of problems.	☐	☐	☐
6	I sometimes don't know what to do next with this software.	☐	☐	☐
7	I enjoy my sessions with this software.	☐	☐	☐
8	I find that the help information given by this software is not very useful.	☐	☐	☐
9	If this software stops it is not easy to restart it.	☐	☐	☐
10	It takes too long to learn the software commands.	☐	☐	☐
11	I sometimes wonder if I am using the right command.	☐	☐	☐
12	Working with this software is satisfying.	☐	☐	☐
13	The way that system information is presented is clear and understandable.	☐	☐	☐
14	I feel safer if I use only a few familiar commands or operations.	☐	☐	☐
15	The software documentation is very informative.	☐	☐	☐
16	This software seems to disrupt the way I normally like to arrange my work.	☐	☐	☐
17	Working with this software is mentally stimulating.	☐	☐	☐
18	There is never enough information on the screen when it's needed.	☐	☐	☐
19	I feel in command of this software when I am using it.	☐	☐	☐
20	I prefer to stick to the facilities that I know best.	☐	☐	☐

SUMI questionnaire

This is a questionnaire which measures how you feel about a piece of software. It was developed by the Human Factors Research Group at Cork University and has been used in several large companies throughout Europe to assess people's perception of the software they are using.

It is being used here as part of an exercise to try to determine what factors influence how people feel about the software they use.

Please read each statement and indicate by putting a check in the left side box if you agree with the statement or think it is generally similar to what you think, or put a check in the middle box if you are undecided or you find yourself partly agreeing and partly disagreeing with the statement, or put a check in the right side box if you disagree with the statement. Some of the questions may seem a little strange but please do answer all of them.

The questionnaires will be treated as confidential and it is not necessary to put your name. The name of the software has been filled in already.

It is estimated that it takes between 10 and 20 minutes to complete the questionnaire. The developers advise that it be completed in the work setting during, or soon after, use of the software and (probably most difficult of all!) without interruption.

Your co-operation in the exercise is greatly appreciated.

Many thanks.

Thanks very much for completing the SUMI questionnaire on xxx and apologies for the long time gap. If you remember, the aim of the questionnaire was to get some information on how people generally feel about a product and to see if there are areas which are perceived as particularly good or bad. The results are not intended as a reflection on the software itself but on how people feel about using it. There are actually 36 pages of stats produced for study. For information, the table below gives the global results. As you can see xxx has not come out particularly well in any of the 5 headings. The standard median would be 50. A summary of the meaning of each heading is given on the other sheet.

I'm still working through the report but am already finding that it raises some further questions. This may mean that I need to come back to some of you - I hope this is o.k.

As you are probably aware, we will be looking at the whole xxx question in more detail in October this year.

Profile Analysis

Scale	UF	Ucl	Medn	Lcl	LF
Global	64	39	33	27	-2
Efficiency	56	34	28	22	2
Affect	81	47	41	35	0
Helpfulness	71	47	41	35	10
Control	52	33	29	25	7
Learnability	87	43	35	27	-12

Note:

The Median is the middle score when the scores are arranged in numerical order. It is the indicative sample statistic for each usability scale.

The Ucl and Lcl are the Upper and Lower Confidence Limits. They represent the limits within which the theoretical true score lies 95% of the time for this sample of users.

The UF and LF are the Upper and Lower Fences. They represent values beyond which it may be plausibly suspected that a user is not responding with the rest of the group: the user may be responding with an outlier.

SUMI Scoring Report from SUMISCO 7.380
Time and date of analysis: 15:45:18 on 12-15-1996

Files used in this analysis:
SUMI English (UK) Language Items
SUMI Version 2.1 Scoring Keys
distributions from January 1996 standardisation
weights from January 1996 standardisation
population parameters from January 1996 standardisation

Appendix 2

Example of results from the Item Consensual Analysis

The Item Consensual Analysis (ICA) shows the items from the SUMI questionnaire for which the results differ from the standardised database. The January 1996 standardisation was used throughout this research.

In order to check the author's interpretation of the ICA results from the SUMI an exercise was carried out with two versions of a wordprocessing package and the data from 28 people was sent to the HFRG in Ireland for analysis.

The following pages are an example of the individual item analysis from the MM1 software. This exercise at the individual item level was not carried out for the MM2 software as the interest was mainly on the Global results at that stage.

Item 44 *It is relatively easy to move from one part of a task to another.*

	Agree	Undecided	Disagree	
Profile	1	0	14	
Expected	10.71	2.28	2.01	
Chi Sq	8.8	2.28	71.35	82.43***

It is evident from the scorings here that users do not find it at all easy to move from one area of the software to another. The one user who checked "agree" to this item (user 13) had a heavy involvement in keyboard input on the previous system and has very good typing skills. However, others who checked the "disagree" box also have very good typing skills and the same background experience. Perhaps the response indicates a different method of working.

Item 1 *This software responds too slowly to inputs.*

	Agree	Undecided	Disagree	
Profile	11	3	1	
Expected	2.89	2.19	9.92	
Chi Sq	22.77	0.3	8.02	31.09***

The number of people who have ticked the "agree" box here is interesting. The only user to have ticked "disagree" is user 14. The software is DOS based running over Local Area Networks and the response is actually faster than would be achieved on a standard 486 PC running a Windows application. However, there have been requests for hardware upgrades from these locations (the hardware has not been changed in three years, whereas Pentium PCs were being installed in other areas at the time of the questionnaires) and it is possible that the perception of being "left behind" is what is really being reflected here particularly as the software in question is what is run on the PCs more than 80% of the time.

Item 47 *This software is really very awkward.*

	Agree	Undecided	Disagree	
Profile	5	6	4	
Expected	0.96	2.71	11.32	
Chi Sq	16.92	3.98	4.74	25.64***

What is interesting here is the number of people who are undecided. One would expect there to be either a strong agreement or a strong disagreement and yet the scores from each of these boxes are is lower than the undecided. Perhaps people were not really sure what was meant by the question.

Item 21 *I think this software is inconsistent.*

	Agree	Undecided	Disagree	
Profile	8	5	2	
Expected	1.94	3.82	9.24	
Chi Sq	18.92	0.36	5.67	24.95***

More than half the users agree with this, although a third are undecided. One of the users who disagrees is user 14 and the other user is the person who checked "agree" to item 44 ("It is relatively easy to move from one part os a task to another"). Evidently, this user finds it easier to navigate around the software than the others.

Item 49 *Getting data files in and out of the system is not easy.*

	Agree	Undecided	Disagree	
Profile	8	7	0	
Expected	2.12	5.63	7.25	
Chi Sq	16.36	0.33	7.25	23.95***

Not all users are involved in importing in exporting data files which perhaps explains the high number of undecideds. User 14 is involved in this. In fact everyone who has had to do it, has had some difficulty in getting data files in or out. There are two areas where this may be done. In one, it is a menu selection and there is no other action required until a message appears to the effect that the transfer has completed. The other is in the report writer which allows the export of Lotus files. This data is used as part of the reporting by local to central management. It would appear that this is the main problem area.

Item 34 *The software allows the user to be economic of keystrokes.*

	Agree	Undecided	Disagree	
Profile	1	7	7	
Expected	9.54	3.46	2.0	
Chi Sq	7.65	3.63	12.51	23.79***

Again it is the user who seems to have little difficulty with navigation (item 44) who has ticked "agree" to this. The split between the undecideds and disagree is equal on the first report. Given the number of people who felt that it is difficult to move from one task to another (Item 44), it is perhaps surprising that the number who disagree with this statement is not higher. Perhaps the answer to Item 44 reflects rather more than just dissatisfaction with the mode of navigation.

Item 42 *The software has a very attractive presentation.*

	Agree	Undecided	Disagree	
Profile	2	5	8	
Expected	8.7	4.07	2.23	
Chi Sq	5.16	0.21	14.9	20.27***

Normally the number of agrees and disagrees ought to swap place on this item. Evidently, the look of the software is not satisfactory.

Item 45 *It is easy to forget how to do things with this software.*

	Agree	Undecided	Disagree	
Profile	11	1	3	
Expected	3.81	2.64	8.55	
Chi Sq	13.58	1.02	3.6	18.2***

Most people agree that this is the case even though they use it on a daily basis. This re-enforces the feeling expressed that it is not easy to move from one task to another (Item 44).

Item 26 *Tasks can be performed in a straight forward manner using this software.*

	Agree	Undecided	Disagree	
Profile	4	4	7	
Expected	10.48	2.57	1.96	
Chi Sq	4.0	0.8	13.02	17.82***

Given the response to Items 45 and 44 it is surprising that the number of people who disagree isn't even higher. However, it is evident that most people do not find it easy to use the software.

Item 41 *The software hasn't always done what I was expecting.*

	Agree	Undecided	Disagree	
Profile	15	0	0	
Expected	7.09	3.09	4.82	
Chi Sq	8.82	3.09	4.82	16.73***

A resounding agree! Everyone has been taken by surprise at some time. This could be due to insufficient training but, given the response to Item 26, perhaps it is more likely that people just find it difficult to remember what they need to do next.

Appendix 3

The Ravden and Johnson Checklist

There are nine criterion based sections in the Ravden and Johnson Checklist and two sections on general usability (Ravden & Johnson, 1989). The checklists from the nine criterion based sections were used for the independent evaluation. Standard tasks were performed in order to carry out the evaluation.

An example of the questions from Section 3 (Compatibility) and Section 4 (Informative Feedback) are reproduced on the next two pages.

	Always	Most of the time	Some of the time	Never	Comments
11. Is information presented in a way which fits the user's view of the task?					
12. Are graphical displays compatible with the user's view of what they are representing?					
13. Does the organization and structure of the system fit the user's perception of the task?					
14. Does the sequence of activities required to complete a task follow what the user would expect?					
15. Does the system work in the way the user thinks it should work?					

16. Are there any comments (good or bad) you wish to add regarding the above issues?

17. Overall, how would you rate the system in terms of compatibility?
(Please tick appropriate box below.)

Very satisfactory	Moderately satisfactory	Neutral	Moderately unsatisfactory	Very unsatisfactory

SECTION 4: INFORMATIVE FEEDBACK

Users should be given clear, informative feedback on where they are in the system, what actions they have taken, whether these actions have been successful and what actions should be taken next.

	Always	Most of the time	Some of the time	Never	Comments
1. Are instructions and messages displayed by the system concise and positive?					

	Always	Most of the time	Some of the time	Never	Comments
2. Are messages displayed by the system relevant?					
3. Do instructions and prompts clearly indicate what to do?					
4. Is it clear what actions the user can take at any stage?					
5. Is it clear what the user needs to do in order to take a particular action? (e.g. which options to select, which keys to press)					
6. When the user enters information on the screen, is it made clear what this information should be?					
7. Is it made clear what shortcuts, if any, are possible? (e.g. abbreviations, hidden commands, type-ahead)					
8. Is it made clear what changes occur on the screen as a result of a user input or action?					
9. Is there always an appropriate system response to a user input or action?					
10. Are status messages (e.g. indicating what the system is doing, or has just done): (a) informative?					
(b) accurate?					
11. Does the system clearly inform the user when it completes a requested action (successfully or unsuccessfully)?					
12. Does the system promptly inform the user of any delay, making it clear that the user's input or request is being processed?					
13. Do error messages explain clearly: (a) where the errors are?					
(b) what the errors are?					
(c) why they have occurred?					
14. Is it clear to the user what should be done to correct an error?					
15. Where there are several modes of operation, does the system clearly indicate which mode the user is currently in? (e.g. update, enquiry, simulation)					

Appendix 4

ITU questionnaire

When the ITU was used within the organisation under study the name of the application of concern was used in the PEOU and PU constructs. However, this was replaced by the term "new software" when it was used in the external organisations as the name of the software was not known. The contacts at the organisations ensured that people were aware of which software was meant. An example of the ITU is printed on the next page. An example of the note which accompanied the ITU is printed on following page. The wording was altered slightly depending on which group was completing the ITU.

ITU QUESTIONNAIRE

Please tick the box which is closest to the way you feel. The scores are intended to try to distinguish between a definite feeling (e.g. Strongly Agree) and a "more this side than that side" feeling (e.g. Agree). Ticking box 4 would indicate that you have no opinion one way or the other. The score range is as follows:

1- Strongly Agree, 2- Quite Agree, 3-Agree, 4-Neither Agree Nor Disagree, 5-Disagree, 6- Quite Disagree, 7-Strongly Disagree

Name: ____________________

	Strongly Agree 1	2	3	4	5	6	Strongly Disagree 7
Learning to operate the software would be easy for me	☐	☐	☐	☐	☐	☐	☐
I would find it easy to get the software to do what I want it to do	☐	☐	☐	☐	☐	☐	☐
It would be easy for me to become skillful at using the software	☐	☐	☐	☐	☐	☐	☐
I would find the software easy to use	☐	☐	☐	☐	☐	☐	☐
Using the software would improve my performance at work	☐	☐	☐	☐	☐	☐	☐
Using the software would increase my productivity	☐	☐	☐	☐	☐	☐	☐
Using the software would enhance my effectiveness	☐	☐	☐	☐	☐	☐	☐
I would find the software useful in my job	☐	☐	☐	☐	☐	☐	☐
I felt sufficiently involved in the decision process	☐	☐	☐	☐	☐	☐	☐
I would like to have been more involved in the decision process	☐	☐	☐	☐	☐	☐	☐
I had no involvement in the decision process	☐	☐	☐	☐	☐	☐	☐
My co-workers would think I should use the software	☐	☐	☐	☐	☐	☐	☐
Generally speaking, I want to do what my co-workers think I should do	☐	☐	☐	☐	☐	☐	☐
My line manager would think I should use the software	☐	☐	☐	☐	☐	☐	☐
Generally speaking, I want to do what my line manager thinks I should do	☐	☐	☐	☐	☐	☐	☐

ITU questionnaire

This questionnaire has been developed as part of a research project and is designed to assess your feelings about the software that you will be using in the future. It is not an assessment of the functionality of the software but of your attitude towards it. Your assistance in this exercise would be greatly appreciated.

Please read each statement and indicate by putting a check in the box which is nearest to your current feelings. A check in the first box would mean that you strongly agree with the statement. A check in the second box would mean that you are pretty much in agreement and so on. The fourth box would mean that you have no feelings one way or the other and then the balance changes towards disagreement with the seventh box indicating strong disagreement. Some of the questions may seem a little strange but please do answer all of them. It generally takes between 5-10 minutes to complete.

All responses will be treated in confidence but it would be helpful if you would put your name.

Once again, many thanks for your help.