THE DEVELOPMENT OF A PSYCHOMETRIC INSTRUMENT FOR HUMAN-COMPUTER TRUST

M. MADSEN

Master of Business

2000

The Development of a Psychometric Instrument for Human-Computer Trust: An Investigation of Trust Within the Context of Computer-Aided Decision-Making.

Maria Madsen
Master of Business

Central Queensland University
Faculty of Business and Law

March, 2000

2334390

Abstract

The definition and measurement of human-computer trust (HCT) was addressed in this study in an exploration of trust as a general human experience and as an outcome of computer-aided decision-making. Trusting computer systems is an increasingly important issue for systems researchers, developers and users due to current market demands to provide and access information and business electronically and the trend toward automation through the use of intelligent systems. There has been little consistency to date in existing HCT research. Neither a robust definition of HCT nor a well-designed psychometric instrument for HCT could be found. This study, therefore, aimed to contribute to a better understanding of the nature and structure of HCT and to provide a carefully designed and tested instrument for the measurement of HCT within the context of computer-aided decision-making. The proposed theoretical model for HCT is based on social attribution theory and the theory of ethopoeia.

The study followed a rigorous process for instrument development from the initial definition of HCT to the final validation of the psychometric instrument. First, a taxonomy of trust and a research framework were derived from an analysis of the existing literature on both human-computer trust and interpersonal trust. Within this framework, HCT can be clearly defined as a specific example of a micro-trust relationship, which can be investigated from both subjective and behavioral perspectives. In addition, as a subjective construct it is shown to be comprised of both cognition-based and affect-based components.

The next stage of the research involved the identification of the underlying dimensions of HCT and indicators for each dimension. The Nominal Group

Technique was used with a group of experienced computer users to identify factors believed to be correlated with HCT. The resultant constructs were compared to constructs from previous research. Nine constructs emerged as the basis for the new instrument. Since multiple indicators improve the content validity of psychometric scales, as many items as possible were selected for each construct.

This set of constructs and items then underwent a series of refinements. The method used for refinement was a modification of the Thurstone scaling technique (Moore & Benbasat 1991, Neuman 1994) with four rounds of sorting. The inter-rater reliabilities for each round were calculated as Cohen's kappa (Moore and Benbasat 1991, Cohen 1960, Fleiss, Cohen and Everitt 1969). The initial instrument with 9 constructs and 74 items was introduced into the first sorting round. Two constructs and 31 items were deleted from the instrument in the first round. This process of sorting and refinement was repeated through the remaining three sorting rounds. The average inter-rater reliability of the scales improved from 0.40 in the initial round to 0.83 in the final round. The instrument was finally reduced to a parsimonious 5 constructs and 25 items.

A pilot study was then performed as a preliminary test of the instrument under controlled conditions with first year students. The pilot study used the grammar checking function in Microsoft Word '97© as the test vehicle. The results from the pilot study suggested that the survey instrument was easily understood by most respondents and that it was able to be completed in a reasonable amount of time. There were no apparent difficulties with any of the items with most negative comments pertaining to the "look" of the survey rather than to the items themselves.

Finally, the instrument was tested for construct validity and scale reliability in a field study with users of operational Taxi Dispatch Systems. Several principal components analyses were performed on the data controlling for the number of factors produced. The results of the one factor model clearly showed that there was one overarching factor, human-computer trust, to which all the variables were highly correlated. The result of a two factor model suggested that the affect-based constructs were more significant than the cognition-based constructs in this study. Lastly, the results of a five factor model suggested that the underlying factors, *faith*, *personal attachment*, *perceived reliability*, *perceived technical competence* and *perceived understandability*, as initially proposed, existed, although not all items loaded on the factors expected.

The scale reliabilities of the instrument, calculated as Cronbach's alpha, were also tested with this data. The reliability of the HCT scale as a whole was found to be 0.94. The alpha's for four of the five sub-construct scales were above 0.84 and the fifth scale had a reliability of 0.74, which may be considered adequate. This finding suggests that the instrument is as concise as possible with each of the scales adding to the overall reliability of the instrument.

The new instrument, which has been shown to be both a reliable and valid measure of HCT, may now be used to more fully investigate the structure of HCT in the context of computer-aided decision-making and the dynamics of HCT development. For example, differences in HCT development among various user groups, particularly cultural differences could be investigated. In addition, the instrument and the development process used here provide a basis for the investigation of HCT in other contexts, such as electronic commerce.

Table of Contents

List of Figures

List of Tables

Acknowledgements

There are many people who have contributed in various ways to the completion of this thesis during the time it has taken to accomplish. These include fellow research students, colleagues, friends and family. Their continued support and encouragement has been greatly appreciated.

I am especially grateful to my principal supervisor Associate Professor Shirley Gregor for her support particularly during some of the difficult times. Her guidance and encouragement through the years helped me to set attainable goals and to achieve them.

I would also like to thank the Faculty of Informatics and Communications for the support they have provided in terms of resources and funding and the Faculty of Business of Law, which provided this support through the early stages of this research.

I must also thank Dr. Petra Lietz who graciously shared with me her enthusiasm for statistics, her knowledge and her instrument development experience. Petra's generosity and willingness to help was very much appreciated.

My thanks also go to the Gladstone Campus of Central Queensland University, especially the Administrative Staff for their interest and encouragement. Thanks also go to all the staff members from whom I borrowed books and who listened to me and advised me from time to time on various matters, including statistics. Special thanks go to Karin Simon (Gladstone Campus Library) who always managed to find even the most obscure sources. Very special thanks go to Pam Brown (Counseling, Careers and Health) who was brave enough to read the thesis in its entirity. Her insight and interest in the topic gave me confidence to complete the task.

I would also like to gratefully acknowledge the efforts of my research assistant Cheryl O'Keefe without whom I might still be entering data. Her tireless efforts and enthusiastic assistance greatly facilitated the completion of this research.

An invaluable contribution was made by the Office Managers and Boards of Directors of Gladstone Blue and White Taxis, Rocky Cabs, Geelong Radio Cabs Co-Operative, Adelaide Independent Taxis, Bundaberg Taxis and Suburban Taxis who encouraged their members to participate in the study.

I would like to thank in particular my husband Christopher Madsen who had the idea to investigate the Taxi Dispatch Systems and my uncle Lawrence Wolff (formerly a Member of the Board of Gladstone Blue and White Taxis) who was a valuable source of information throughout the final stage of this study.

Finally and most importantly, I would like to thank my parents, George and Carmel Zafiropoulos, for instilling in me the desire to learn and for helping me to achieve my goals and a very special thank you to my son Allen George Madsen who has been a constant source of inspiration and who was willing to help whenever he could.

My love and gratitude goes to all of these people, especially my family without whom this could not have been achieved.

Declaration

I certify that this thesis is my own work, based on my personal study and research, and that I have acknowledged all material and sources used in the preparation of this thesis whether they be books, articles, reports, or electronic or personal communication.

I also certify that this thesis has not been previously submitted for assessment at any other University or Institution.

Signature Redacted

20/9/2000

Chapter 1

Introduction

Intelligent systems are often used to assist people with decision-making tasks. Such systems are referred to in this study as intelligent decision aids[1] (IDA). The type of assistance provided by these systems varies according to the needs and requirements of the users. Woods, Roth and Bennett (1989), suggest two distinct paradigms of computer-aided decision making, *instrumental* and *prosthetic*.

Instrumental support systems[2] allow the user to make the final decision while prosthetic control systems make and execute decisions automatically (Muir 1987). Although the two paradigms are clearly distinct, they are not a simple dichotomy. Intelligent decision aid implementations result in a continuum of interactive styles bounded by these two extremes. Thus, depending upon the application, the human user has varying degrees of control over the tasks and processes performed by the machine, from complete to none at all (Sheridan 1988, Turban 1995, Sprague & Watson 1996).

An intelligent decision aid is designed to assist the decision-maker to make better task decisions more efficiently. However, there remains one decision which only the user can make – "Do I trust this machine's advice?" An incorrect or inappropriate response to this question may have serious consequences for the user depending on the task they are performing (Muir 1987). It is arguable that it is more

---

[1] Intelligent decision aids (IDA) may be stand alone systems or integrated components of other systems such as Decision Support Systems (Turban and Aaronson 1998).

[2] For example, a grammar-checking component in a word processing application is an instrumental decision aid because the user is simply provided with information in the form of a suggested correction and an explanation stating the reason for the suggestion. The user of such a system has complete control over how to use that information. On the other hand, an autopilot control system in

important for a pilot to have an appropriate level of trust in an autopilot, in order to use it effectively and safely, than it is for an author to appropriately trust a grammar checker. After all, the immediate safety of the author, as an IDA user, is not at risk if their grammar is incorrect, although, their credibility as an author may well be damaged.

Another question immediately becomes of interest: "How much trust does a user have in the IDA they are using?. In order to answer this question it is necessary to be able to measure the level of trust that a user has in a particular IDA.

Thus, the focus of this study is the measurement of the user's trust in the decisions and advice provided by IDA. This trust is more commonly referred to as human-computer trust (HCT). HCT has been studied using both instrumental and prosthetic systems. For example, instrumental systems have been investigated by Lerch, Prietula and Kim (1993), and Will (1991, 1992) while prosthetic process control systems have been investigated by Lee and Moray (1994) and Muir and Moray (1996).

The nature of trust is discussed in Chapter Two with a review of previous work on both interpersonal and human-computer trust. It is shown that trust, generally, is a complex concept that is related to, yet not completely analogous with, *confidence*. Human-computer trust is defined in this study to be,

> *the extent to which a user is confident in, and willing to act on the basis of, the recommendations, actions, and decisions of an artificially intelligent decision aid.*

---

an aircraft is a prosthetic intelligent decision aid because once it is activated it has control of the aircraft, unless the pilot chooses to manually override the system.

This definition, adapted from McAllister (1995, p.25), has been chosen as that which most clearly and fully states the concept of HCT as it is to be studied here. It encompasses both the user's *confidence* in the system and their *willingness* to act on the system's decisions and advice.

Understanding the nature and dynamics of human-computer trust is an important step toward improving the effectiveness and quality of computer-aided decision-making (Muir 1987, Lee & Moray 1994, Sheridan 1988, Turban 1995).

1.1 Background to the research

The definition, model and measurement instrument for human-computer trust, as proposed in Chapter Two, are based on the theories of social attribution, ethopoeia[3] and the two joint-cognitive systems paradigms discussed above. These theories have been previously tested in human-machine contexts (Lerch, et al. 1993, Nass et al.1993, Woods et al. 1989). The model also takes into account results from previous empirical studies and discussions on human-computer trust including Connors (1994), Dassonville et al. (1994), Lee & Moray (1994), Lerch et al (1993), McDermid (1991), Muir (1987), Muir (1992), Muir & Moray (1996), Nass (1993), Sheridan (1988), Will (1992) and Woods et al.(1989).

HCT is believed to be an outcome when there is human interaction specifically with an intelligent decision aid irrespective of the underlying technology used in the implementation of the system. Whatever the underlying technologies[4],

---

[3] Ethopoeia is, "the assignment of 'selves' and human attitudes, intentions, or motives to non-human entities," (Nass et al. 1993, p.543).

[4] Decision IDA may be implemented using various technologies including ES, ANN, CBR and EA/GA.

IDA have one feature in common which is central to the current research — the human-computer decision making interface (Woods et al. 1989).

At this interface the user interacts with the machine to provide task information, to guide the decision making process or to monitor the actions of the machine. It is at this interface, that the joint-cognitive system exists between the human and the machine. During this interaction the human forms a close relationship with the machine within the decision-making environment. This close interaction between the user and the IDA elicits responses from individual users based on how they perceive the system within the context of that environment.

Nass et al. (1993) demonstrate that these responses are not unlike those elicited when individuals interact with each other. The user necessarily and naturally makes various judgments about the machine, about the decision task, and about themselves while engaged in the decision-making process. Specifically, HCT may be seen to be the outcome of the user's judgment of the *trustworthiness* of the IDA within a particular context.

Lerch et al. (1993) found that at various stages of the decision-making interaction users attributed the system's trustworthiness to various causes, either internal or external to the computer. Their findings are thus consistent with previous work on social attribution theory Kelley (1973). Kelley discusses the level of information available to the attributor within a particular context as the basis upon which attributions will be made about another's behavior.

Just as decision-makers are known to exhibit varying levels of judgement confidence (Pincus 1991) while decision-making, they are also known to exhibit varying levels of trust in the IDA they use (Lerch et al. 1993, Muir 1987, 1996).

Lerch et al. (1997) found that user attributions about an IDA affected their level of confidence in the machine's advice. HCT is thus based on the user's perception of certain characteristics of the IDA,or *causes* of its behavior, which they have attributed to the computer based on their experiences. Greater knowledge and a better understanding of HCT, therefore, will assist efforts to improve computer-aided decision- making and the adoption of new decision support technologies (Muir 1994).

1.2 The research problem

A psychological construct, such as human-computer trust is "an abstract theoretical variable that is invented to explain some phenomenon which is of interest to scientists." (Young 1996). Given that the phenomenon known as human-computer trust exists and that computer scientists are interested in it, the principle problems arising for researchers in this field are: "What is it?"; and "How can it be measured?".

Previous HCT researchers have used a variety of definitions and a variety of measurement instruments some of which have been adapted from interpersonal trust studies. This work has provided a variety of results which are, at times, difficult to interpret. The literature review in Chapter Two demonstrates that the current state of human-computer trust research does not adequately explain the human-computer trust construct. This raises the essential research question in this study which is: *are there as yet unidentified variables involved in this phenomenon which, when measured appropriately, better explain the nature of human-computer trust?*

The aim of this research is, therefore, to develop a psychometric instrument specifically designed to measure a user's level of trust in an intelligent decision aid.

In an attempt to address some of the inadequacies of previous research a rigorous instrument development process was followed in this study.

The intelligent decision aids of interest in this study fall within a range which begins with less specialized, readily available systems, such as the grammar checker sub-systems embedded in word processing software, to more highly specialized, purpose-built systems such as loan assessment, financial planning and scheduling systems. The latter systems may be stand alone systems or components of larger systems such as Intelligent Decision Support Systems (Sprague & Watson 1996, p.19).

Finally, it is noted that the systems studied here do not include those systems which are fully automated such as critical control systems and do not encompass those systems which are fully manual. HCT, as defined here, is irrelevant to fully manual, non-computer-based systems. However, it is believed that the concept of human-computer trust as described in this study may be applicable even to the most fully automated system in the event that a human operator is charged with monitoring that system (Muir 1987).

The objectives of the current research are:

- to define the concept of human-computer trust;
- to develop a theoretical model of human-computer trust;
- to develop an instrument to measure human-computer trust based on the model;
- and to demonstrate the effectiveness of a rigorous development process.

Specifically, the research addresses the following questions:

1. What is human-computer trust?
2. What are the components of human-computer trust?
3. Is the new instrument a valid and reliable measure of HCT?

4. How effective is the multi-staged process used here for instrument development?

The first question is answered in Chapter Two Section 2.6 where HCT is defined within a general framework of trust, which is developed from existing trust theories in Section 2.3 of the same chapter. As such, the definition is a robust one which is consistent with other aspects of this study.

In order to answer the second question, the major components of trust are identified in Chapter Two based on the proposed trust framework and from the findings in the first stage of the research, reported in Chapter Three, Section 3.1, these underlying variables are identified, defined and discussed.

The third question addresses the primary objective of this study specifically, which is the content and construct validity and the scale reliabilities of the instrument. The reliability and construct validity of the new instrument is determined in the final stage of the research and is reported in Chapter Four, Sections 4.5.2 and 4.5.3, respectively.

Finally, the usefulness of the rigorous, multi-staged instrument development process is considered and discussed in Chapter Five Section 5.3.3.

1.3 Justification for the research

The past decade has seen the proliferation of IDA which utilize artificial intelligence technologies to enhance organizational productivity, efficiency and effectiveness. These technologies have enabled organizations to distribute decision making to lower levels of management, particularly through the use of Expert Systems (ES). ES make expertise available to workers at all organizational levels allowing them to make decisions which would otherwise have to be channeled back through the organization to the top management level (Sprague et al.1996).

Artificial intelligence technologies, such as ES and Artificial Neural Networks (ANN), have also allowed repetitive, dangerous and complex processes to be automated. Computerized critical control systems, such as the these, while usually monitored by human operators, are effectively in continual control of the operating decisions that need to be made. These decisions are required to keep the overall process functioning at an optimal and safe level. These control systems are intended to be much more efficient than a human operator, sampling and analyzing system data continually and consistently. However, it is not unusual to have human operators monitoring these systems with the ability to override decisions which they perceive to be incorrect.

Thus, it can be seen that decision making is a primary function in both business and process management (Turban & Aronson 1998). The study of decision makers' attitudes and behaviors (Pincus 1991, Power, Meyeraan & Aldag 1994) is of particular interest in management science research. The study of decision makers' knowledge and problem solving methods, or expertise, is also of interest in the area of intelligent systems research, particularly with regard to knowledge based systems such as ES (Clancey 1983, Dhaliwal & Benbasat 1996, Freeman, Jones & Field 1994, Gregor 1996, Turban & Aronson 1998).

Following from these interests in human behavior and attitudes during decision making, is the desire to understand how decision making is affected, if indeed it is affected, with the introduction of computerized decision making tools and new technologies. The phenomenon of trust that a decision-maker has in a machines' advice is product of the introduction of intelligent systems into business and industry. Human-computer trust research performed to date, as briefly discussed above, examines both business management decision making as well as operator responses to critical control systems.

In HCT research, as in most areas of human behavior, including information systems research, there are many independent variables. The large number of factors involved in human relationships makes it difficult to identify those independent variables likely to be correlated with a concept such as human-computer trust. The inability to test for or control every possible variable however, does not preclude attempting to measure human phenomenon, such as HCT, as well as possible. With a well-planned approach to the investigation it is possible to design measurement instruments which account for some of the more highly correlated independent variables (Cronbach 1970). While such an instrument will never measure every correlation between all variables, these instruments are often sufficient to provide valuable insights into the dynamics of relationships and the attitudes of those involved. A valid and reliable instrument may even be able to predict future behavior and outcomes.

Without a valid, reliable HCT measurement, however, there can be little advance made in the understanding of the role and relevance of HCT in the adoption of and patterns of use of IDA beyond that which has already been accomplished. With the ever increasing use of internet technologies for commerce and the distribution of many organizational systems, including decision support systems, understanding and measuring human-computer trust is now more than ever an area of concern to systems developers and users alike.

In order to begin to address these issues and clarify the nature of HCT, the present study is based on a clear, concise definition of HCT, which is relevant to the context of IDA use. This concept of trust is based upon the comprehensive framework of trust types and trust relationships, which is developed in Chapter Two from a review of existing trust literature. The more specific HCT definition and broader trust framework provide a foundation upon which to further investigate the components of HCT and quantify their measurement as is reported in Chapters Three and Four of this report.

1.4 Methodology

The measurement instrument was refined and tested through several stages in this study. The research design aimed at improving the construct and content validity of the instrument prior to the main analysis. Each stage was designed to address at least one of the fundamental requirements for the development of a valid and reliable instrument. Each stage necessitated the use of a different research method and analysis. Therefore, several different investigative methods, both qualitative and quantitative, were employed depending on the research objectives at each stage.

First, the theoretical construct or constructs to be measured were precisely defined. A careful and precise definition of the construct is the foundation upon which construct validity is built. Any ambiguity in the definitions may cause ambiguity later in the instrument. The validities relevant to this research are discussed in more detail in Chapters Three and Four of this report.

Next, because a construct such as human-computer trust is a complex, abstract variable, a direct measurement, such as, "How much do you trust your decision system?", will not provide sufficient information about the nature of the construct (Cohen, Cohen, Teresi, Marchi & Velez 1990 , p.187) It is therefore preferable to identify factors, other variables, that are involved in the formation of the construct of interest. These factors are also referred to as the dimensions of the construct and may themselves be constructs or they may be observable behaviors or other concrete measures.

It might be thought that the more factors measured the more fully the construct can be described. To the extent that these factors can be measured and that they correlate with the construct of interest, this is true. However, the number of factors which might be measured, especially when using a questionnaire to measure psychological constructs such as attitudes, is limited by the cognitive ability of participants to distinguish between the factors (Lietz 1998). The identification of these dimensions is the first step toward assuring content validity. Content validity is a judgement about how well the constructs and later, the items, belong to the universe of discourse. The process of factor identification necessitates an inductive approach. Grounding the theoretical constructs within a particular context, in this case intelligent systems use, enhances the content validity of the instrument. This stage of the research is discussed in greater detail in Chapter Three.

In order to measure constructs, other than those for which observable behaviors are available, questions, or items, must be created which "tap into" the attitudes of interest. To achieve this, items for each of the dimensions must be found or created. Several items are created for each construct. These items are grouped according to the dimensions they are intended to measure. These groups of items are referred to as scales. The quality of the groupings can be demonstrated quantitatively through analysis of the responses to the questions later in the research. This is a measure of the way items relate to other items in the scale and is known as the internal consistency of the scale or the scale reliability. Chapter Three provides a complete description of the process of scale development used in this study.

Validation of the instrument was the final stage in this study in which the instrument was tested quantitatively for construct validity and scale reliability. This step, however, is just the beginning of the overall validation cycle of a psychometric instrument. The initial validation is considered to be an exploratory analysis from which the instrument can be further refined and tested in confirmatory studies with various sample populations. A confirmatory study is able to demonstrate other forms of reliability such as criterion reliability and representative reliability. Confirmatory analysis of this kind is outside the scope of the current research.

1.5 Thesis Outline

The thesis is comprised of the following chapters.

Chapter One gives an introduction to the topic, the aim of the study, a brief summary of previous research, a discussion of the scope and significance of the study, a statement of the research questions and an overview of the remainder of the thesis.

Chapter Two discusses the nature and structure of various types of trust. Reviews of previous interpersonal and human-computer trust studies are examined in terms of definition, theory, measurement scale and findings. In this chapter a taxonomy of trust is developed which is used as a framework within which human-computer trust is defined. Finally, a definition and model of human-computer trust are presented. The theory that was chosen as being most suitable to describe human-computer trust is one which combines social attribution models with the theory of ethopoeia presented by Nass et al. (1993).

Chapter Three addresses instrument development issues and describes in detail the stages involved in the creation of the measurement instrument. Construct identification and content analysis is performed using techniques such as the nominal group technique and Thurstone scaling. Findings from the initial field investigations and preliminary scale reliability tests are reported along with the refined instrument.

Chapter Four presents the methodology used for instrument validation and discusses the operationalization of the research model. This was accomplished by administering the new scales to groups of users who use an IDA regularly in their work. The validity of the instrument in terms of construct convergence and divergence and the scale reliabilities are reported.

Chapter Five summarizes the findings from the various stages of the research and concludes the report. A synthesis of the results from all investigations and a discussion of the implications of the findings and the value of the instrument development process is presented here. This chapter also discusses the limitations of the instrument, the model and the study as well as opportunities for future research some of which have been identified above.

1.6 Limitations and key assumptions

This study deals specifically with intelligent systems which are designed to aid decision-making. These systems might be designed with various technologies but only those which either provide advice, or make decisions subject to the user's discretion, are relevant to this investigation.

It is assumed that the user may not have knowledge of the underlying technology nor the specifics of how the system arrives at its output. It is also assumed that the users may or may not be experts in the task domain. Thus they will have more or less task knowledge depending on their level of experience. User experiences with the system being investigated may range from a few months to several years.

Since one objective of this study is to determine quantitatively the validity and reliability of the new measurement instrument it is necessary to consider what constitutes a sufficient sample size. Sufficiency of sample size is related to the type of quantitative analysis to be undertaken. In this case the measurement instrument, was analyzed using Principal Components Analysis (PCA). PCA is reasonably robust to sample size as long as the sampling adequacy, to be discussed in detail later in Chapter Four, can be shown to be relatively large. As a rule of thumb, it is prescribed that the researcher should have a ratio for the sample size to the number of items of five to one (Neuman 1994).

Finally, the rsults of this study may not be generalized too widely. The first caution against generalizing the results is that this study is context specific. The second caution is that this study involves only one sample group and only one method of measurement. Thus the representative reliability of the instrument was not tested nor was it possible to test the criterion validity of the instrument (Neuman 1994, p.131).

1.7 Conclusion

This first chapter has laid the foundation for this research report. The research problem and the research questions have been introduced. The research topic has been justified and the research methodology to be undertaken has been briefly described. An outline of the report was presented and limitations and assumptions were addressed.

The remainder of this thesis describes and discusses the research process and data analysis in detail.

Chapter 2

Background to the Research

"The amount of knowledge necessary for trust is somewhere between total knowledge and total ignorance (Simmel, 1964). Given total knowledge there is no need to trust, given total ignorance, there is no basis upon which to rationally trust." (McAllister 1995)

In Chapter One the aims and objectives of this study have been described and the design of the instrument development process has been broadly introduced. In this, the second chapter of the thesis, the background to the current research is investigated more fully. First the task of defining of HCT as a concept is undertaken through an exploration of existing trust definitions and research studies. A framework of trust is then developed which comprehensively describes trust in all contexts. It is within this framework that HCT is fully defined, both as a psychological construct and as a unit of study. Next, previous HCT research is reviewed and compared with respect to definitions, perspectives, instruments and methodologies. Finally, a model of HCT is proposed which reconciles a number of previous studies.

2.1 What is Trust?

The definitions of trust are as many and varied as the number of studies investigating trust. Trust means different things to different people and what it means to trust another is dependent upon circumstances and the extent to which trust is required. Thus, it is difficult to find one definition that *fits* all situations. Trust needs to be defined within the context of the relationship being investigated. Yet, we use the term trust as if we implicitly understand what it means. Moreover, we assume that we instinctively know how to trust and why we trust. Even a cursory examination of the existing trust literature soon belies this assumption. Trust is a complex human experience complicated by our beliefs, emotions, comprehension and expectations. Some time will be spent in the following sections discussing trust relationships and types of trust in an effort to bring order and clarity to, "this most problematic topic", (Misztal 1998).

2.1.1 Trust as expectation

Barber (as cited in Muir 1987 p.528, italics added) has suggested that trust can be defined in terms of three specific expectations:

1. our very general expectation of the *persistence* of the natural (physical and biological) and the moral social orders (i.e. we expect natural physical laws to be constant, human life to survive, and mankind (and computers) to be good and decent, respectively);
2. our expectation of *technically competent* role performance from those involved with us in social relationships and systems;
3. our expectation that partners in an interaction will carry out their *fiduciary obligations and responsibilities,* that is, their duties in certain situations to place others' interests before their own.

Rotter (1980) has defined trust as a, "generalized expectation held by an individual that the word, promise, oral or written statement of another individual or group can be relied on.", and Shaw (1997) has defined trust to be a, "Belief that those on whom we depend will meet our (positive) expectations of them."

Of these three definitions only the latter avoids introducing specific dimensions into the definition. Given that trust needs to be defined within a particular interactive context, the latter definition is the only one of this set which will lend its use to various relationships. The other two definitions set boundaries on the elements of trust that can be measured and as such are not sufficiently robust definitions as they can not be applied across various trusting relationships. They narrow the field of investigation prematurely.

Another difficulty with these expectation theories is that defining trust as an expectation necessitates studying trust from a pre-test point of view. One can not, it seems, measure an individual's initial expectations either during or after there has been interaction. Once an experience has occurred the individual's expectations must necessarily be changed in some way. Whether they are changed positively, that is, the other's behavior reinforced or was consistent with the individual's expectations, or negatively, that is, the other's behavior was contrary to what was expected of them, the original expectations will be changed.

There is no existing study which has been designed to measure or describe the participants' initial expectations of their partners. Researchers have defined trust as being an expectation but have not incorporated this into their research designs. In order to measure an expectation in an experimental design, participants would necessarily be asked questions about their expectations prior to having had any type of experience with a perspective trustee. All studies reviewed to date have used either post-test instruments or in-situ, behavioral observation as their method of measurement. Expectations, therefore, have not yet been truly investigated.

It is clear that trust defined as an expectation or set of expectations held by an individual about another is insufficient as a definition of the concept of trust. However, Barber's definition will be revisited later in this report as it has been incorporated, via Muir's (1987) thesis, into several other human-computer trust studies.

2.1.2 Trust through experience

Most trust research has focused on a participant's experiences with another. One of the most common experimental designs in early trust research was that of game play. Scenarios were created such that the participants were placed in some specific situation which necessitated that they trust their gaming partner. Their behavior in this situation was monitored and conclusions drawn about the level of trust that they had in their partner.

The remainder of the studies, that adopt the experience perspective utilize questionnaires that ask participants about their perceptions of some specific other with whom they have a relationship. These studies focus on the participants' experiences with another and for the most part have used multi-dimensional models of trust and multi-item scales.

Many of these studies have been based on the work of Rempel, Holmes and Zanna (1985). While Rempel et al. do not offer an explicit definition of trust, they proposed and tested three dimensions of trust. The three dimensions are:

- Faith - "trust that is not securely rooted in past experience," (p.97)
- Predictability - "the ability to forecast a partner's future actions," (p.96)
- Dependability - "attributions about a partner's behaviours that are seen to reflect the partner's dispositional qualities of trustworthiness," (p.97).

On of the outstanding features of this theory, which has been most widely adopted by HCT researchers to date, is that trust is defined to be multi-dimensional. Another is that the dimensions themselves, with the exception of faith, are characteristics attributed to one's partner based on experience. The individual's prior expectations of their partner are not considered in this theory.

Once again the researchers have focused on specific dimensions of trust, which indeed may be applicable to the close personal relationships that they investigated. Although, little explanation is offered as to how these three dimensions were identified and no suggestion was made in their study that these dimensions might be applicable outside the context for which the model and instrument were designed.

2.1.3 Experience versus Expectation

It seems likely that an individual's level of trust is in fact due to a combination of both their expectations of the other and, over time, the experiences that they have with the other person. Shaw (1997, p.22) believes that, "it is the inconsistency between what one expects and what one observes that raises doubts about the motives or ability of those in whom we place trust."

This observation is consistent with the Theory of Planned Behavior (Ajzen 1991, Taylor & Todd 1995) which "postulates that behavior is a function of salient information, or beliefs, relevant to the behavior (in this case trusting another)," (Ajzen 1991, p.189). Behavioral beliefs, for example, are beliefs formed about the value of certain behaviors based on associations made between the behaviors and expected outcomes. The behavioral beliefs and and the other belief sets in this theory, normative beliefs and control beliefs, are said to directly influence an individual's behavioral intention. This theory also suggests that behaviorial intention precedes the manifestation of the behavior.

The correlation between prior expectation and experience in the development of trust has yet to be studied in either interpersonal or human-computer trust research, although Muir and Moray (1996) have included variables from trust research based on both expectation and experience in their study of human-computer trust. There remains a great deal of work to be done before one could be confident that this issue has been dealt with adequately. The definition of trust, however, does not have to be encumbered by issues which better belong to theories about the structure and dynamics of trust.

It is clear that what is needed is a definition of trust that does not specify dimensions, is not focussed on a specific perspective, and which will remain applicable at any stage in an investigation. Since many definitions of trust exist, it was reasonable to explore these for a suitable one before inventing yet another which could further add to the confusion.

The following definition of trust was chosen as one which most clearly defines the concept of trust while avoiding the encumbrances of the issues discussed above (McAllister 1995, p.25).

> *Trust is the extent to which a person is confident in, and willing to act on the basis of, the words, actions, and decisions of another.*

2.2 The nature of trust relationships

Trust relationships are dynamic relationships. As such, the level of trust that, an individual, the trustor has in another, the trustee, changes over time. It is likely that there are many antecedents to the development of trust. These antecedents continue to operate throughout the life of the relationship thus becoming mediators in the development process (Johns 1996). Although trust levels may be difficult to change once established (Muir 1987, Einhorn & Hogarth 1978) they can and do change with time.

2.3 Taxonomy and framework of trust.

Although there has been no systematic classification of the various types of trusting relationships to date, researchers have identified a variety of trust types. There are two seemingly disparate views of trust relationships based primarily on the point of view of the researcher. Alternatively, researchers from the social sciences view trust as a social experience which necessitates a broad description and discussion of social interactions (Lewis & Weigart 1983, Barber 1983, Misztal 1998). These researchers argue that trust can not be fully understood through simple psychometric testing. Researchers from a psychological tradition view trust as being internal to the individual and are proponents of psychometric testing as a reliable means of elucidating the nature of trust.

The sociological and psychological perspectives are here to more fully describe trust as a phenomenon of human activity. Trust exists in a specific person and must be directed toward another person or entity, whether those others are large groups, institutions, individuals or machines. Trust may be shaped by the person's nature, their social experiences and their acquired knowledge, none-the-less it remains that trust does not exist without the individual involved.

The following investigation of the nature of trust and the integration of existing research perspectives resulted in the parsimonious framework illustrated in Table 2.1. An inclusive, top-down approach was adopted in the analysis of existing research beginning with the broad social perspective of trust and ending with the specific, psychological perspective of trust.

Table 2.1 Trust Taxonomy and Framework

Trust type / Relationship Type	Perceived		Manifest
	Cognition Based	**Affect Based**	
Macro (general)	This society is lawful and we enjoy personal freedoms protected by these laws.	I like living in a free, orderly society.	I choose to continue to live in this society and I obey the laws that protect my personal freedom.
Micro (dyadic, specific)	My partner has the same goals as I do.	I love my partner.	I allow my partner to make decisions that will effect my life.

2.3.1 The Sociological Perspective of Trust: A Classification of Trust Relationships

Barber (1983) in his book *The Logic and Limits of Trust,* distinguishes between several types of general trusting relationships including trust in society as a whole, trust in social institutions such as governments and trust in professionals. Other researchers have investigated specific types of interpersonal trust. Among these are trust in physicians (Anderson & Dedrick 1990, Carterinicchio 1979), trust in academic relationships (Imber 1973), children's peer trust (Rotenberg 1995), trust in nursing staff (Johns 1996), trust in various organizational relationships (Hosmer 1995, Mayer 1995, McAllister 1995, Ross & Weiland 1996, Schindler 1993, Zand 1972), as well as trust in relationship marketing (Morgan & Hunt 1994) and trust in negotiation (Ross 1996).

From the nature of these studies, trust relationships can be seen to belong to either one of two major types based on their scope. The first is a generalized relationship between an individual and a larger entity such as society as a whole or nature itself. This type of trust relationship, in economic terminology, might well be referred to as a *macro*-trust relationship.

Macro-trust is generalized and extends to the trust that people have for society, for the government, for the legal system and in humanity as a whole (Barber 1983). These a priori beliefs and social disposition may, in fact, impinge on the person's overall willingness to trust (Yamagishi 1986) in any environment. The individual's macro-trust may influence their micro-trust relationships discussed below.

The second type of trust relationship is a specific relationship between individuals, a dyad. This may be referred to as a *micro*-trust relationship. Micro-trust is specific and dyadic, involving only two actors, the trustor and the trustee. These relationships are among the most highly investigated trust relationships and of most interest in this investigation since human-computer trust is a micro-trust relationship.

2.3.2 The Psychological Perspective of Trust: A Classification of Trust Types

It is evident from the discussion that has preceded that the level of trust a person has in another is a subjective judgement. At some point in time this subjective trust can manifest itself as behavior, with the trustor acting in a trusting way toward the trustee. Kee and Knox (1970), discuss the transition from subjective to behavioral trust as being one of importance that has thus far been neglected by many researchers.

> "The subjective state … is important since it would be of interest to know at which point or threshold *subjective trust* becomes manifest as *behavioral trust*, i.e. to what extent P must 'feel' that he trusts O before he will in fact make a trusting decision. Even more interesting, the threshold will undoubtedly vary with a variety of situational, structural, and/or dispositional factors, eg. incentives of P's own trustworthiness."

The failure of some researchers to distinguish between these two types of trust has led to, perhaps erroneously, the results of behavioral studies being used as indicators or predictors of perceived trust. A clear distinction can and should be made between perceived, or subjective, trust and manifest, or behavioral, trust (Lewis & Weigart 1983).

2.3.2.1 Perceived Trust

Perceived trust is an individual's perception of the trustworthiness of their partner, whether that partner is another human or in the case of this study a computer system. The individual's perceptions are based on both intellectual or *cognition*-based knowledge and emotional or *affect*-based knowledge of the trustee (McAllister 1995), as illustrated in Table 2.1. Perceived trust most probably arises from both the trustor's expectations of and experiences with the trustee (Shaw 1997).

Many interpersonal trust studies are specific to very intimate relationships such as marriage and investigate perceived trust rather than the trusting behavior of the individual toward their partner (Anderson & Dedrick 1990, Johnson-George & Swap 1982, Larzelere 1980, Rempel 1985, Ross 1996, Rotenberg 1995, Rotter 1968, Rotter 1980, Schindler 1993).

Perceived trust may change with new evidence although it is thought by many investigators that once a level of trust is established it may be difficult to alter. This may be explained by the *persistence of belief* theory (Begg, Anas & Farinacci 1992, Rosenbaum & Levin 1969).

2.3.2.2 Manifest Trust

Manifest trust is the way in which an individual acts or behaves in situations that require trust in their partner. "The manifestation of trust on the cognitive level of experience is reached when social actors no longer need or want any further evidence or rational reasons for their confidence in the objects of trust," (Lewis & Weigert

1983, p.970). Manifest trust is, therefore, the manifestation of perceived trust as an individual's behavior toward a specific other, within the context of a specific situation in which trust must be given. Variables which might mediate the subjective to behavioral transition may include the risk involved in trusting a partner, the individual's self-confidence, and as discussed above the individual's general willingness to trust.

Many of the predictors of trust in interpersonal trust research are based on observations of participant behavior in two person games such as the Prisoner's Dilemma Game (see for example Wrightsman 1966). These game playing models have been challenged as being incomplete measurements of interpersonal trust due to the fact that the experimental designs base trust measurement on observations of behavior alone ignoring the individual's attitudes or perceived trust (Kee & Knox 1970).

2.3.3 Implications for continuing trust research

A trust scale designed for specific interactions and tested under specific conditions may not necessarily be valid in all relationships (Johnson-George & Swap 1982). There are many different types of trust situations possible within the above framework and it is reasonable to expect that each type is characterized by a different set of parameters that influence the level of both subjective and behavioral trust. An examination of previous research strongly suggests that there is a need to design trust scales that are tailored to measure trust within specific relationships and situations.

In light of the above discussion, the results of all interpersonal studies need to be treated carefully within the constraints of their particular research designs as either measures of perceived or manifest trust; and not necessarily as valid general measures of interpersonal trust. Thus the practice of adopting these interpersonal trust scales and adapting them to human-computer trust situations by rewording questions is of doubtful value in studying human-computer trust. It appears preferable to design studies aimed specifically at HCT, that can add to our understanding of this phenomenon.

> "Despite the variety of motivations underlying trust relationships, they show ... some common characteristics. These common features of trust relationships result from human beings, as emotional, cognitive and instrumental oriented agents, seeking to ensure that their social relations and arrangements meet their emotional, cognitive and instrumental needs and conform to their sense of what is appropriate in each context.", (Misztal 1998).

In the following section the similarities between human-computer relationships and interpersonal relationships are considered.

2.4 The relationship between a human and an intelligent decision aid

The decision making relationship between a human and an IDA goes beyond the basic interaction between human and computer as studied by many information systems researchers. An IDA may be perceived by the user to be more than a mere tool. Because these systems behave as experts, providing advice or even taking action without intervention by humans they may be, more than any other type of computer system, attributed with human characteristics by their users.

2.4.1 Human responses to intelligent systems

It is well known that humans attribute characteristics to other humans based on their experience with each other (Brickman 1979, Elig & Frieze 1979, Kelley 1973, Rempel et al. 1985, Russell 1982). It is possible that such attributions are made by humans about the computer systems that they are using, particularly intelligent computer systems. This type of response is known as ethopoeia (Nass et al. 1993) and is the equivalent to social attribution except that the object of the attribution is inanimate rather than human.

Social attribution theory states that a person will attribute to a partner certain personal characteristics, such as reliability, based on information gathered through their experience with that partner. The type of information gathered can be described by three criteria: *consensus, consistency,* and *distinctiveness*. Consensus is the degree to which other people's behavior is similar to that of the person in question. Consistency is the degree to which the behavior occurs repeatedly in similar situations. The last criterion, distinctiveness, is dependent upon the predictability of behavior in various situations (Kelley 1973).

It is the combination of these three criteria that establish the foundation for attribution. For example, low consensus, high consistency and low distinctiveness results in an internal attribution. While a combination of low consensus, low consistency and high distinctiveness results in an external attribution (Bernstein, Roy, Srull and Wickens 1991). Kelley (1973) also states that there is a tendency for individuals to make attributions before they have all the information they need to make an accurate assessment of the causes of another's behavior. In many cases it simply isn't possible for an individual to gather all the facts before making a judgment about the cause of another's behavior. This tendancy to make attributions based on insufficient information is described as the *fundamental attribution error* (Bernstein, Roy, Srull and Wickens 1991) and usually results in *internal attributions*.

The theory of ethopoeia states that a computer user attributes "human characteristics" to the computer as if the computer was a cognisant being. This theory that humans interact with computers similarly to the ways in which they interact with other humans is supported by several authors (Kennedy, Wilkes, Elder, & Murray 1988, Nass et al. 1993, Steuer, Henriksen, & Dryer 1993).

Conflicting theories such as the deficiency theory and the proxy theory suggest that computers are not, normally, attributed with human characteristics by users and should be considered simply as tools. The user's attributions of characteristics to the computer are explained in terms of dysfunctional behavior or in terms of indirect attributions to other referents, such as the programmer.

Nass et al. (1993), in particular, considered these theories. Briefly stated, the deficiency theory maintains that quasi-social or para-social behaviors exhibited by individuals toward machines result from some psychological or social dysfunction. The proxy theory maintains that because machines are human artifacts the users may "adopt a perspective that they are interacting with a human creator or programmer … because (the) machines embody and reflect the attitudes, conceptions and intentions of their producers", (Nass et al., 1993, p. 544). The results of the Nass et al. (1993) study did not support either the deficiency or proxy theory. People displaying ethopoeia were shown to be neither dysfunctional nor did they have an imagined relationship with the developer of the machine (Nass et al., 1993, p.556). They found that when interacting with a computer, "individuals … use social heuristics that are inconsistent with their espoused beliefs about machines".

Participants made evaluations about the information presented by various computers based on social rules, which were inconsistent with the characteristics of computers indicating that there are more complex reactions taking place than have previously been considered. In fact, these authors suggest, individuals may assign to the computer a separate and unique identity, especially when the technology is perceived to change initial information or to provide additional information such. The computer may be seen as a *messenger* and not the *medium* and may be attributed with social qualities similar to those that we attribute to each other (Nass et al., 1993, p. 556).

Much research within the information systems research community suggests, that a user's perceptions mediate the way in which they accept and use technology. Some of these perceptions appear to be based on causal attributions that the user has made about the system. Furthermore, various attributions appear to affect the user's level of confidence in IDA in various ways (Lerch et al. 1993, 1997). Information systems researchers have also found that psychometric instruments are useful tools for measuring user perceptions about computer systems. Well designed instruments are able to a large extent to predict behavior (Moore & Benbasat 1991, Taylor & Todd 1995).

2.4.2 The computer as a decision-making partner

As discussed briefly in Chapter one, intelligent decision aids can be designed using a number of technologies. Each of these technologies give the computer *artificial intelligence* and each has particular advantages and disadvantages in specific applications. Irrespective of the underlying technology the resultant system is one which gives the user the illusion that the system is *intelligent*. That is that the system is "thinking" about the problem and arriving at a solution, which in the case of expert systems, can even be explained to the user via the explanation subsystems (Gregor 1996, Turban & Aaronson 1998).

Intelligent decision aids provide users with support in solving problems. They provide the user with advice based on the input provided by the user themselves and data gathered from various other sources. This advice is generated by the system by applying pre-determined logical and mathematical models to the data and other inputs. The human and the computer are thus in partnership and are sometimes regarded as a joint-cognitive system. DeGreef & Neerincx (1995) suggest that the failure to fully analyze the needs of the human-computer relationship may result in computer systems that are inappropriate or incomplete.

The theory of joint-cognitive systems is comprised of two paradigms based upon the type of decision support provided by the computer. As explained by Woods, Roth and Bennett (1989) there is the cognitive-tool-as-prosthesis paradigm and the cognitive-tool-as-instrument paradigm.

> "In the cognitive-tool-as-prosthesis paradigm, the *system* is defined as the machine expert, and *effective* means usually correct machine solutions. In the cognitive-tool-as-instrument paradigm, the system is defined as the combination of human and machine (the human-machine cognitive system) and *effective* means maximizing joint performance; i.e. performance of the whole should be greater than the performance possible by either element alone," (p.129).

Thus, there are two major types of IDA, prosthetic and instrumental. The amount of interaction and more importantly the amount of control that the user has over the final decision differs considerably between the two as does the amount of knowledge that the user must have in order to use the machine appropriately and maximize the effectiveness of the decisions being made.

2.5 Trust in the context of computer-aided decision making

One of the concerns for system's owners and developers is that users may be lulled into a false sense of security when using an IDA. Their use may be *dysfunctional* in the sense that while they may be more confident in and less anxious about their decisions having had the support of an IDA, the actual quality of these decisions may not be any better than those which might have been achieved without the IDA (Will 1992).

The user of an IDA may or may not have sufficient knowledge to judge whether the computer is providing useful or even correct advice. In cases where the user has a choice whether to accept the computer's solution or not it is their *trust* in the computer, based on their perceptions of the system, that influences their choice. An IDA that is not trusted will not be used as its designers intended (Muir 1987).

The desired level of trust is a level that is consistent with the accuracy and correctness of the IDA. A level that is inconsistent with the accuracy and correctness of the IDA is termed *mistrust*. A complete lack of trust in the system at all is termed *distrust*. The ramifications of this are that mistrust may result in misuse and distrust may result in disuse (Muir 1987).

In extreme instances of mistrust there may be serious consequences for both the users and the organizations for which they work. For example, in the case of mistrust in a correct and accurate IDA the advice may be ignored. For the user the consequences of their mistrust may result in missed opportunities to improve decision quality or their own decision-making processes. Furthermore the system may be under-utilized resulting in costs to the organization of opportunities, money, and customers. In the case of a *critical control system* either a good decision may be manually overridden or a bad decision may be accepted resulting in critical system failures. The costs to the organization in these cases may be equipment, productivity or even lives. Ultimately, the liability for such loss will rest upon a particular person or persons – not the machine.

Muir (1987) believes that inappropriate levels of trust in users can in fact be "calibrated", that is aligned to the system's accuracy and correctness. Being able to better align human-computer trust levels with IDA accuracy and correctness could lead to increased decision quality, system productivity and reduced risk. Alignment might be achieved through better system design, better implementation protocols, or better training and retraining methods.

"One of the paradoxes of trust is that trust cannot grow unless we take risks that may result in distrust," (Shaw 1997 p.24). As the level of human control over the system decreases and the criticality[5] of the functions performed by the system increases, the amount of risk associated with the decision increases. Likewise as the complexity of the decision task increases the amount of knowledge required by the user to decide whether the system is correct or not increases.

Furthermore, the greater the amount of knowledge the user needs to have for a particular decision task the less likely it is that they will have complete task knowledge and the more likely they are to need to use the IDA to assist them (Turban 1995). Also, the less knowledge the user has about the task the more their need to *trust* the IDA they are using. To continue Shaw's argument, the user must, "risk being wrong" before they can determine whether it is safe to trust the system. "Trust and risk give rise to each other; it is rare to find one without the other," (Shaw 1997).

[5] The criticality of the system is determined by the situation in which it is used and the importance placed on the system's function by owners or users of the system.

2.6 What is Human-Computer Trust?

> *Human-computer trust is the extent to which a user is confident in, and willing to act on the basis of, the recommendations, actions, and decisions of an artificially intelligent decision system (adapted from McAllister 1995, p.25).*

This definition fits well with the results of the literature analysis already presented. Human-computer trust is not simply an *expectation* held by the user about the performance of the IDA they are using, nor is it simply the *act* of trusting, which in the case of an IDA translates into the *actual use* of the system's advice, that is manifest trust.

HCT is defined to be, in part, a level of *confidence* on the part of the user to act on, or accept, the advice and decisions generated by the IDA and, in such cases where it is applicable, to allow the IDA to take action without intervention (for example, process control systems). HCT is also defined to be a *willingness* on the part of the user to act on the advice of the system. The user's willingness to act may result from their level of confidence in the system when they have sufficient evidence to make a judgement about the trustworthiness of the system. However, when sufficient evidence does not exist something more than simple confidence is at work. Shaw (1997, p.21) explains this as follows,

> "Confidence (alone) arises as a result of specific knowledge, it is built on reason and fact. In contrast, trust is based, in part, on faith. We sometimes give our trust in spite of evidence that might suggest we should feel some caution, if not outright suspicion, about relying on another."

Thus, the definition of trust suggested above encompasses yet distinguishes between the user's *confidence* in the system and the user's *willingness* to use the machine to perform the decision task as intended. The duality of this definition corresponds well with the trust framework presented in Table 2.1, in as much as *confidence* may be seen to be the primary outcome from the cognition-based component of HCT and *willingness* (Yamagishi 1986) may be seen to be an outcome of both the cognition-based and the affect-based components of HCT. The affect-based component necessarily plays a greater role in situations where the user has insufficient knowledge upon which to base a cognitive decision.

At this time, it might be helpful to clarify some terminology used in this study. Distinguishing between *trust* and *trustworthiness* is particularly important. It should be understood that the level of trust that the user has in a particular IDA is internal to the user and is based upon the individual's perceptions of the trustworthiness of the system. This level of trust perceived by a user for an IDA is separate and distinct from the true trustworthiness of the system, which results from intrinsic qualities of the system. The latter is external to the user. Furthermore, the level of trust that the user has in the system may not be congruent with the system's true trustworthiness (Will 1992).

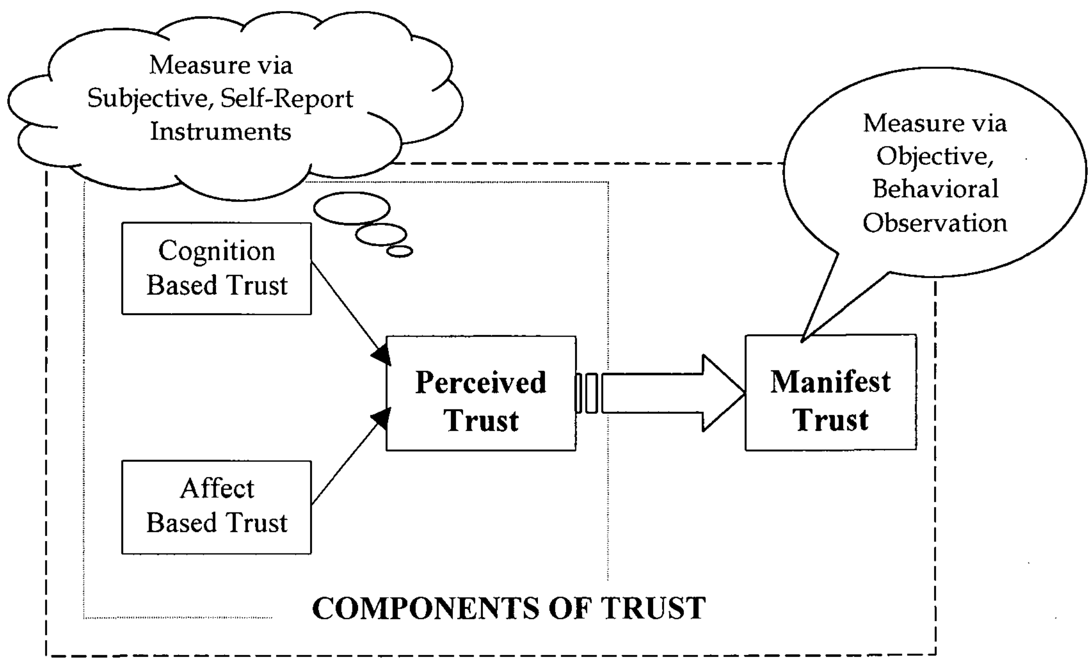
Factors that contribute to the overall levels of trust and trustworthiness may share common names and should not be confused during this discussion. For example, *predictability* is both a characteristic of the IDA, i.e. a dimension of trustworthiness, and an attribution made by the user based on their perception of the predictability of the IDA, which is a dimension of trust.

Predictability as a characteristic of the IDA exists at a specific level within the system which can be measured objectively as a function of the system's performance. Predictability as a dimension of the user's trust in the system exists at a level attributed to the system by the user and as such can only be measured subjectively via a self-report mechanism. This thesis focuses on the user's perceptions of trust and trustworthiness rather than the objective trustworthiness of the IDA. Therefore, dimensions of interest will be prefaced in this report by the term 'perceived' to indicate that these are characteristics attributed to the IDA by the user.

2.7 A conceptual model of human-computer trust components

It is proposed, based on the previous discussion, that human-computer trust, as a psychological construct, conforms to the nature of psychological constructs. That is, it encompasses both subjective and behavioral aspects of human experience (Bernstein et al. 1991). Manifest HCT, that is the users' actual use of the IDA will not be further considered here. Likewise, the transition from perceived trust to manifest trust, illustrated by the block arrow in Figure 2.1, is outside the scope of this study.

Figure 2.1 Human-computer trust components and measurement



Perceived HCT, which is the variable of interest in this study, is seen to be comprised of both cognitive and affective components as reflected in the definition

(McAllister 1995) and the trust framework (Table 2.1). These two components of perceived trust are intuitively consistent with our understanding of human nature. The premise that as humans we base our decisions on both what we know and what we feel is one which is widely accepted in the community.

2.7.1 Overall Perceived Trust

The overall perceived trust is the level of trust that a user has for a particular IDA being used in a particular decision-making activity. It is subjective in nature and as such can not be measured through any direct observation. Recall that the user must have some amount of knowledge about the decision task and the IDA in order to be able to make judgments about how much to trust the computer. With complete task knowledge the decision makers would have no need to trust IDA. They would know without question whether the machine advice was correct or not. With limited task knowledge, the decision maker may need to trust the IDA. Their level of trust is then based on their knowledge about the situation, the IDA and themselves, and their emotional responses to, or feelings about, the IDA and the situation.

2.7.2 Cognition-Based HCT

Cognition based HCT is based on evidence of trustworthiness of the intelligent systems under a particular set of conditions. Aspects of the IDA's behavior, or performance, such as consistency, reliability, competence, or dependability provide the user with the information with which to make an evaluation of the *trustworthiness* of the system (Lee and Moray 1994, Lerch et al. 1993, Muir 1994, Muir and Moray 1996, see Nominal Group Results in Table 3.1).

For example, a particular IDA provides consistently high quality, accurate responses across similar routine decision problems therefore the system is evaluated by the user to be trustworthy for all similar instances. On the other hand, this same system may fail to provide consistent responses for different or unusual problems. The user may then evaluate the system to be untrustworthy for all problems other than those which are routine. A specific level of cognition-based trust, therefore, is an outcome of experience with a specific IDA.

2.7.3 Affect-Based HCT

Affect-based trust is based on the user's feelings about the IDA. Aspects of the IDA such as the interactive style, its ease-of-use or the style of output used may cause the user to like or dislike using the IDA. The user's feelings about the computer also provide the user with the grounds upon which to make an evaluation of the *trustworthiness* of the IDA.

For example, an IDA may have been designed to provide the user with input screens that conform to a style with which the user is familiar therefore, the user is at ease using the system and feels that the system is trustworthy. On the other hand, an IDA which has a difficult to use or unfamiliar interface may cause the user to be frustrated or distressed while using the system. From this they may perceive the system to be untrustworthy. A specific level of affect-based trust, therefore, is also an outcome of experience with a specific IDA.

Affect-based trust has not been specifically included in any of the existing HCT studies. With the exception of the Faith variable (Muir 1994, Muir and Moray 1996, Lee and Moray 1994, Lerch et al. 1993), investigators concentrated on the cognitive perceptions of their participants. This is possibly because it seems unnatural to propose that a user has *feelings* for a particular computer they are using. But in general conversation it is not uncommon to hear someone say, "I hate computers!", or "I love this new program.". It may indeed seem unnatural, but it is not unreasonable, given that humans have been shown to attribute human characteristics to intelligent systems (Nass et al. 1993), to consider that a person's feelings about the computer system they are using influence the level of trust they have for the system.

2.8 Measurement Issues

Measuring manifest trust is perhaps more straight forward than measuring perceived trust. Trust is manifested as behaviors which can be observed. In the case of human-computer trust the user's trust in the system is manifest in the actual use of the system's advice as opposed to the use of the system. This type of measurement can be performed by observing and monitoring the way in which an intelligent decision aid is used over some pre-determined period of time under specific conditions.

Measuring perceived trust on the other hand is not as direct. One can not observe a user's perceived trust of an intelligent decision aid. The only way of learning about a user's perception of trust in an intelligent decision aid is to ask them. This is the reasoning behind the use of questionnaires and scaling techniques, such as Likert scaling in social and psychological research. It would be possible, for instance, in the case of this study, to ask user's to rate their level of trust in a particular system. However, given the complexity of human relationships and the vast number of different situations in which human-computer trust is likely to be a determinant of behavior, the concept of human-computer trust is unlikely to be sufficiently investigated via a single item questionnaire.

In order to gain a better and more complete understanding of human-computer trust it is necessary to investigate the essential elements which inlfuence or comprise human-computer trust in particular contexts. These elements are referred to as the underlying factors or dimensions of the overall construct of interest.

For example, one may expect that their doctor is a skilled diagnostician so a construct such as *technical competence* may be an essential underlying factor, or dimension, in trusting a physician. On the other hand one may not expect their physician to be loving toward them or indeed to be lovable, therefore love may not be an essential underlying factor in trusting a physician. Love, however, may be a particularly important underlying factor in trusting one's life partner.

Perceived trust, generally, and perceived human-computer trust in particular, may thus be seen to be based on several essentially important underlying factors or dimensions depending upon the context in which it develops. The HCT construct is therefore considered to be multi-dimensional. It is the purpose of this investigation to identify both the cognitive and affective dimensions involved in human-computer trust and to develop a psychometric instrument which may be used to measure these factors thereby providing a composite measure of overall perceived trust.

2.9 Empirical Research into Human-Computer Trust

There remains a paucity of research in the study of human-computer trust. This may in part due to the fact that a valid, reliable measurement instrument does not yet exist. The following review and discussion of existing research thus includes unpublished papers such as, "Measuring Trust in Machine Advice" (Lerch, Prietula and Kim 1993) and preliminary discussion papers such as, "Trust between humans and machines, and the design of decision aids," (Muir 1987). Table 2.2 summarizes these investigations.

Table 2.2 Summary of previous human-computer trust research

Researcher	Year	Instrument Used	Constructs Used	Scale Test / Sample Size	Methodology / Sample Size	Instrument Type
Muir	1987 and 1994	None	Predictability, Dependability, Faith, Persistence, Technical Competence and Fiduciary Responsibility	Not applicable	Not operationalized	None
Will	1991	Created for this study	Decision confidence	Validation/ 385	Experiment/12	Single item
Will	1992	Created for this study	Decision Confidence	Not applicable	Experiment/14	Single item
Lerch, Prietula and Kim	1993	Modified Remple, Holmes and Zanna 1993	Predictability, Dependability, Faith	Pilot Survey / 70	Experiment / 95 Experiment / 82	3 items per construct
Dassonville, Jolly & Desdot	1994	Created for this study	Reliability, Predictability, Performance	None	Experiment / Not Reported	1 item per construct
Lee & Moray	1994	Created for this study	Self-confidence and Trust in various aspects of life and the test system	None	Experiment / 12	1 item per construct
Muir & Moray	1996	Created for this study	Predictability, Faith, Technical Competence and Fiduciary Responsibility	None	2 Experiments / 6 in each	1 item per construct

Initially, investigators of human-computer trust turned to the existing interpersonal trust models as a starting point for their studies. These researchers have attempted to describe and explain the user's development of trust in an IDA by using adaptations of interpersonal measurement instruments or simply creating their own scales (Lee & Moray 1994, Lerch et al. 1993, Muir 1987, 1994, Muir & Moray 1996, Will 1991, 1992).

Numerous interpersonal models have been borrowed from both sociological and psychological studies. Unfortunately, there have been few confirmatory studies performed on the available interpersonal trust scales within either the existing body of interpersonal trust research or the existing body of human-computer trust research. Existing instruments, for the most part, multi-item Likert scales, have been used without confirmatory testing by continuing researchers or altogether ignored and replaced with single item scales.

Researchers in the field of human-computer trust have primarily focused on the interpersonal trust theories and models of Remple, Holmes and Zanna (1985) and Barber (1983) either directly or indirectly. The remainder of the studies have simply asked participants if they trust the particular systems that were used as test vehicles.

The following sections discuss the problems found in interpreting the results from these studies, as a basis for this research.

2.9.1 Definitions of Human-Computer Trust used in previous research

Lerch, Prietula and Kim (1993) adapted Rotter's (1980) definition of trust as, "a generalized expectancy held by the *user* that *information or advice generated by a machine* can be relied on"[6]. By adopting this definition the authors have accepted an expectancy perspective for their research. However, at no time do they address this issue by trying to measure the expectations held by their participants. As will be seen later in this discussion, they in fact measure variables which relate to their participants' experience. Any differences between participants' expectations about the machine and their experiences with using the system are simply ignored. It seems incongruous to define Human-Computer Trust as an expectation, necessarily a pretest condition, and then measure it as an experience, necessarily a post-test condition.

Muir (1987, p.528) considers several definitions of trust including both Rotter and Barber's definitions as discussed above in section 2.1.1. Barber's (1983) definitions in, *The Logic and Limits of Trust,* are developed with respect to an individual's trust in different actors such as authority figures, government and society as a whole. Muir's decision to incorporate Barber's definition into her own

---

[6] italcs added and indicate the substitutions made to accommodate the machine as the trustee

was based on the fact that this definition was one which expressed the multi-dimensional nature of trust. The difficulty with this decision is that the author has confused the problem of defining trust with the problem of identifying the underlying dimensions of the concept of HCT. In any case, Muir arrived at a very interesting matrix of trust dimensions by crossing Barber's expectation model with Rempel, Holmes and Zanna's (1985) dimensions of trust from their study of close personal relationships. This matrix is reproduced in Figure 2.2.

Figure 2.2 Experience-Expectation Matrix

Taken from Muir (1987, p.529)

	Basis of expectation at different levels of experience		
Expectation	**Predictability** (of acts)	**Dependability** (of disposition)	**Faith** (in motives)
Persistence			
Natural Physical	Events conform to natural laws	Nature is lawful	Natural laws are constant
Natural Biological	Human life has survived	Human survival is lawful	Human life will survive
Moral social	Humans and computers act "decently".	Humans and computers are good and "decent" by nature	Human and computers will continue to be "good" and "decent" in the future
Technical competence	*j*'s behavior is predictable	*j* has a dependable nature	*j* will continue to be dependable in the future
Fiduciary responsibility	*j*'s behavior is consistently responsible	*j* has a responsible nature	*j* will continue to be responsible in the future

This led her to define trust as, "Trust (T) is the expectation (E), held by a member (*i*) of a system, of persistence (P) of the natural (*n*) and moral social (m) orders, and of technically competent performance (TCP), and of fiduciary responsibility (FR), from a member (*j*) of the system, and is related to, but not necessarily isomorphic with, objective measures of these qualities. Or alternatively, $T = [E_{i(\mathrm{Pn}+\mathrm{Pm})}] + [E_{i\mathrm{TCP}i}] + [E_{i\mathrm{FR}j}]$" (Muir 1987, p.531). It is clear that this definition is complicated by the author's attempt to include the underlying dimensions of trust in the definition. The inclusion of the dimensions into the definition assumes that these dimensions are the only ones which exist in the Human-Computer relationship. This assumption is unlikely to withstand scrutiny since it is clear from interpersonal trust research that as the context changes so do the elements which most affect the user's perception of the machine's trustworthiness. In fact, Lerch et al. (1993) were able to demonstrate this in their HCT work. In a later study conducted by Muir and Moray (1996) it was found that only the experience dimensions in Figure 2.3, were correlated with the users' trust in the systems they used.

The Muir matrix was adopted by other researchers such as Lee and Moray (1994) and Dassonville, Jolly and Desodt (1994) with little further consideration or testing. The Expectation versus Experience matrix has the appearance of a neat, concise, mathematically definable model of trust and as such it has proven to be appealing to researchers from a science and technology background. Unfortunately, it is not based on sound psycho-social procedures for defining concepts and identifying underlying variables with no effort made to "ground" the dimensions in reality.

2.9.2 Psychometric Instrument Selection

Lerch et al. (1993) studied the three dimensions of trust as first described by Remple, Holmes and Zanna (1985): predictability, dependability and faith.. The first two of these dimensions are characteristics attributed to the trustee by the trustor based on the latter's perception of the former. The third of these, faith, is a characteristic of the trustor alone and functions under circumstances in which the trustor has a need to trust and does not have sufficient information to make a cognitive judgment about the trustworthiness of the trustee. Perceived trust, in their study, is based upon the predictability of the other person's behavior, their dependability and later in the relationship faith that the other person will continue to behave and respond in a consistent manner. Lerch et al. adapted the Rempel, Holmes and Zanna scales to computer use by simply changing the wording of the questions to suit a computer system.

The scales were tested in a pilot study and items which did not load on their respective factors were dropped from the scales. This resulted in a scale of ten items from the original 26 item scale originally developed by Rempel et al. (1985). Items were unevenly distributed among the three factors of faith, reliability and dependability. The number of participants in this study were sufficiently large to support quantitative analysis. Their findings however, could not be considered to be conclusive, in terms of the appropriateness of the dimensions selected, since these dimensions were simply adopted from an existing interpersonal trust instrument. Thus, the content validity and reliability of this scale remains in question.

Muir and Moray (1996) published the scale that was used in Muir's original study of human-computer trust. The scale consists of nine items created specifically for their study. In this scale there is one item for faith, there are three items directly asking the participant to rate their "degree of trust" in the pump, the display and the system overall. This leaves five items to be distributed among the six factors left from the original 3x3 trust matrix in Figure 2.3. These five items included two items for competence, two for responsibility and one for predictability. The dimensions of persistence and dependability were not investigated. Even if sample numbers used in this study had been large enough to support quantitative analysis the validity of these scales would be difficult to establish since there are so few questions for each construct with some constructs missing and existing questions unevenly distributed among the remaining constructs.

Lee and Moray (1994) investigated the correlation between the user's perceived trust in a system and their self-confidence. In spite of the fact that these researchers worked with Muir they chose to create their own set of questions to measure trust. These questions can be found in Appendix A. In addition, they measured the user's self-confidence having determined that the user's self-confidence and their trust in the system are negatively correlated mediators of the user's behavior while operating the system. Once again, their scale was limited to direct questions about trust and particular aspects of the process control system they were using. This scale had three items for trust and three items for self-confidence. Once again, this scale is not one which could be shown to be valid nor is it one which could be useful outside of the study for which it was created.

Dassonville et al. (1994) based their investigation on the work of Rempel et al. (1985), Barber (1983) and Muir (1987). Dassonville et al. developed their own trust scale specific to their research needs. Items were as follows: Is the joystick reliable?; Does the joystick have a high performance?; Is the joystick's behavior predictable? (Dassonville et al. 1994, p.199). Considered in conjunction with these specific trust items were measures for user's a priori self-confidence, trust in others and trust in machines in general.

In terms of the current study this research is a good example of the problems that have been discussed above. The researchers, finding no existing scale of measurement, simply created their own. Unfortunately, the scale they devised was too simple to be of any real value to an investigation of human-computer trust. They investigated only two of the many parameters in the literature they cited, which inlcuded Rempel et al. (1985), Barber (1983) and Muir (1987). These two variables were *reliability* and *predictability*.

In his study *True and False Dependence on technology: evaluation with an expert system,* Will (1992) investigated differences between experts and novices using the same expert system. He didn't measure human computer trust, per se, but rather measured the user's decision confidence "as an indicator of whether the subject was confident that the problem was solved correctly," (Will, 1992, p.176). As a measurement for decision confidence participants were asked to rate their confidence "from no confidence (zero) to very high confidence (100)", (Will, 1992, p.177).

Once again the researcher relied on a single item measure for decision confidence. This may have been sufficient for this investigation as the researcher was studying several other parameters as well, however, it is not sufficient to elucidate the dimensions that underlie the user's perception of trust in the system.

While there was little value in the measurement scale, this study was of particular interest as it was designed to use a system that gave incorrect decisions and the results suggested that neither task experts nor novices were able to tell that this system was false. Furthermore, the experts displayed higher confidence in the wrong decisions that they made using the decision aid than did the novice users. This study did little however to further our knowledge about how and why these levels of trust were observed.

The other variables that were considered included state anxiety, dogmatism, and system success. It is possible that the measure of decision confidence was confounded by other variables such as the self-confidence of the participants. The experts participating were reported to have stated that the system was of no particular use and that they could solve the problem without the aid of an expert system. This anomaly was not further investigated by the researchers.

2.9.3 Methodologies and Approaches

Most HCT studies have concentrated on measuring subjective or perceived trust of their participants under experimental conditions, and most investigators have not specifically differentiated between perceived and manifest trust.

Previous investigators studied the development of trust with users who had, at most, three separate experiences with the system. At best these results could only be considered to be preliminary in nature since it is widely agreed that trust takes considerable time to develop. It is unlikely that after only two or three uses of a particular system under experimental conditions, where little risk exists in the event of incorrect decisions, the user would have had time to really evaluate the factors such as the competence or the dependability of the systems. They also studied such things as the system characteristics and personal characteristics of the users without first constructing a valid, reliable measure for the dependant variable they were investigating, human-computer trust.

With the exception of Lerch et al. (1993), all other investigators used small numbers of participants precluding any quantitative analysis of the results. Thus their results can not be generalized to other samples nor relied on by other researchers.

A valid and reliable psychometric instrument specifically designed for human-computer trust is thus yet to be found or used by any previous researcher in this field.

2.10 Conclusion

The practices of borrowing existing scales designed specifically to measure interpersonal trust rather than human-computer trust and creating new scales in an ad hoc fashion have led to several identifiable problems.

First, there is little agreement among researchers as to the nature and structure of human-computer trust. Some treat trust as uni-dimensional, others believe that it is multi-dimensional. The experience model was found to be better able to predict user's manifesr trust in intelligent systems than expectation model (Muir 1996). But no attempt is made to establish why this might be the case.

Second, there remains an inconsistency between the definitions of trust chosen and the type of trust being investigated. The ad hoc style of the existing human-computer trust research means that there is no comprehensive examination of interpersonal trust studies to establish their applicability to human-computer trust research.

Third, confirmatory analysis of borrowed scales does not exist in human-computer trust studies and HCT researchers fail to provide evidence to support their selection of measurement instrument and creation of or selection of scale items.

Finally, there has been no field work done in human-computer trust research to date. Existing HCT data has been collected via experimental methods and is related to simulated or prototype systems rather than operational systems.

The inconsistency in both method and measurement among existing human-computer trust studies and the absence of a measurement instrument designed specifically for human-computer trust motivated the current research.

Chapter 3

Instrument Development

"... thanks to trust one avoids having to take account of some possibilities and is able to embrace some action, which without trust would be impossible," (Misztal 1998).

In Chapter Two trust in general and human-computer trust in particular are defined in terms of a trust framework which encompasses both the sociological and psychological perspectives of trust. A new model for human-computer trust was developed based on this definition and framework. In addition, existing studies into the nature and dynamics of human-computer trust were reviewed. Many of these existing studies were found to have failed to address some of the most fundamental issues of instrument development.

This chapter reports the development of a new psychometric measurement instrument for HCT. There are several criterion of quality with which psychometric instrument development practitioners are concerned. Primarily these are the equivalence reliability of the instrument, does the instrument provide an internally consistent measure of the construct; the content validity of the instrument, does it measure representatively from the universe of interest; and the construct validity of the instrument, does it measure the construct it is intended to measure and nothing else. Assessment of these criterion for a particular instrument can be accomplished in several ways (Cronbach 1970, Straub 1989). This study was designed to address these criteria in a logical, rigorous progression through the various stages of the study.

This chapter first details the stages of instrument creation, which were the identification of the latent variables underlying both the cognitive and affect-based components of perceived human-computer trust and the creation and selection of items for each of the constructs identified. These intial stages specifically address the issue of content validity. Then details of the instrument refinement stages are reported. The initial draft of the scales underwent a process of refinement through which both constructs and items were eliminated so that the final scale was as succinct as possible while having improved internal consistency, construct validity and content validity. The complete survey instrument was then created with the addition of information for the participants, instructions and demographic items. The results of a trial with this instrument are then presented.

3.1 Stage 1: Identification of factors via the Nominal Group Technique

The purpose of this stage is to identify the factors involved in human-computer trust from the user's perspective. In the model presented in the previous chapter the overall perceived trust is described as a latent variable comprised of two other latent variables, cognition-based and affect-based trust which in turn are comprised of several underlying variables. It is the underlying variables which are able to be measured through subjective rating scales because the respondent can identify how strongly they believe a particular system has specific characteristics and how they feel about the system (Lerch et al. 1993, Muir 1987, Nass et al. 1993, Neuman 1994).

In order to measure attitudes of this nature it is customary for researchers to design and test psychometric instruments which allow individuals to self-report their personal attitudes. The construction of such instruments can be a rigorous undertaking although this has not necessarily been the case in previous HCT research. A rigorous approach to instrument construction involves several steps. The first step is to identify and define the concepts of interest as clearly as possible.

3.1.1 Participants

The participants in this stage of the study were four members of the general public who had worked or were currently working with computers. The members of the group all had an interest in and experience with various types of intelligent systems from grammar checkers to process control systems.

3.1.2 Materials

The materials needed to conduct a Nominal Group are: tape recording equipment, whiteboard; Butchers paper; and marking pens. The Information Sheet, and the Consent Form that were used can be found in Appendix B.

3.1.3 Design

The nominal group technique (NGT) was chosen as the method by which this exploratory Stage would be accomplished. The nominal group technique is a structured discussion technique developed to be used to assist small groups, generally six to twelve participants, in decision making. It allows for participants to have both individual and group input into the final outcomes within a manageable amount of time. Each part of the process is time limited which keeps the group moving toward a final outcome. The technique was chosen here because it is an effective, widely used process which is adaptable to the requirements and resources of particular studies without compromise to its effectiveness (Delbecq, Van De Ven & Gustafson 1975).

3.1.4 Procedure

Participants were given an Information Sheet, a consent form and a brief introductory talk about the research during which they were invited to ask questions. The group was then given an overview of the procedure that would be followed in the session and again invited to ask questions for clarification. The group was then provided with two scenarios of interpersonal trust in order to set the context of the discussion. Finally the group was asked to consider their own experience using a computer system and what aspects of this experience would most likely lead to their trust in the computer system. Details of the above instructions can be found in Appendix B. The researcher facilitated the discussion process and the research supervisor observed and took notes.

The group was then guided through the following Stages of the NGT (Delbecq et al. 1975):

1. Brainstorming – individual generation of ideas
2. Round robin listing of ideas and group discussion and clarification of ideas. Participants were also asked to define their items at this time.
3. Selection of the seven most important items from the complete list by the group[7]
4. Individual ranking or rating of ideas from most important to least important, with two or more items of equal importance rated equally.
5. Group discussion of individual ratings until consensus resulting in a rated list of items agreed to by all participants.

3.1.5 Results

Table 3.1 presents the results from the nominal group process. The constructs are identified by the names and definitions given to them by the group and have been placed in rank order according to the rank of importance indicated by the group. The

[7] In the case of this nominal group ten items comprised the entire original list from the Round Robin and the group chose to include all in the rating step.

second column indicates whether the construct is most likely associated with cognition-based or affect-based trust and has been added by the author.

Table 3.1 Nominal Group Results

Construct	Trust Type	Definition	Rank
Discriminating	C	the system can discriminate between different degrees of information correctly and accurately records it	10
Reproducibility of results	C	the system outputs the same result given the same input	10
Feedback	C	meaning that messages and results are meaningful and understandable	10
Conciseness of display	C	such that as much information is displayed as concisely as possible	9
Rcliability	C	the system performs repeatedly and reproduces decisions from one time to another	8
Output of the system	C	The output of the system is correct and accurate.	8
Ability to be edited	C	in the sense that the input can be changed when errors occur or new information is known.	8
Friendliness	A	In the sense of user friendliness such that the system is natural and instinctive to use.	7
Ease of use	A	in the usual sense of the ease with which the instructions and functions can be learned and used effectively	6
Speed of the system	C	the system performs within a time frame that is acceptable to the task and situation in which it is being used	5

C - Cognition-based trust
A - Affect-based trust

3.1.6 Discussion

The participants in this nominal group were highly motivated to participate in the research and demonstrated a great deal of enthusiasm for well designed computer systems. It is clear from the list that was generated by this group that the most highly regarded systems were those that performed correctly, accurately, reliably, were relatively easy to use and allowed the user reasonable control over both inputs and outputs.

Items such as ease-of-use and speed were ranked lower than others because it was believed by the group that if the system performed its job particularly well then the user could adapt to some extent and learn to live with a system that was a bit slower or a bit harder to use than one would like. Clearly though it was very important for this group to know what the system was doing as feedback from the system was one of their most highly ranked requirements.

Participants during the final steps, in the nominal group activity, are usually asked to reduce their list, with group consensus, to the 5 to 7 most important concepts. This group, however, had some difficulty discarding any of these items. After a few attempts to get the group to reduce the list, they were allowed to keep and rank all ten of the items they had generated. This decision was made because there were only ten and this was the first step in a rigorous process of development and refinement. Had this been a final stage in the research the group would have been encouraged more strongly to discard the least important items so that the ranking might have been more stringent. In this case however, it is believed that the results from the group would not necessarily have been improved by forcing them to discard items.

The ten constructs produced by the group were found to be consistent with constructs from previous research[8] in spite of the fact that there was only one group of four participants. It was decided that these results combined with those from previous human-computer trust research would provide an adequate starting point for the next stage.

3.2 Stage 2: Selection of factors and scale items

The second stage developed questions to elicit responses from participants appropriate to the constructs being investigated. For example, if one is asking about

[8] In research such as this it is the underlying concept and not the name given to it by various people which is important. At this stage in information systems research very few concept names with the exception of the term "user-friendly" have been widely agreed upon and accepted by the research community.

trust one could simply ask a participant how much they trust the system. This may provide knowledge of the participant's level of trust but it does not provide any indication of why or how this level was achieved. More importantly, it may not be accurate if the respondent does not interpret the meaning of trust in the same way as the investigator. To obtain richer information it is necessary to ask a series of questions rather than just one which help to eliminate some of the error of interpretation (Cohen et al. 1990).

The selection of the initial set of factors to be investigated consisted of several steps. The first was to aggregate all the available constructs both from the nominal group and from previous literature including both cognitive constructs and affective constructs. Constructs from sources other than the Nominal Group were first culled on the basis of whether or not they were within the scope of this study. This produced a preliminary set of constructs which was as broad as possible while remaining within the scope of the current study.

For example, Sheridan's constructs *usefulness* and *dependence* were thought to be outside the scope of this study as they related to task performance rather than specifically to the system's perceived trustworthiness. Constructs which were too close in meaning to be easily discriminated were either merged into single constructs or dropped. Constructs such as *liking*, taking pleasure in using the computer system and finding it agreeable to one's taste, and *loving*, a partiality or preference for using the system, after inspection by two independent judges were considered to be too difficult to distinguish from each other. Since the discriminant validity was thus in doubt the constructs were combined into one construct labeled, *personal attachment*.

One other concept emerged after consultation with colleagues which appeared to be of particular importance. This is the *integrity* of the system meaning that the system is able to recover from technical failures or user errors without loss of data in the same way that a word processor saves a back up copy of an open document when system errors occur.

These two stages of reduction resulted in the following nine constructs:

1. ***Reliability*** of the system, in the usual sense of repeated, consistent functioning.
2. ***Robustness*** of the system, meaning demonstrated or promised ability to perform under a variety of circumstances.
3. ***Familiarity***, that is the system employs procedures, terms, and cultural norms which are familiar, friendly and natural to the trusting person.
4. ***Understandability*** in the sense that the human supervisor or observer can form a mental model and predict future system behavior.
5. ***Explication of intention***, meaning the system explicitly displays or says that it will act in a particular way (as contrasted to its future action having to be predicted from a model).
6. ***Technical Competence*** of the system meaning that the system is perceived to perform the tasks accurately and correctly based on the information that is input.
7. ***Integrity*** of the system in the sense that the system is able to recover from technical failures or user errors without loss of data.
8. ***Personal Attachment*** to the system comprised of: *liking* meaning that the user finds using the system agreeable and it suits their taste and *loving* meaning that the user has a strong preference for the system, is partial to using it and has an attachment to it.
9. ***Faith*** meaning that the user has faith in the future ability of the system to perform even in situations in which it is untried.

3.2.1.1 Assigning items to each construct.

Questions which would elicit responses from participants about the various constructs, had to be either selected from previous interpersonal and human-computer trust research, or created with the aim of developing at least five questions for each construct. The resultant groups of items, each expected to measure one construct, would become the preliminary scales of the new HCT psychometric instrument. This preliminary set of scales consisting of the nine constructs above and 75 items, as yet unevenly distributed among the constructs, was then introduced into stage three for reduction and refinement.

The preliminary set of scales was kept as large as possible to ensure that as many aspects of each construct were considered as possible. It was thought that it would be unwise to restrict the pool of items prematurely. As is discussed in the following section, this made the task of sorting items into construct groups quite challenging for the first group of judges in particular.

3.3 Stage 3: Reduction and Refinement of HCT items via Thurstone Scaling.

The reduction of the number of constructs and their related items was undertaken because in survey research where the concept being measured is one of human perception or opinion it is better to have as few constructs as possible, without losing the validity of the measure. The aim is to limit the instrument to those constructs and items which are the most easily discriminated from each other and most convergent on the concept being studied with three to five items per construct (Neuman 1994).

The number of scales in the instrument must be reduced to those which can be easily distinguished from each other and the number of items must be reduced to those which best fit each construct. Because the type of questionnaire developed in this study is a summative scale, it is advisable to keep the number of items per construct consistent so that the weighting of each scale in the overall measure is approximately equal. Therefore, while one may begin with a large pool of items and numerous constructs from which to choose, the aim of the reduction stage is to refine the scales to their most parsimonious while maintaining the construct validity of the instrument.

This reduction process was performed via a series of card-sorting exercises similar in design to the Thurstone Scaling technique (Neuman 1994) and used by Moore and Benbasat (1991) in their study, "The Development of a Measure for the Adoption of Technology".

3.3.1 Participants

The judging panels were designed to include people from various backgrounds with varying degrees of knowledge of computing because the measurement instrument being developed is to be used in the field with people of varying ages, backgrounds and levels of computer experience and knowledge. It was necessary to refine the scales so that the questions were easily understood by people from these various demographic groups. A possible difficulty in the development of an instrument such as this for HCT is that the questions can become so specific to a particular sample of respondents that the instrument is not useful outside that sample for which it was designed. There were four judging panels with each panel comprised of at least four volunteers.

3.3.2 Materials

The materials used here were cards with one item printed on each. For the groups with the construct definitions cards, the names of the constructs and corresponding definitions were also provided. Participants also received information sheets and consent forms prior to participating and were given the opportunity to ask questions about the research or the process that they were about to undergo. These items can be found in Appendix C.5.

3.3.3 Design

During this stage of development the set of scales becomes more and more reduced and refined as it passes through each judging panel to the next. Individual judges were asked to sort the items into groups and to discard any items which are ambiguous or did not seem to belong to any one of their groups. An example of the results from this process is provided in Appendix C.

Each group of items created by the judges is considered to describe a single construct to which all the items in the group are related. In two of the panels, judges were asked to sort items into the already defined constructs. In the other two panels, judges were asked to create their own constructs based on their groupings of items. Judges in these latter cases were also asked to name and define the groupings they had created.

Once the sorting process was completed, judges were able to discuss their choices with each other. However their original groupings were those which were analyzed and upon which reductions and refinement of items and constructs were made. The inter-rater reliability of the judges' item placements was calculated as Cohen's Kappa.

It is recommended that there be at least three to five items per construct and that each construct has the same number of items (Neuman 1994, Lietz 1998). In order to ensure that this is the case in this study new items were created when original items failed to survive the sorting process.

3.3.4 Procedure

The following steps were followed with each judging panel:

1. Judges were given an information sheet, consent form and demographic survey to complete before sorting began.
2. The trial sorting exercise was performed.
   a. Judges were provided with either the constructs and the items or just the items for the trial sorting depending on which group they were participating in.
   b. Judges were asked to sort the trial items and encouraged to clarify the task as they went. They were also encouraged to discuss their solutions with the other judges to satisfy themselves that they knew what was expected of them for the actual task of sorting items into groups.
3. The actual sorting exercise was performed.
   a. Judges were then provided with either the constructs and the items or just the items for the actual sorting depending on the group in which they were participating.
   b. They were then asked to complete the sorting individually. At the completion of the sorting they were invited to discuss their solutions with the other judges on the panel. Their original individual solutions however were those that were analyzed.

3.3.5 Results

3.3.5.1 Sorting Round One

This panel was given the set of constructs and items as defined in stage one. Judges were asked to sort the items into these constructs or discard any items which did not fit or were ambiguous. Judges were also asked to make comment on the wording of questions.

This first judging panel had 74 items and 9 constructs to consider and found that it was quite difficult to sort this number of items into groups. One of the judges in later discussion revealed that they had sorted simply on the wording of the items. This particular judge proceeded easily through the task while the others appeared to give greater consideration to the meaning of the questions and had more difficulty completing the task. The two approaches while perhaps altering the ease of the task, did not appear to make a noticeable difference to the quality of the resultant placements.

The results of this first sort, calculated as the "hit ratio" among all the judges, are shown in Figure 3.1. The hit ratio, unlike Cohen's kappa, does not account for the probability that corresponding placements will be made by chance. The hit ratio is useful during the development process because it provides the overall percentage of correctly placed items from all judges in a single matrix. Cohen's kappa requires a separate matrix for each pair of judges (see Figure 3.2). The hit ratio method, therefore, facilitates the refinement process.

Figure 3.1 Sorting Round One Hit Ratio

	OBSERVED PLACEMENT											
TARGET CATEGORY	**Re**	**Ro**	**U**	**E**	**Fm**	**T**	**I**	**P**	**F**	**D**	**Total Items Placed**	**% Items Correctly Placed**
Reliability	15	5	1	2		2			5	2	32	47
Robustness	5	5				5			10	3	28	18
Understandability	4	2	8	2	2				1	1	20	40
Explication of Intention			4	9	1	1			4	1	20	45
Familiarity	4		4	3	18			2		1	32	56
Technical Competence	15	12	2			17			11	7	64	27
Integrity	1		1	1		1	10		4	2	20	50
Personal Attachment	2		1			2		39	2	8	54	72
Faith	3		1		2	1		1	21	1	30	70
Total Item Placement: 300						**Total Hits: 142**				**Overall Hit Ratio: 47%**		

D - Discarded items

Figure 3.2 Round One Inter-Rater Reliabilities for Judges 1 & 2

	OBSERVED PLACEMENT											
TARGET CATEGORY	**Re**	**Ro**	**U**	**E**	**Fm**	**T**	**I**	**P**	**F**	**D**	**Row Total**	**q**
Reliability	7	3	1	1		2			2		16	2.45
Robustness	2	2				3			5	2	14	2.15
Understandability	3		2	1	2				1	1	10	0.67
Explication of Intention			3	4	1				1	1	10	0.67
Familiarity	2				12	1				1	16	1.81
Technical Competence	7	5	2			9			4	5	32	3.63
Integrity							6		2	2	10	0.40
Personal Attachment	2		1		2	1		14		8	28	2.80
Faith			1			1		1	10	1	14	2.33
Column Totals =	23	10	10	6	17	17	6	15	25	21	150	16.91

Diagonal totals (d) = 66 **N = 150**
q = n(row) x n(col)/N **$\kappa = (d-q)/(N-q) = 0.37^{\dagger}$**

The inter-rater reliability in this round was poor with a hit ratio of 47% and a kappa of 0.37. These results indicated that the items were neither discriminant nor convergent on their constructs. The raw placement results were used as the basis for reducing the items and constructs into a smaller and hopefully better set.

As seen in Figure 3.3 the *robustness* construct was effectively removed with items placed into other categories depending upon the initial judges' placements or dropped from the instrument.

Figure 3.3 Sorting Round One Hit Ratio - Poorly Placed Items Deleted

	OBSERVED PLACEMENT											
TARGET CATEGORY	Re	Ro	U	E	Fm	T	I	P	F	D	Total Items Placed	% Items Correctly Placed
Reliability	20	2	1						1		24	83
Robustness											0	0
Understandability	2		10	2	2				2	2	20	50
Explication of Intention				7	1						8	88
Familiarity			2	2	15			1			20	75
Technical Competence	1	4	1			12				2	20	60
Integrity	1					1	9			1	12	75
Personal Attachment	1				2			33		4	40	83
Faith									54		54	100
Total Item Placement: 198						Hits: 160				Overall Hit Ratio: 81%		

D - Discarded items

Decisions to delete items and constructs were made based on the consistency among the judges' placements for each item. For example, refer to Item Placement Table for *robustness* in Appendix C.5. Items placed as expected would havea 2 in the judges placement cells for that item. Item 2.i in this table was misplaced by all judges with two placing the item into the Reliability (1) construct and two placing this item into *technical competence* (6). Item 2.vii on the other hand, was consistently placed into the *faith* (9) category by all judges.

Since only one item in this set had more than one correct placement, this construct was deleted. Its items were either distributed to other categories or dropped. For example, item 2.vii was placed into the *faith* construct. Item 2.v was placed into the *technical competence* construct. The rest of the items in this set were eliminated, since there was little consensus among judges.

With the reduction completed the modified and reduced placements were analyzed again as a check on the refinements made and the inter-rater agreement increased to 81% (Figure 3.3). This improved hit ratio is meaningful only in terms of an interim indication that the refinements applied to the scales may have improved them to some extent. It can not be used as an independent measure of the content validity or reliability of the scales. The refined set of scales, now consisting of 43 items and 7 constructs, was input to the next sorting round.

3.3.5.2 Sorting Round Two

The second judging panel received the 43 items without the constructs and were asked to create their own groups. The judges were not told how many groups there were originally. The judges in this group all worked in the computing field and it was thought that the sorting task would be easier for them than for the other panels. However, it seemed as though the sorting task was harder for these computing professionals. This is perhaps because they brought with them preconceived ideas about the concepts that might be measured in a study such as this and tried to fit the constructs to their a priori ideas rather than allowing their understanding of the items, as they first read them, to form new constructs. This group had great difficulty completing the task.

All of the judges expressed concern that they would be making incorrect placements and although the trial sort alleviated some of their misgivings they continued to express concerns about the "correctness" of their results. One of the judges also expressed concerns about the nature of the study. This judge indicated that in their opinion and based on Actor Network Theory[9] it was inappropriate to suggest that users attribute human characteristics to computer systems. It was obvious that this judge was uncomfortable with the wording of the questions. At this point it was restated to the panel that their participation was completely voluntary and that if they would like to withdraw at any time they could. The judge having difficulty chose to stay and persevered with the task to completion.

Again the results from this panel on the reduced set of scales was below the target of 80% inter-rater agreement and so further reduction and refinement was performed. The results of this round demonstrated that the constructs of *familiarity* and *integrity* were not easily discriminated by the judges from other constructs. Items expected to converge on *familiarity* were spread across *personal attachment* and *understandability*. Items expected to converge on *integrity* were spread across *reliability, personal attachment and faith*. Given the difficulty that judges had with these constructs they were discarded. Some of the items from the deleted constructs were reassigned to other constructs if there was good agreement among the judges' placements. The remaining items from the deleted constructs were discarded.

At this point the number of items for each construct was equalized because the overall measure is an additive scale. It is, thus, preferable to have an equal number of items for each construct so that there is an equal representation of each construct in the final measure. Cohen et al. (1990) recommended having at least three items per construct. Since the purpose of this study was to refine the scales, it

---

[9] Actor Network Theory (ANT) would suggest that there should be no disctinction made between the human and the non human actors in the decision-making network. This theory however, is relatively new and precludes the study of social attributions and ethopeoia upon which this current research is based. Whether the ANT perspective would add to the understanding of IDA use is unknown at this time and outside the scope of this study (refs).

was decided to maintain a set of five items for each construct with the expectation that at least three of these items would survive the final factor analytic stage.

During the remainder of this current stage, however, those items which performed best in each round were kept and for those constructs in which less than five items remained, new items were created. Some items were simply reworded where it was felt that verbal cues such as "I think" or "I feel" were interfering with the interpretation of the item.

This sorting round resulted in a set of five constructs with five items each. This set of scales was given to the third judging panel.

3.3.5.3 Sorting Round Three

The third judging panel again received the items without the constructs. It was felt that greater knowledge of the items requiring adjustment was to be gained by having the panel perform the sort without the constructs than with them since good discriminant validity should result in the expected number of constructs being created by the judges. This indeed was the case with the trial items.

The panel performed well with the trial sort and were confident to proceed with the actual sorting task. At this time the panel still had more difficulty with the actual task than with the trial task although there was less discrepancy between the number of constructs expected and the number of constructs created by the judges. Again the reduction and refinement process was undertaken with new items created to replace deleted ones in order to maintain five items per construct. New items were created to be consistent with the items that placed well. The resultant set of five constructs was used for the fourth and final judging panel.

3.3.5.4 Sorting Round Four

The fourth judging panel was provided with both the items and the constructs along with their definitions. This panel also experienced the trial sort and once again most judges achieved consistent results with only one or two items being misplaced. The actual sorting task was much easier for this group because of the reduced number of items they had to sort and the fact that they had the constructs provided. The inter-rater reliability of this group was above the 80% target (Moore & Benbasat 1991). Some minor changes in wording were then made to those few items that were misplaced. The resultant set of scales was then formatted into a survey using a seven point Likert scale, Appendix D, and introduced into the fourth stage for further testing.

The inter-rater reliability results for this and the preceding rounds are summarized in Figure 3.4.

Figure 3.4 Inter-Rater Reliabilities from all sorting rounds

Inter-Rater Reliability							
Round 1		**Round 2***		**Round 3***		**Round 4**	
Judges	κ	Judges	κ	Judges	κ	Judges	κ
1,2	† *0.37*	1,2	0.26	1,2	0.55	1,2	0.88
1,3	0.43	1,3	0.33	1,3	0.43	1,3	0.78
1,4	0.43	1,4	0.26	1,4	0.35	1,4	0.90
2,3	0.37	1,5	0.25	2,3	0.53	2,3	0.76
2,4	0.38	2,3	0.39	2,4	0.45	2,4	0.88
3,4	0.44	2,4	0.27	3,4	0.33	3,4	0.78
		2,5	0.31				
		3,4	0.37				
		3,5	0.39				
		4,5	0.31				
Average	**0.40**	**Average**	**0.32**	**Average**	**0.44**	**Average**	**0.83**

***Judges created their own constructs in these rounds**
†Refer to Table 3.2

3.3.6 Discussion

All groups were given a trial at the sorting procedure. The constructs and their corresponding items for the trial were taken from Moore and Benbasat (1991) and are included in Appendix C.4. This process gave those who felt that they didn't know enough about computers an opportunity to compare their sorting abilities with those who did have computer experience. Aside from this unexpected benefit, the purpose of the trial sort was to give all participants the opportunity to ask questions about the process and to become familiar with the sorting task before beginning.

All participants indicated that the trial sort was very valuable to them and that they would not have felt confident without it. For those who doubted their ability because of their lack of computing knowledge, the trial demonstrated clearly that computing knowledge was not a prerequisite to having the ability to sort **like** concepts into named groups whether those concepts are computing related or not. All groups of judges were given the opportunity to discuss their results with each other before concluding the judging session. However, none of the judges felt that it was necessary to do this once they had decided on their placements.

The four judging panels were treated in two distinct ways. Two of the groups had the constructs and their definitions and the other two groups had only the items. This process resulted in some differences in the way in which the judges performed their task and in the way the results were prepared for the inter-rater reliability calculations.

3.3.6.1 Groups with constructs

Being given the research constructs and their definition in some ways simplified the sorting as these judges did not have to create, name and define their own categories.

A problem that became apparent with this method was that the judges, suspecting that all the items had originally been placed in one of the categories tried to force items to fit one of the categories even if they were unsure. Most judges were indeed reluctant to discard any items. None-the-less, misplaced rather than discarded items in most cases were dropped from the set after inter-rater reliabilities had been calculated.

This misplacement of items could account for the low reliability coefficients found for these groups. Dropped items were either reworded for clarity or replaced with newly created items depending on how these items had been misplaced with respect to the expected placements. Replacement of dropped items was necessary as the minimum number of items per construct was held at five with a view to having no less than three items per construct after the final quantitative validation stage of this study.

3.3.6.2 Groups without constructs

The two groups without constructs were provided with blank paper and pens on which they could name and define the categories that they created for their item groupings. These judges seemed to more easily group items once they had decided on the categories they would use. All judges performed well in the trial sort and it could be expected from this that given well constructed scales they would perform equally as well with the research scales. The results of the actual sorting were not as good as was expected at the outset. This indicated that the scale items were neither as discriminant nor as convergent as it was thought they would be during their development.

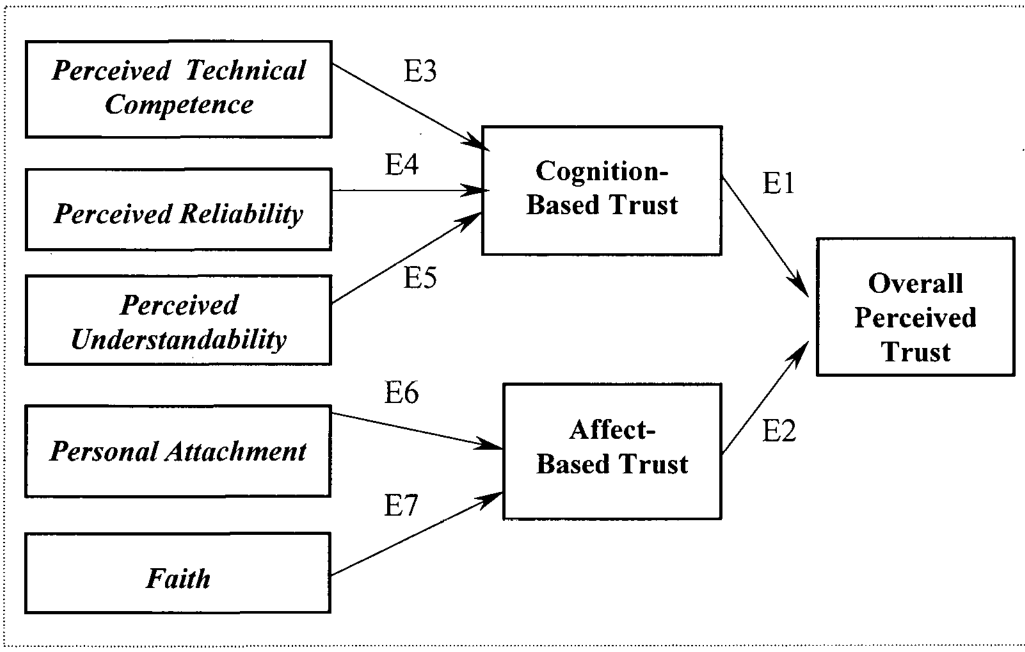
A particular difficulty with this treatment was relating the judges constructs to those expected. In cases where judges' categories differed from those expected decisions had to be made about how to relate these created constructs to the research constructs based on the definition provided by the judges. These decisions were made on a "best-fit" basis. Additional difficulties arose when the judges created fewer categories than expected. The need to relate judges' categories to the expected research constructs in order to calculate Cohen's Kappa resulted in some expected categories having no hits. It is suspected that the overall reliability of these results was lowered by this incongruence between the number of expected and the number of observed categories.

It should be remembered here that the purpose of the judging panels was to assess the discriminant and convergent validity of scale items prior to testing the scales quantitatively. To this end, the series of panels proved to be most helpful and enlightening. Caution should be taken, however, with regard to interpreting the results from this exercise as anything other than a means of refining the new scale. The empirical test of the instrument's validity as a measure of HCT is discussed in the following chapter.

3.4 Summary of the constructs in the new HCT instrument

The dimensions identified during this stage may now be added to the subjective part of the HCT model (Figure 2.1), as illustrated in Figure 3.5. These dimensions, which form the basis upon which "subjective probability judgements" (Rempel et al. 1985) about the future behavior of the IDA can be made, are illustrated along with the expected relationship between each construct and either *affect-based* or *cognition-based trust.*

Figure 3.5 A model of human-computer trust in the context of IDA use.



It is not suggested that these are the only variables that might be of concern to individual users of IDA. It is, however, suggested by the research carried out for this study and discussed in the remaining chapters of this thesis that:

1. these variables are common to users of IDA;
2. these are the variables which were most easily identified by the users of IDA sampled for this research as being important to their judgement of the trustworthiness of the system;
3. measuring these few variables may be sufficient to provide a valid, reliable measure of perceived trust in the context of using an IDA.

Definitions of each of the constructs are given below.

3.4.1 Perceived Technical Competence

The perceived technical competence of the system is defined to be the degree to which the system is perceived to perform the tasks accurately and correctly based on the information that is input. The dimension of technical competence was first introduced into the realm of HCT by Muir (1987) who adopted it from Barber (1983) (see Figure 2.2). There was, however, inconclusive evidence from later studies by Muir and her associates to indicate that this variable was in fact correlated with a user's overall level of trust in an IDA. This variable is included in this study because it encompassed the concepts of *output,* meaning the system is correct and accurate, *speed,* meaning that the system performs within a time frame acceptable to the task and situation in which it is being used and *discriminating,* meaning that the system uses inputs appropriately, resulting from the Nominal Group Stage of this study.

3.4.2 Perceived Reliability

The perceived reliability of the system is defined to be the degree to which the system functions repeatedly and consistently (Sheridan 1988). Reliability is not found as a variable in previous HCT research. However Muir (1987) does include Barber's *persistence* construct and Rempel et al.'s constructs of *predictability* and *dependability.* Although these constructs do not correspond exactly with perceived reliability as it is defined here there are aspects of each of these included in the present definition. Barber's persistence construct for example, implies that a particular behavior continues or is repeated over time and Rempel et al.'s predictability and dependability are both based on a consistency of behavior. The predicament of fine tuning and comparing construct definitions is one of the many problems which plagues trust research. The first data collection stage of this study was designed to identify constructs relevant to HCT specifically. The results of this stage indicated that *reliability,* in the sense that the system functions repeatedly, and *reproducibility of results,* in the sense that the system functions consistently, were important factors and these are encompassed in the definition of reliability as it is used here.

3.4.3 Perceived Understandability

The perceived understandability of the system is defined to be the degree to which a human user or observer can form a mental model of the system's behavior and predict future system behavior (Sheridan 1988). Understandability as it used here suggests that the behavior of the system in terms of "how it does things" is understandable to the user. Understandability is akin to Rempel et al.'s predictability in the sense that both constructs suggest the ability to predict future behavior. The current definition of understandability is believed to include the factors of *conciseness,* meaning that information is displayed as concisely and understandable as possible, and *feedback,* meaning that the system's messages for the user are meaningful and understandable as described by the Nominal Group in the second stage of this study.

3.4.4 Personal Attachment

A user's personal attachment to the system is defined to be the degree to which the system is perceived to be familiar and natural to the user so that using the system is agreeable and preferred to performing the task manually. This variable arose from two proposed variables of liking and loving and two variables from the Nominal Group *friendliness* and *ease-of-use* which were originally assigned to a construct named *familiarity*. Familiarity and personal attachment were not able to be discriminated by judges during the third Stage of this research. The constructs were thus merged into the personal attachment construct according to the results from the judging panels in the third Stage of this study.

3.4.5 Faith

The user's faith in the system is defined to be the degree to which the user has faith that the system will be able to perform future tasks even in a situation in which it is untried. This definition presupposes that the user does not have specific knowledge about this future situation nor about how the IDA will perform in this situation. The user then does not have sufficient knowledge to judge whether the system will perform well or not.

The Faith construct is originally from Rempel et al. (1985) and was used in studies by Lerch et al. (1993), Muir (1994), Muir and Moray (1996) and Lee and Moray (1994). Unfortunately it was adopted by these HCT researchers without sufficient explanation as to how or why it should be involved in trusting a computer system.

Shaw (1997, p.21) has perhaps best explained the construct of faith as a component of trust in his book, *Trust in the Balance*, when he states that,

> "... trust is not always rooted in past experience with others. ... trust is based, in part, on faith. We sometimes give our trust in spite of evidence that might suggest we should feel some caution, if not outright suspicion about relying on another. Trust, however, is not absolute faith. ... Pure faith is beyond reason: those with such faith can justify any event or view ... . Trust, then, is more than simple confidence and less than blind faith."

Users may rely on *faith* that the system will meet their needs when they lack sufficient knowledge to confidently judge the trustworthiness of the system in a particular decision-making situation.

3.4.6 Expected Research Outcomes

It proposed here that each of the constructs identified and defined in this stage of the development process relates to one of the two major sub-constructs of HCT, as illustrated in Figure 3.5 above. The following research propositions will thus be investigated in the final stage of this study, through a principal components analysis.

Table 3.2 Expected relationships among latent and manifest variables

Expected Relationship	Definition
User's overall level of trust in an IDA	
E 1	The user's *cognition-based trust* is a component of the user's *overall perceived trust* in the system.
E 2	The user's *affect-based trust* is a component of the user's *overall perceived trust* in the system.
User's cognitive-based trust in an IDA	
E 3	The user's perception of the *technical competence* of the system is a component of the user's *cognition-based trust* in the system.
E 4	The user's perception of the *reliability* of the system is a component of the user's *cognition-based trust* in the system.
E 5	The user's perception of the *understandability* of the system is a component of the user's *cognition-based trust* in the system.
User's affect-based trust in an IDA	
E 6	The user's *personal attachment* to the system is a component of the user's *affect-based trust* in the system.
E 7	The user's *faith* in the system is a component of the user's *affect-based trust* in the system.

Table 3.3 The scales in the new HCT instrument

Key	Item	No.
R1	The system always provides the advice I require to make my decision.	6
R2	The system performs reliably.	13
R3	The system responds the same way under the same conditions at different times.	14
R4	I can rely on the system to function properly.	16
R5	The system analyzes problems consistently.	28
T1	The system uses appropriate methods to reach decisions.	4
T2	The system has sound knowledge about this type of problem built into it.	12
T3	The advice the system produces is as good as that which a highly competent person could produce.	19
T4	The system correctly uses the information I enter.	22
T5	The system makes use of all the knowledge and information available to it to produce its solution to the problem.	29
U1	I know what will happen the next time I use the system because I understand how it behaves.	10
U2	I understand how the system will assist me with decisions I have to make.	15
U3	Although I may not know exactly how the system works, I know how to use it to make decisions about the problem.	17
U4	It is easy to follow what the system does.	20
U5	I recognize what I should do to get the advice I need from the system the next time I use it.	30
F1	I believe advice from the system even when I don't know for certain that it is correct.	5
F2	When I am uncertain about a decision I believe the system rather than myself.	8
F3	If I am not sure about a decision, I have faith that the system will provide the best solution.	18
F4	When the system gives unusual advice I am confident that the advice is correct.	25
F5	Even if I have no reason to expect the system will be able to solve a difficult problem, I still feel certain that it will.	26
P1	I would feel a sense of loss if the system was unavailable and I could no longer use it.	3
P2	I feel a sense of attachment to using the system.	7
P3	I find the system suitable to my style of decision making.	9
P4	I like using the system for decision making.	21
P5	I have a personal preference for making decisions with the system.	24

Key:

R1 to R5: Perceived Reliability

T1 to T5: Perceived Technical Competence

U1 to U5: Perceived Understandability

F1 to F5: Faith

P1 to P5: Personal Attachment

Number: the number of the item on the final draft of the survey since items were ordered randomly.

3.4.7 Summary

The new model proposed for HCT (Figure 3.5) defines human computer trust by combining the theories of ethopoeia and social attribution in the context of the human-computer joint-cognitive-system. The model further describes the multi-dimensional nature of human computer trust in a way that is intuitively consistent with our knowledge of human experience by taking into account the duality of our intellectual and emotional responses. It is, therefore, more complete and more easily understood than previous models and likely to be generalizable across a wide spectrum of situations in which a human may have need to trust an intelligent computer system.

3.5 Stage 4: Refinement of the HCT Instrument in a Pilot Study

The fourth and final stage in the development of the HCT measurement instrument was a pilot study conducted under experimental conditions. The grammar checking facility incorporated in Microsoft Word '97© was chosen as the IDA for this study because it is a relatively simple instrumental support system to use (refer to Chapter One p.8) and most people, especially university students, are familiar with it. The expected outcome from this study was a further refinement of the set of scales in terms of wording and of the overall survey format to ensure that the written information, instructions and questions were easily understood by respondents.

3.5.1 Participants

Participants in this study were volunteers from the first year student population attending the Gladstone campus of Central Queensland University. There were 10 male and 4 female participants most of whom, being first year students, were under

24 years of age. The computer experience of these participants ranged from one to twenty years with most having over five years experience with computers generally.

The study was performed once with one group of fourteen first year students. The number of participants forming one group was limited by the capacity of the computer room being used for the experiment.

3.5.2 Materials

The materials required for this experiment were Word '97 ©, a piece of prose with grammatical errors inserted and the survey itself, as well as the information sheet and consent form as used previously. The materials for this stage of the research are included in Appendix D.

3.5.3 Design

The primary purpose of this study was to determine the quality of the survey in terms of the instructions given, the information provided and the wording of the questions.

The study was designed to ensure that participants used the grammar checker with a piece of prose to which documented grammatical errors had been inserted, also included in Appendix D. The errors inserted were ones which the MS Word '97© grammar checker could identify. After using the grammar checker to check the test piece for errors the participants were asked to complete the survey based on the experience they had just had with the IDA.

3.5.4 Procedure

The procedure was simple. The group was given verbal information about the research and invited to ask questions. They were then asked to sign the consent form and informed that they could withdraw at any time. The group was then given verbal instructions on how to use the grammar checker and the online help and a printed copy of the test piece (Appendix D.3). They were then asked to work through the following steps:

1. Read through the test piece.
2. Check for grammatical errors with the word processor.
3. Read the on-line Help information and the suggestions presented by the Grammar Checker.
4. Underline each error on the printed copy of the test piece that was underlined in the electronic copy of the test piece by the grammar checker.
5. Circle on the printed copy those of the underlined errors that they decided to change according to the suggestions made by the grammar checker.
6. Once these steps were completed for all the discovered errors, they were to read carefully through the survey instructions and complete the questions.

Participants were encouraged to make written comment on the style, layout and the ease of understanding of the survey as a whole as they went through it. It was these comments that were of particular interest in this stage of the study.

3.5.5 Results

It was not possible to perform quantitative analysis on the data set from the pilot study since the sample size was limited. From the descriptive analysis on this data set, it was seen that the responses for most items were normally distributed, although with such a small sample this can not be taken as anything but an indication that the items were relatively unbiased and understandable.

Of the fourteen students who participated, twelve fully completed the survey. The two who did not answer all the questions had some difficulty identifying what the phrase “the system” referred to although this was stated in writing at the beginning of the survey.

Of the others, one of the participants commented on the color of the cover sheet as being too bright. Another indicated that they would have preferred to have the Likert scale presented in reverse. One indicated that they answered the questions to the best of their ability suggesting that there were some questions that they were not sure of but failed to specify which questions they had trouble with.

Finally, another felt that their own knowledge of grammatical rules and the function of the grammar checker were insufficient to judge whether the grammar checker was correct or not. They would ask someone else before making a decision to accept the grammar checker's suggestion in those instances where they didn't know the grammatical rules themselves.

3.5.6 Discussion

The results of the pilot study indicated that the format of the survey was acceptable and that most of the items by this stage were easy to read and understand. Those items that had caused problems for any one of the students were reconsidered. Changes to these items were made and the Likert scale was reversed. Further minor adjustments to other items were made so that the survey better suited the system which was to be investigated in the final stage which was a scheduling support system for taxi-fare allocation. Particular care was taken with this final reworking of the questions so as not to lose the meaning of the items which had survived the rigorous development process. The color of the cover sheet was also changed from bright yellow to an understated green.

Participants with little confidence in their own knowledge of grammatical rules found that the grammar checker was of little use since they had to either accept or reject the advice from the system without the ability to evaluate its correctness. Those participants who believed they had a good knowledge of grammatical rules found that several of the system's suggestions for improving the grammar in the test piece did not make sense and therefore rejected the suggestion and made their decisions to change the sentence structure based on their own knowledge.

Most of the participants did not know about or rarely used the help facility which provides explanations for the grammatical suggestions. When these participants were asked to use this facility they did not find it particularly useful. These users felt that they needed to have better task knowledge of grammatical rules, in order to be able to know if the system was correct or not.

3.6 Conclusion

The psychometric instrument for human-computer trust was prepared for use in a final test stage of this study. The development and refinement stages resulted in an instrument with five constructs, three which appear to be components of *cognitive-based trust* and two which appear to be components of *affect-based trust*. Based on the results of the sorting rounds each construct has five indicators which also appear to belong to their respective constructs. These stages have thus addressed the issue of the **face validity** of the instrument, that is, **that the constructs and items appear to be measuring what they are supposed to be measuring** (Neuman 1994). The sorting rounds have also address the **equivalence reliability** of the indicators which is "**whether or not the measure yields consistent results across different indicators**" (Neuman 1994). They have also addressed the **content validity** of the instrument, which is **whether or not the indicators used capture the entire meaning of the construct** (Neuman 1994). In order to maintain the content validity of the instrument, the initial pool of items and constructs were sourced from both grounded research and a review of existing studies. The item pool was kept large initially so that many different aspects of the constructs would be investigated in the sorting rounds. The items that survived these rounds are expected to be a reasonably good representative sample of items from this pool ensuring satisfactory content validity (Cronbach 1970).

For an empirical measure of the construct validity and scale reliabilities of the instrument, the survey was tested in the final stage of this research, the field study, in which data was collected from users of operational IDA. The results of this stage are reported in the following chapter.

Chapter 4

Instrument Testing

Chapter Three describes how a psychometric instrument for HCT was developed in a rigorous, multi-staged process. The instrument was then compiled into survey format and a trial of the survey was performed through a pilot study. At this point, it was known that the items comprising the measures could be discriminated by impartial judges, that the items converged on their respective constructs as well as could be determined without quantitative measurement and also that the survey format, including the instructions, was understandable.

The next and final stage of this research was the quantitative testing of the reliability of the scales by calculating Cronbach's alpha[10], as well as the construct and content validity of the instrument through a Principal Components Analysis (PCA)[11].

This final investigation demonstrates the quality of the instrument and helps to elucidate the structure of human-computer trust.

4.1 Participants

The 75 participants in this study were taxi drivers and taxi base operators in several Australian states using automated taxi dispatch systems (TDS).

The mean age of respondents was 44.41 years. The sample was comprised of 71% male and 29% female respondents. 78% of respondents were taxi drivers and 22% were base operators. The majority of respondents, 71%, had less than one year

---

[10] Cronbach's alpha is based on the average co-variance among items in a set and is interpreted as a correlation coefficient (Coakes and Steed 1996).

experience using the TDS while 59% reported having performed the same work without the TDS for more than one year. 35% of respondents reported general computer experience of between one and four years, with 42% having had less than one year experience using any type of computer system.

4.2 Materials

The materials for this part of the study included, the information sheet and consent forms used previously (Appendix E). These forms were included with the survey questions as in the pilot study. The resultant data were analyzed using Principal Components Analysis in SPSS8©.

The selection of an IDA as the research vehicle was particularly challenging. While there are many IDA now in operation in various business areas and industrial applications, it was necessary to find a system which was used by sufficient numbers of people to support the quantitative analysis techniques to be used in this study. This is, however, not unusual in field study designs, which are often difficult to implement due to the need to have the cooperation of the institutions selected for study.

Taxi dispatch systems were selected because they met the criteria for large numbers of users, with every driver and base operator using the system to perform their work.

TDS have been designed to replace the human base operators who allocated fares based on the voice response of the drivers. Problems with the manual system included the ability of drivers to respond to fares which were intended for someone

---

[11] PCA is the linear transformation of a large set of variables into a smaller set of uncorrelated variaables that represent the variation in the original set (Dunteman 1989).

else; the ability of drivers to station themselves in busier areas selectively thereby creating a long waiting period for fares outside these areas; and the possibility of unfair selection of drivers for more lucrative fares such as long distance or account customers. The automated systems were intended to alleviate these difficulties as well as improving the efficiency of the overall system in terms of response time to the customers and the work performed by the operators.

The system was designed to queue the available cars according to the area in which they were operating and the sequence in which they entered the area. By doing this, the fares are distributed more evenly to the cars and the response time to the customer is improved because the cars are at all times in the pickup area. Special fares such as long distance and account customers are also dealt with via a queue. The system, ideally, eliminates the need for the operator to make the selection decisions and the need for the taxi driver to choose for which fares they will bid. At all times the operator can override the system and the driver can reject the fare.

Thus, the taxi dispatch system falls within the range of IDA discussed in Chapter Two. They are prosthetic decision support systems which are monitored by human operators. The users of the dispatch system, both the drivers and the operators, can either accept or reject the decisions made by the system.

The drivers' interaction with the system includes entering the current location according to the area in which they are driving and actively accepting a fare as it is assigned to them. They can reject the assignment by not responding. Non-response places the job back into the queue and it is reallocated to another car.

The operators' interaction with the system includes answering customer calls and manually entering all relevant fare information so that the system can queue the job, monitoring the system continually for any problems or difficulties, overriding the system by amending the queuing data when necessary, providing additional verbal information to drivers about particular fares and other cars when queried, and switching to manual dispatch should the system fail. The operators and drivers also engage in personal informal discussions on non-work related topics from time to time.

The TDS has automated the allocation of fares which has alleviated the need for drivers to decide for which fares they will bid, and for the operators to decide to whom a fare will be allocated. However, both drivers and operators have had to learn additional, new computer related skills in order to continue to perform their work. The automated dispatch system is generally seen to be more equitable and less competitive than the manual dispatch system.

4.3 Design

A field study design was used because it is believed that in order for a user to really make judgments about their trust in an IDA the user needs to be using the system in their regular work and must have had time to make judgments about the various aspects of the system which are being asked about in the instrument. While an experimental design could have assured an adequate sample size it would have necessitated using participants novice to the system chosen.

An IDA thus had to be identified which was in wide use and which had been in use for some time. The taxi dispatch systems met both of these criteria. There are approximately 900 users of these systems among the various taxi companies that were approached. In total, 400 surveys were distributed. Some of these companies have used TDS for up to five years and others have only recently changed over from manual, radio systems to the automated systems. Thus, users with a wide range of both task experience and IDA experience were included.

The purpose of this study was to determine the validity of the measurement instrument as well as to explore the structure of the human-computer trust construct. SPSS8© provided the appropriate factor analytic technique, principal components analysis, with which to investigate the construct validity of the overall HCT construct and the five sub-constructs: *perceived reliability, faith, personal attachment, perceived technical competence and perceived understandability.* It also provided the ability to calculate Cronbach's alpha as a measure of the internal consistency for each of the individual scales.

The construct validity of the HCT instrument is illustrated by the factor loadings, or correlations, between the **items**, which are also referred to as indicators or manifest variables, and the **factors**, which are also referred to as the principal components, sub-constructs or latent variables. The factor loadings also provide evidence of the underlying structure of the constructs and of the convergent and discriminant construct validities of the individual indicators in the scales.

The content validity of each of the scales is demonstrated by the scale reliabilities to the extent that Cronbach's alpha provides an indicator of how well the various items in the scale correlate with each other as a group. Content validity, however, is impossible to prove completely, since when one is investigating psychological constructs there will always be aspects of the construct that remain unknown. Content validity begins with a sufficient definition of the constructs of interest and was primarily addressed in the preliminary stages of this research, described in chapter three, in which a rigorous development process was employed to ensure that the content of the new scales was as valid as possible. It is sufficient, in most cases, that one can demonstrate that the items for a construct are, as indicators, representative of the construct as it has been defined which leads to the ability to perform reliable measurements of the construct of interest (Cronbach 1970, p.145, Neuman 1994, p.131).

The seven point Likert scale used in the pilot study was used here, with 1 being strongly agree and 7 being strongly disagree. A seven point scale was chosen because it offers the respondent a wider range of choices and a median choice of 4 which is neither a positive nor a negative response for those respondents who did not have an opinion on a particular item (Neuman 1994). Missing values were substituted with the sample mean for that item. The random ordering of items used in the pilot study was also used here to minimize precedence effects (Neuman 1994). As discussed in the previous chapter, the random ordering of items was accomplished by assigning random values to each of the items then sorting on these values using Microsoft Excel 97©.

4.4 Procedure

A majority of the 400 surveys were distributed by the Taxi Base managers who then collected and returned both the completed and unused survey forms. Some surveys, however, were distributed by the researcher directly to participants in an effort to improve the rate of distribution. Two further strategies were employed in order to improve the survey response rate. The first was used with those groups that were easily accessed by the researcher. For these groups a raffle ticket was attached to each survey and every respondent was entered into a draw for a dinner for two to the value of $100.00. Groups outside the local area received surveys with $1.00 coins attached to the cover sheet.

The data was collated using Microsoft EXCEL 97©. Missing values were coded as 9, and the randomly ordered items were re-grouped into their constructs to facilitate the analysis. Other items in the survey, such as demographic data, were encoded according to their groupings on the survey. For example, ages were categorized into five groups and each was assigned a digit from 0-4 to facilitate the descriptive analysis of the respondent sample. The data were then imported into SPSS8©. First the reliabilities of the five scales were checked and then an exploratory factor analysis was performed using Principle Components Analysis.

4.5 Results

4.5.1 Data screening

The response rate for this survey was less than 25% overall, with only 78 completed forms of 400 surveys returned. From the original sample of 78 respondents, 3 were removed from the data set. Two of these had the same responses to all questions. Since at least two of the non-trust items could not be answered in the same way, as they were mutually exclusive, it was decided that these responses were not sufficiently reliable to include in the analysis. The third deletion was one in which the respondent had apparently answered the entire questionnaire in response to an accounting system. This was evident from question 3 which asked participants to identify the system about which they were responding.

The remaining 75 responses were deemed to be useable. Items with missing responses were few although R5, and F5, had 7 and 5 missing responses respectively. Mean substitution was used to deal with missing responses in these cases. Two respondents failed to identify their job title. There was no reasonable method of replacing these values. This variable was not used in the main principal components analysis. Missing values were thus of little concern. Eight respondents did not disclose their age, however, since age was simply a demographic variable for the purpose of profiling the respondent sample, this did not compromise the results.

There were two ways in which the data in this set could be grouped. The first was based on the type of job the respondent performed, either operator or driver. The second was based on the type of TDS being used. A between groups comparison of means using the Mann-Whitney U test was performed to ensure that there was no significant differences between group in the case of either of these variables.

First the data was split according to the type of TDS used. The two TDS sampled in this study were the Expertech and Raywood systems. An analysis of the means for these two groups, Table 4.1 below, shows that there is no significant differences between these groups with $p > 0.05$.

Table 4.1 Comparison of HCT sample means split by system

Expertech TDS (64 cases) v. Raywood TDS (11 cases)
Total Cases: 75 Missing Cases: 0

Statistic	Result
Mann-Whitney U	315.000
Wilcoxon W	381.000
Z	-0.554
Asymp. Sig. (2-tailed P)	**0.579**

Second the data was split according to the type of job the respondent performed with the TDS. Although, there is some difference in the way in which base operators and drivers use the TDS, there was no significant differences found between these two groups in their responses to this survey with $p > 0.05$, see Table 4.2 below.

Table 4.2 Comparison of HCT sample means split by job

Operators (16 cases) v. **Drivers (57 cases)**
Total Cases: 73 Missing Cases: 2

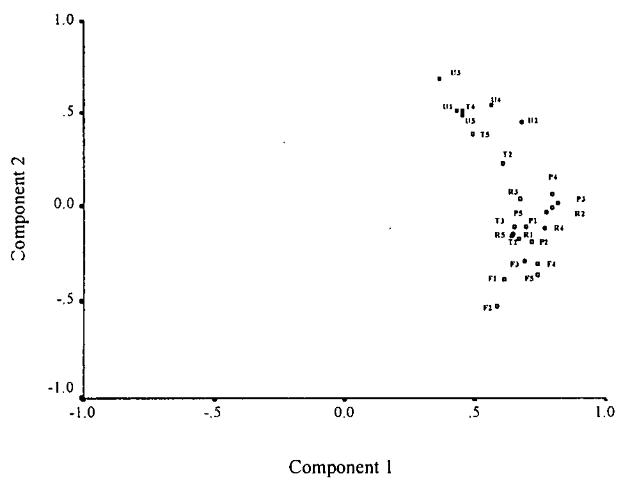
Statistic	Result
Mann-Whitney U	355.000
Wilcoxon W	491.000
Z	-1.347
Asymp. Sig. (2-tailed P)	**0.178**

Since there were no between group differences for either of these variables the data set consisting of 75 respondents was analysed during the remainder of this study without splitting the sample.

Principal Components Analysis is robust to assumptions of normality; therefore it was not necessary to adjust variables to normal distributions by transformation or re-coding prior to analysis. However, PCA can be sensitive to bivariate outliers, since the analysis is based on the correlation or covariance matrix of the original data. Once an analysis has been performed, a plot of the factor scores for the first two factors, which generally comprise the greatest explained variance in the data set, can be checked for outliers[12]. The component plot for the first two factors in this analysis, Figure 4.1, shows that there are no outlying variables in this set of data.

[12] Outliers are not necessarily influential to the analysis. If found, outliers should be removed from the data and the PCA analysis should be performed again. If there is a substantial change in the results then the outlying variable is influential and may be deleted from the data set (Dunteman 1989).

Figure 4.1 Plot of First Two components from PCA of Taxi Data



Although, the size of this sample fell short of the ideal minimum of 125 responses, five respondents to each item as discussed in Chapter One, its was adequate for principal components analysis, as indicated in Table 4.1 below. SPSS8© provides tests for sampling adequacy such as Kaiser-Meyer-Olkin measure of sampling adequacy[13] and Bartlett's test of spherecity[14] prior to performing a principal components analysis as a means of determining the suitability of the sample results for factor analysis. It can be seen in Table 4.3 below, that the KMO test for sampling adequacy is above 0.6 and Bartlett's test of spherecity is significant. These results indicate that the sample, although small, is suitable for factor analysis.

---

[13] "KMO measure of sampling adequacy tests whether the partial correlations among variables are small." SPSS8©

[14] "Bartlett's test of spherecity tests the null hypothesis that the correlation matrix is an identity matrix. The data must be a sample from a multi-variate normal population. If the null hypothesis cannot be rejected, and the sample size is reasonably large, ...(then) the use of multi-variate analysis (is not appropriate) since the dependent variables are not correlated. " , SPSS8©

Table 4.3 Sampling Adequacy

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.863
Bartlett's Test of Sphericity	**Approx. Chi-Square**	1268.912
	Degrees if freedom	300
	Significance	0.000

4.5.2 Scale reliability Analysis

The results of the reliability analyses based on the calculation of Cronbach's alpha were promising. The reliability calculated for the overall HCT construct from all items was **0.94**, with four of the five individual scales having alpha's of at least **0.84, as** illustrated in Table 4.4. Only one of the items, T4, which is highlighted in Table 4.1, was seen to improve the alpha of its own scale by its elimination. This indicates that all other items were necessary to the overall reliability of their respective scales. It also suggests that the convergent construct validity and the content validity of the scales, with the exception of the *perceived technical competence* scale is satisfactory.

Table 4.4 Scale Reliabilities reported as Cronbach's alpha (α)

Scale	Standard α	Item	α if item removed
Reliability	0.85	R1	0.84
		R2	0.79
		R3	0.80
		R4	0.79
		R5	0.83
Understandability	0.84	U1	0.81
		U2	0.78
		U3	0.79
		U4	0.80
		U5	0.82
Technical Competence	0.74	T1	0.70
		T2	0.64
		T3	0.70
		T4	0.75
		T5	0.68
Faith	0.88	F1	0.86
		F2	0.83
		F3	0.86
		F4	0.85
		F5	0.84
Personal Attachment	0.90	P1	0.89
		P2	0.89
		P3	0.87
		P4	0.86
		P5	0.87
Overall HCT	**0.94**	All	Appendix E.5

These results support the rigorous process undertaken in the creation of the scales, described in Chapter Three, as a sound method to use to select and refine measures of psychological variables such human-computer trust.

4.5.3 Principal Components Analysis

Principal components analysis(PCA) is a structural equation modeling (SEM) technique that searches for the principal components comprising a data set. It is both a method of data reduction which allows a large data set to be explained in terms of fewer variables and an exploratory method with which one can investigate the structural relationships among a set of variables (Dunteman 1989, Kim & Mueller 1978). PCA is relatively well suited to exploratory studies where the underlying structure of the variables is unknown. It was well suited to the current study, providing rotated component matrices that were much more easily interpreted than those produced by principal axis factoring (PAF) with this same data.

4.5.3.1 Preliminary Analysis

The initial principal components analysis allowed the SPSS program's default method of displaying factors with eigenvalues[15] greater than one. This method resulted in four factors (Table 4.5) with the *perceived reliability* factor missing and its items spread across the other scales. This analysis suggested that there was one overall dominant factor as expected, easily seen on the scree plot (Figure 4.2). Since this study was of an exploratory nature, it was decided to use PCA to explore models with the number of factors that were expected from the proposed model, Figure 3.5. Hence, one, two and five factor models were specified in the analysis criteria.

---

[15] Eigenvalues are the squared correlations from the factor matrix.

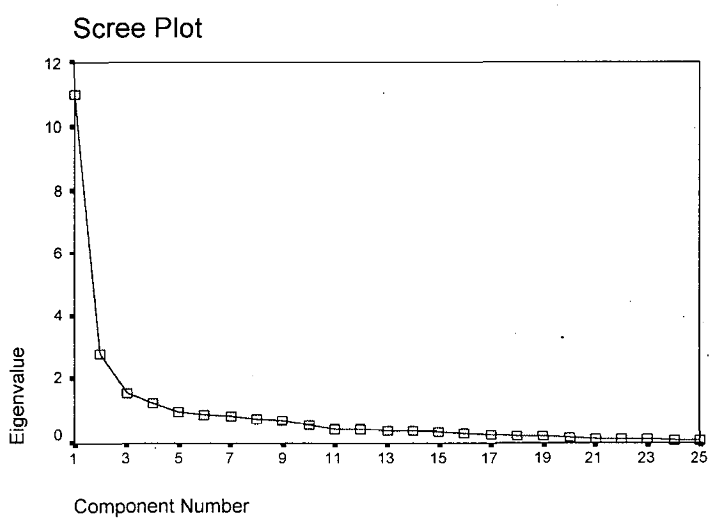
Table 4.5 SPSS Default Model for TDS data: Four Factors displayed.

Observed Component - Expected Variable correspondence based on item loadings:

1 - Personal Attachment
2 - Faith
3 - Understandability
4 - Technical Competence

	Component			
	1	2	3	4
P1	0.690			0.328
P2	0.703	0.337		
P3	0.702	0.335	0.316	
P4	0.776		0.392	
P5	0.765			
F1		0.755		
F2		0.767		
F3		0.759		
F4	0.572	0.589		
F5	0.409	0.664		0.312
U1			0.750	
U2	0.385		0.740	
U3			0.812	
U4			0.680	0.304
U5			0.555	0.457
T1	0.494	0.393		0.325
T2				0.734
T3		0.694		
T4	0.394		0.429	0.410
T5			0.318	0.775
R1	0.329	0.534		
R2	0.453	0.516	0.349	
R3	0.356	0.342		0.493
R4	0.560	0.489		
R5	0.459			0.646

Figure 4.2 Scree Plot from Prinicipal Components Analysis for TDS Data



4.5.3.2 Single Factor Model

All items loaded with correlations above 0.3 on one overall dominant factor (Table 4.6). This was seen to be a positive outcome in terms of the construct validity of the instrument as a whole, since the scales were developed to measure one overall construct, human-computer trust.

Table 4.6 Single Over-arching Factor of HCT

Component Matrix(a)	
	Component 1
F1	0.614
F2	0.587
F3	0.687
F4	0.739
F5	0.741
P1	0.697
P2	0.717
P3	0.815
P4	0.796
P5	0.772
R1	0.642
R2	0.794
R3	0.673
R4	0.767
R5	0.648
T1	0.667
T2	0.609
T3	0.652
T4	0.432
T5	0.490
U1	0.452
U2	0.679
U3	0.361
U4	0.565
U5	0.453

Key:

R1 to R5: Perceived Reliability

T1 to T5: Perceived Technical Competence

U1 to U5: Perceived Understandability

F1 to F5: Faith

P1 to P5: Personal Attachment

4.5.3.3 Two Factor Model

The two factors proposed in the initial model for HCT were affect-based trust and cognitive-based trust. It can be seen clearly, from the rotated component matrix (Table 4.7) that two factors exist and that the items that were considered to be affect-based items load well on the first factor. This factor can thus be considered to be the affect-based trust latent variable as proposed in chapter two. An unexpected result was that items belonging to *perceived reliability*, which were originally thought to be related to cognitive-based trust, can be seen here to load strongly on the affect-based component. The second factor, cognitive-based trust, is comprised of all the *understandability* and three of the *perceived technical competence* items.

Table 4.7 Affect-Based and Cognitive-Based Trust Components

	Component	
	1	2
F1	0.724	
F2	0.775	
F3	0.737	
F4	0.790	
F5	0.818	
P1	0.652	
P2	0.708	
P3	0.688	0.438
P4	0.649	0.465
P5	0.673	0.379
R1	0.63	
R2	0.681	0.408
R3	0.553	0.386
R4	0.716	
R5	0.629	
T1	0.659	
T2	0.403	0.508
T3	0.613	
T4		0.660
T5		0.585
U1		0.649
U2	0.347	0.735
U3		0.772
U4		0.757
U5		0.673

Component	-	Expected Variable
1	-	Affect-based trust
2	-	Cognition-based trust

Key:

R1 to R5: Perceived Reliability

T1 to T5: Perceived Technical Competence

U1 to U5: Perceived Understandability

F1 to F5: Faith

P1 to P5: Personal Attachment

4.5.3.4 The Five Factor Model

The fifth factor, which was not displayed in the default model (Table 4.3) had an eigenvalue of 0.979. However, using a particular eigenvalue as a cutoff point is an arbitrary decision and in this case not very useful[16]. Often the scree plot (Figure 4.2) can help to decide how many variables should be considered. However, in this case where there is a single overarching factor it is difficult to make a decisive judgement about at which point, after the first two, the plot levels out. The decision to investigate the five factor model was based then, not on a particular eigenvalue that might easily be debated, nor on the scree plot which suggests only two variables, but rather based on the expected outcome of the analysis. As such, it was a means of

[16] There are a number of methods suggested for interpretation of the principal components results. Eigenvalues above one is one method that it widely accepted, another is an interpretation of the scree plot, another is to choose the number of factors based on the cumulative amount of variance explained (Dunteman 1989).

exploring the nature of the sub-constructs or latent variables, and the discriminant validities of the items, or manifest variables.

The recommended threshold at which factor loadings are considered important enough to display is generally set at correlation values of 0.3 and above, (Table 4.8).

Table 4.8 HCT Five Factor Model - 0.3 cut off

	Component				
	1	2	3	4	5
P1	0.680			0.337	
P2	0.693				
P3	0.696				
P4	0.776				
P5	0.762				
F1		0.769			
F2		0.819			
F3		0.681		0.318	0.330
F4	0.557	0.540			
F5	0.383	0.654		0.337	
U1			0.667		0.384
U2	0.396		0.700		
U3			0.876		
U4			0.620		0.327
U5			0.660	0.413	
T1	0.475	0.310		0.353	
T2				0.735	
T3		0.438			0.683
T4	0.405		0.450	0.386	
T5			0.330	0.767	
R1	0.317	0.608			
R2	0.438				0.628
R3	0.340			0.512	0.326
R4	0.546	0.303			0.533
R5	0.436			0.669	

Component	-	**Expected Variable**
1	-	**Personal Attachment**
2	-	**Faith**
3	-	**Understandability**
4	-	**Technical Competence**

Key:

R1 to R5: Perceived Reliability

T1 to T5: Perceived Technical Competence

U1 to U5: Perceived Understandability

F1 to F5: Faith

P1 to P5: Personal Attachment

With the output forced to five factors, it became apparent that the display of factor loadings above 0.3 included loadings that were not necessarily useful to this analysis. It can be seen in this table that there is a great deal of noise from items that load on their own scale strongly but also load between 0.3 and 0.4 on other constructs. These extraneous loadings make this matrix rather difficult to interpret. Generally, one examines the difference between the expected loading and the other loadings and if there is a great enough difference (0.2 or greater) then the extraneous, smaller loadings may be disregarded. To demonstrate this point further the threshold was set at 0.1 (Table 4.9) and it can be seen that there are a great number of extraneous, small loadings.

Table 4.9 HCT Five Factor Model - 0.1 cut off

	Component 1	2	3	4	5
P1	0.680	0.288		0.337	
P2	0.693	0.385		0.164	
P3	0.696	0.275	0.270	0.191	0.263
P4	0.776	0.217	0.336		0.261
P5	0.762	0.268	0.240	0.101	0.162
F1	0.247	0.769	0.100		0.180
F2	0.246	0.819			
F3	0.152	0.681		0.318	0.330
F4	0.557	0.540			0.283
F5	0.383	0.654		0.337	0.175
U1	0.191		0.667		0.384
U2	0.396	0.140	0.700		0.277
U3			0.876	0.105	
U4	0.249		0.620	0.289	0.327
U5		0.184	0.660	0.413	-0.165
T1	0.475	0.310		0.353	0.266
T2	0.167	0.177	0.244	0.735	0.143
T3		0.438	0.137	0.272	0.683
T4	0.405	-0.209	0.450	0.386	
T5			0.330	0.767	0.148
R1	0.317	0.608	0.221	0.185	
R2	0.438	0.283	0.211	0.258	0.628
R3	0.340	0.220	0.134	0.512	0.326
R4	0.546	0.303	0.123	0.156	0.533
R5	0.436	0.265	-0.106	0.669	

Component	-	Expected Variable
1	-	Personal Attachment
2	-	Faith
3	-	Understandability
4	-	Technical Competence

Key:

R1 to R5: Perceived Reliability

T1 to T5: Perceived Technical Competence

U1 to U5: Perceived Understandability

F1 to F5: Faith

P1 to P5: Personal Attachment

An inspection of the matrix thus led to a decision to use 0.4 as the threshold (Table 4.10). This value eliminates much of the noise while not eliminating any item completely. It should be noted that loadings of 0.6 and above are considered to be strong and reliable (Dunteman 1989). Therefore, 0.4 is still low enough to include less reliable correlations highlighting some of the problematic variables in this study.

Table 4.10 Five Factors - 0.4 cut off

	Component 1	2	3	4	5
P1	0.680				
P2	0.693				
P3	0.696				
P4	0.776				
P5	0.762				
F1		0.769			
F2		0.819			
F3		0.681			
F4	0.557	0.540			
F5		0.654			
U1			0.667		
U2			0.700		
U3			0.876		
U4			0.620		
U5			0.660	0.413	
T1	0.475				
T2				0.735	
T3		0.438			0.683
T4♦	0.405		0.450		
T5				0.767	
R1		0.608			
R2	0.438				0.628
R3				0.512	
R4	0.546				0.533
R5	0.436			0.669	

Component	-	Expected Variable
1	-	Personal Attachment
2	-	Faith
3	-	Understandability
4	-	Technical Competence

Key:

R1 to R5: Perceived Reliability

T1 to T5: Perceived Technical Competence

U1 to U5: Perceived Understandability

F1 to F5: Faith

P1 to P5: Personal Attachment

♦**T4** is the only item for which the scale reliability would have increased by its removal.

It can be seen in Table 4.10, which has been ordered so that expected loadings fall along the diagonal, shaded cells, that the five scales do in fact appear with the *personal attachment, faith,* and *understandability* scales being comprised of all five of the items which were expected to load on these factors. From the table of Total Variance Explained (Table 4.11) it can be seen that the first five variables account for **68.9%** of the variance in this data set. Items such as R3 and T4 , however, do not show good discriminant validity. The *perceived reliability* scale is particularly problematic with no items loading cleanly on this scale. These items are discussed in more detail in the following section.

Table 4.11 Total Variance Explained by each of the first five factors

	Initial Eigenvalues		Rotation Sums of Squared Loadings			
Comp.	**Total**	**% of Variance**	**Cumulative %**	**Total**	**% of Variance**	**Cumulative %**
1	10.674	42.694	42.694	4.788	19.151	19.151
2	2.798	11.194	53.888	3.958	15.833	34.984
3	1.562	6.250	60.137	3.333	13.332	48.316
4	1.222	4.890	65.027	2.998	11.990	60.306
5	0.979	3.914	**68.941**	2.159	8.636	68.941

4.5.4 Analysis of individual items based on PCA results

The findings from the five factor model provide the most detailed picture of the problems with individual items. As discussed above, the choice of threshold at which to display and examine the loadings varies among researchers with little agreement. Within the loose guidelines that do exist it is best left to the individual to choose the limit based on the nature of the data and the aim of the research. The matrix display of factor loadings of 0.4 and above was primarily used for the following analysis of individual items. The minor loadings, illustrated in Table 4.9, provided the broader context within which the items were examined. For example, only in Table 4.9 is it evident that all of the *perceived technical competence* items did load on their own construct to some extent, while only three of the *perceived reliability* items loaded on their own construct at even very low levels.

Specifically, it is strong loadings which provide the best information about the nature of the latent variables and variables with multiple low loadings which can be considered to be the most problematic. Furthermore, in a study of this nature variables that do not load at a high level on their expected factors are of particular interest.

Consideration of the item loading patterns discussed above, **strong loadings**, **multiple low loadings**, and **poor loading on expected factors** provides the information necessary to improve the overall quality of the instrument. Analyzing the rotated component matrix (Table 4.10) with these three issues in mind, leads to the identification of eight items which require more detailed examination. These items are F4, R1, R3, R5, T1, T3 and T4 (Table 4.12).

Table 4.12 Problematic items from HCT scales

Key	Item	No.
R1	The system always provides the advice I require to make my decision.	6
R3	The system responds the same way under the same conditions at different times.	14
R4	I can rely on the system to function properly.	16
R5	The system analyzes problems consistently.	28
T1	The system uses appropriate methods to reach decisions.	4
T3	The advice the system produces is as good as that which a highly competent person could produce.	19
T4	The system correctly uses the information I enter.	22
F4	When the system gives unusual advice I am confident that the advice is correct.	25

Key:
R1 to R5: Perceived Reliability
T1 to T5: Perceived Technical Competence
U1 to U5: Perceived Understandability
F1 to F5: Faith
P1 to P5: Personal Attachment
Number: the number of the item on the final draft of the survey since items were ordered randomly.

These items cannot simply be deleted from the instrument because, as seen from the scale reliabilities (Table 4.4) each item, with the exception of T4, helps to improve the overall reliability of its respective scale. Were this not the case, it might be a simple matter of eliminating the items that caused problems. It was expected at the outset the final instrument might be reduced to three items per construct. However, given the reliability results this no longer seems to be a viable solution. Removal of these items without further investigation would undoubtedly have a negative effect on the content validity and reliability of this instrument (Cronbach 1970).

These items will, therefore, be examined more closely from three perspectives in an effort to improve the overall effectiveness of the instrument.

- **Placement:** First, each of these eight items will be examined in terms of where they were placed in relation to other items in the survey as a means of investigating the possible influence of precedence effects.
- **Wording:** Second, the wording and general construction of the question will be considered, since key words such as personal pronouns, can influence the respondents' interpretation of the question and may confound the intended meaning.
- **Meaning:** Finally, the relationship between the question's intended meaning and the observed "misinterpretation" of its meaning will be examined.

These three features of the problematic items lead to the final recommendations for the revision of the HCT instrument in this study. For the following discussion, the reader is referred to these items which are listed along with their ID and question number in Table 3.3.

4.5.5 Faith

There was only one item from the *faith* scale that loaded relatively equally on both its own construct and *personal attachment.* It is therefore the least problematic of the eight items and is considered first.

- **F4: When the system gives unusual advice I am confident that the advice is correct.**

This is the fourth *faith* item, F4, F4 was preceded by item P5 and followed by item F5 in the survey. Given that F4 loaded on both the *personal attachment* and *faith* constructs some effect of placement cannot be discounted here. The 25 HCT items were ordered randomly in the survey to reduce such precedence effects, however it seems that the random placement of F4 between a personal attachment item and a faith item may have influenced some responses.

The wording of the item may also have caused some difficulty with its interpretation. In particular the phrase "I am confident", which in this question was intended to tap into the confidence aspect of trust at a time when the user does not have sufficient knowledge to judge whether the advice is in fact correct, may have led respondents to associate this with the personal attachment items which are all written in the first person, present tense.

Except for the use of this phrase, this question appears to be very similar in meaning to F1 which was interpreted precisely as expected. It seems likely then that through a combination of placement and wording this item divided respondents' interpretations sufficiently to cause the dual loading observed. The recommendation to improve this item is to reword it to clarify its intended meaning and possibly to change its position in the survey. The rewording of the item alone, however, might be a sufficient remedy.

4.5.6 Perceived Technical Competence

The next items that will be examined are the *perceived technical competence* items T1, T3 and T4. From Table 4.6 it can be seen that T1 loaded on everything but *understandability* and loaded most highly on *personal attachment*. T3 loaded on *perceived reliability* and T4 loaded on *perceived technical competence* as well as *personal attachment* and *understandability*. From the reliability analysis, it is known that the removal of T4 from the *perceived technical competence* scale would improve the scale. The removal of T1 however, would lower the overall reliability of the *perceived technical competence* scale.

- **T1: The system uses appropriate methods to reach decisions.**

T1 is preceded by P1 and followed by F1. Upon consideration of the wording of P1 and F1 it seems unlikely that responses to T1 were influenced by either of these items since the content of each is quite distinct. It is more likely that asking the respondents to judge whether the system uses "appropriate methods" or not was too difficult. It may be impossible for most IDA users to know whether the analysis methods are appropriate since they may not know what analysis methods are used. This item was intended to investigate the correctness and accuracy of the problem solving methods used in the system, however, it has obviously been interpreted in various other ways. It is clear that the meaning of the question needs to be clarified. A rewording of this question to remove the reference to "appropriate methods" is therefore recommended.

- **T4: The system correctly uses the information I enter.**

T4 is preceded by P4 and followed by question 23, which is not a trust related question. T4 loaded most highly on the *understandability* construct, although the loadings were too close in strength to be reliably interpreted. The wording of the question is ambiguous in relation to the other items since it contains the words, "correctly", "I", and "information".

The emphasis, from the point of view of this research, was the correctness of the system's use of data. However, some respondents have apparently interpreted the question as being similar to their understanding of what the system does, while others have interpreted it to relate to how they feel about using the system and finally others have taken the intended interpretation and associated the question with the accuracy and correctness of the system's solutions.

An evaluation by this group of users about the correctness of the system may be correlated to their liking of the system. Discussions with several of the respondents after they had completed the survey, indicated that they felt that the system was fairer than the manual system. Those respondents holding to this opinion may indeed believe that the system gives correct solutions while others may believe that the system is not correct because they preferred the manual dispatch system. Therefore there may be a problem with the determination of the correctness of the system rather than an interpretation of the item itself.

The users' responses may not be a true evaluation of the system, but rather an evaluation that is influenced by their a priori preferences and feelings with regard to fare allocation generally. None-the-less, a reliable item is expected to transcend such specific sample issues. This item therefore needs to be reworded to be more explicit and less susceptible to context specificity or better still, replaced with a new item as suggested by the scale reliability results.

- **T3: The advice the system produces is as good as that which a highly competent person could produce.**

T3 is the next item to be analyzed. T3 was preceded by F3 and followed by U4. T3 loaded on the *perceived reliability* construct therefore the item was not confounded by precedence effects.

This is perhaps the most complicated item in terms of its wording and it is strongly believed that this complexity and the nature of the question itself contributed to its poor performance. Apparently, respondents associated the comparison of the system's solutions to those from a highly competent person with the reliability of the system. It does not seem unreasonable to associate competence with reliability, however, this is not the purpose of the item. The item was designed to be the cornerstone of the *perceived technical competence* items. The association between the *perceived technical competence* and *perceived reliability* constructs is considered in greater detail in section 4.6 and again in chapter 5. Should this scale be retained, it is highly recommended that this item be reworded to avoid association of the system with a human expert and to simplify the wording as much as possible.

4.5.7 Perceived Reliability

Finally, the *perceived reliability* items are examined. These items showed good reliability in terms of the overall *perceived reliability* scale with an alpha of 0.85. Yet, items R1, R3, and R5 failed to load on the *perceived reliability* construct. The remaining two items R2 and R4 loaded on *perceived reliability* as well as on *personal attachment* with R4 loading slightly more strongly on the latter than on its own construct. The dual loading of R2 is the least problematic issue. With R2 loading relatively more strongly on *perceived reliability* than on *personal attachment,* and the wording of the item being simple and easily understood this item is to be left unchanged at this time. It is possible that once the other *perceived reliability* items are improved this item's loading on its own construct will also improve.

- **R1: The system always provides the advice I require to make my decision.**

R1 was preceded by F1 and followed by P2. F1 and R1 are in fact similarly worded with both questions asking about "advice" from the system. It is possible in this case that having just answered the *faith* question this *perceived reliability* item seemed to be asking about the same thing. The random placement of these items together may have been a little unfortunate.

The use of the term "always" in this question may also have contributed to the respondents' association of this item with the *faith* items.

This item is attempting to tap into the user's perception of how reliable they find the system at the present time, rather than how reliable they believe the system might be at some future time. The subtle difference between these concepts is not easily discriminated and may have been further obscured by the wording and the placement of this item. In light of these findings this item should also be reworded. The recommendation is to remove the word "always" and to simplify the statement. It may or may not be necessary to reposition the item in the survey once the changes have been made.

- **R4: I can rely on the system to function properly.**

Item R4 was preceded by U2 and followed by U3. If there were precedence effects here then this item surely should have loaded on *understandability.* This effect can thus be discounted for this item.

The wording of item R4 would suggest that it might be confused with *perceived technical competence* given that it contains the phrase "function properly", however this is also not the case. It seems that in this case the fact that respondents believed that they could rely on the system to function properly was directly related to how much they liked using the system. Perhaps this is an artifact of the context within which this system operates, but this does not seem a probable explanation for this observation. The question is concise and so closely related in appearance and meaning to R2, at face value it is almost beyond explanation that this result should have been observed.

The one outstanding feature of this item, which differs from the other *perceived reliability* items, is the use of the personal pronoun "I". It is possible that since the other items were worded as statements about the system, this item in using the personal pronoun elicited the unexpected association of this item with the *personal attachment* items. It is recommended that this item be reworded before being used in subsequent research.

- **R3: The system responds the same way under the same conditions at different times.**

R3 was preceded by R2 and followed by U2 in the survey, yet it was correlated with the *perceived technical competence* construct. Again, precedence effects are not apparent here. The wording of this item is a little cumbersome, yet the meaning of the item seems to have been understood. This item was intended to tap into the consistency aspect of the *perceived reliability* construct. Further discussion of this issue is undertaken in conjunction with the second consistency item, R5 below.

- **R5: The system analyzes problems consistently.**

R5 was preceded by question 27, an item unrelated to HCT, and followed by T5. It is possible that responses to this item were influenced by it proximity to the *technical competence* item T5. There is, however, no clearly apparent similarity between these two items. However, the wording of R5 was obviously a problem to some respondents since this item had the greatest number of non-responses, 7 of the 75 respondents failed to answer this question. The results then are calculated on the mean substituted responses for these 7 respondents. The validity of these results as such remains questionable.

The similar results for items R3 and R5 raise a particularly important issue. For instance, R3 and R5 were meant to capture the consistency aspect of *perceived reliability* under the definition proposed in section 3.4.2 in this study. Both of these items were observed to be associated with the *perceived technical competence* construct by respondents. This suggests that users related consistency with competence rather than with reliability. In light of this, one must ask whether either or both of these construct needs to be redefined or whether it is simply necessary to ask better questions.

For the moment, assuming that the constructs remain unchanged, the recommendation is to simplify the wording of R3 and to reword R5 to ask about consistency of performance rather than analysis. It should be noted that the latter change may alter the underlying variable being measured. If, however, the resultant variable is more easily discriminated than the current R5, this change may improve the overall reliability and construct validity of the scale. The effect of any such change on the content validity of the scale also needs to be considered.

4.6 Discussion

It has been suggested that the *perceived reliability* scale, which demonstrated good convergent validity based on the scale reliability results might need to be redefined because it demonstrated a lack of discriminant validity based in the factor analysis. On first inspection, the results would suggest that the *perceived reliability* construct needs to be redefined to eliminate the consistency component. Would **consistency** then be seen to be an indicator of *perceived technical competence* on which it loaded or is it simply to be discarded? It is not possible from the results of this study to answer this question completely. However, it is worthwhile revisiting the definitions of both *perceived reliability* and *perceived technical competence* at this time.

The definition of *perceived reliability* used in this study was taken from Sheridan (1988). It encompasses both Barber's (1983) concept of *persistence* and Rempel et al.'s (1985) *predictability* and *dependability*. It seems in light of the results discussed above that respondents might have more easily distinguished *perceived reliability* items had they been designed simply to measure Barber's persistence construct. That is, that the system continues to function repeatedly over time. This, however, still appears to be an inadequate definition of reliability since *consistency of function* is a vital aspect of reliability. For example, a system which continues to function, **persistence**, does not necessarily function the same way every time, **consistency**.

The definition of *perceived technical competence* is adapted from Barber's (1983) technical competence construct and has been used here to refer to the system's correct and accurate performance of tasks based on information that is entered. However, associating consistency of function with correctness and accuracy is not always valid. For example, an error-prone system that performs the same way each time could not be said to be correct (Will 1992).

The nature of the relationship between repeated and consistent functioning, *perceived reliability,* and correct and accurate functioning, *perceived technical competence,* of the system obviously requires further examination. The face validity of these constructs in terms of the way in which they have been defined here appears to be high, yet the observed construct validity is not as definitive. It may possible to propose alternatives to the expected constructs based on the concept that the **unexpected combination** of items ***appears*** to be measuring. A closer look at the item loadings that comprised each of these two factors, as observed in the results of this analysis, may provide clues as to whether or not these items tapped into one or more constructs that had not been initially considered.

The *perceived technical competence* factor found from the principal components analysis was comprised primarily of two *perceived technical competence* variables, T2 and T5 and the two *perceived reliability* variables discussed above, R3 and R5. T2 and T5 are items about the knowledge and information that the system has built into it. R3 and R5, as discussed above, are concerned with the consistency of the system's response and analysis respectively. T4, which is clearly an unreliable indicator of any of the constructs examined here, is disregarded in this discussion, as it must be replaced with a more reliable item once the definitions of these constructs have been resolved.

The most obvious problem with the *perceived technical competence* items are their scale reliabilities. Cronbach's alpha for the *perceived technical competence* scale was the lowest of the scale reliabilities found in this analysis, 0.74. This means that the consistency of responses to these items was just adequate and that the convergent validity of the items as indicators of *perceived technical competence* could be improved. This inconsistency is also reflected in the haphazard pattern of multiple loadings displayed by these items as a group, Table 4.8 illustrates this best. This implies that not only are these *perceived technical competence* items relatively unreliable for their own construct, they are not particularly reliable indicators for any other construct investigated here, neither do they suggest any clear alternative construct.

The *perceived reliability* factor found in the five factor matrix is comprised primarily of R2 and R4, which are clearly asking about the reliability of the system, as well as T3, which is asking about the competence of the system. T3 is, in fact, the only item in the *perceived technical competence* set that asks about competence with all other items attempting to tap into other aspects of competence such as correct use of knowledge, and use of accurate problem solving methods. Yet it loaded on the *perceived reliability* construct, instead.

A further complicating issue with the *perceived reliability* construct is that, unlike *perceived technical competence*, its items are more closely associated with other constructs, particularly *personal attachment* and *faith*, as seen in Table 4.6. Any change, therefore, to its definition and subsequently its indicators, may adversely effect the content validity and the reliability of the overall HCT measure.

An examination of the scale reliability for the combined *personal attachment* and *perceived reliability* items, for example, resulted in an alpha of 0.93. This is a much improved result for both scales which were 0.90 and 0.85 respectively (Table 4.4). Combining the *perceived technical competence* items with any of the other scales resulted in lower alphas for those scales except for *understandability*, which marginally improved, from 0.84 to 0.86. However, all results were higher than those for the *perceived technical competence* scale itself. These additional scale reliability results are listed in Appendix D.

It is therefore, not recommended that the *perceived reliability* construct simply be deleted or too quickly re-defined, because it was the weakest of the factors in this analysis. It is clearly an important component of HCT, although it remains at this time poorly understood as a construct in terms of its relationship to the other variables. It is recommended that the *perceived technical competence* construct be reconsidered and if possible more precisely defined. It also appears to be important to the overall validity and reliability of the HCT instrument and therefore can also not simply be ignored. A fuller discussion of these issues is provided in Chapter Five.

4.7 Summary

The new HCT instrument has been demonstrated to be a valid and reliable measurement for human-computer trust, with items showing good convergent validity on one over-arching factor, HCT, and the two underlying factors, *affect-based trust* and *cognition-based trust*.

Items have also been demonstrated to have good convergent and discriminant validities on three of the five scales proposed, *perceived understandability, faith, and personal attachment.*

The remaining two constructs, *perceived technical competence, and perceived reliability,* show sufficient convergent validity among their respective items to support their existence as important underlying HCT factors. Further work is needed to improve the validities of both of these scales and the reliability of the *perceived technical competence* items, in particular.

In spite of residual problems with the discriminant validity of some items, and the questions surrounding the construct definitions of *perceived reliability* and *perceived technical competence*, as an overall measure of HCT, this instrument, as it stands could be used with some assurance. Given that the five factor model accounts for **68.9%** of the variance in the data and that the reliability score for the overall human-computer trust scale was **0.94**, the **new HCT instrument can be considered to be a valid and reliable measure for the construct of *overall perceived human-computer trust* as intended.**

Chapter 5
Conclusion

This chapter presents an overview of the work performed for this study. This chapter begins by summarizing the theory and objectives of the study. Then, the various stages of instrument development undertaken in this study are summarized and the results are discussed in terms of their consistency with established instrument validation objectives. Next, the research conclusions are presented. This is followed by a discussion of the implications of the research findings for theory and practice. Finally, the limitations of this study and future research opportunities arising from this investigation are suggested.

5.1 Summary of the objectives and theory underlying this research

The primary aim of this research was to produce a psychometric instrument to measure human-computer trust. The objectives of the research were to clearly and precisely define human-computer trust, to develop a model for this construct, to develop an instrument based upon on this model and to demonstrate the usefulness of following a rigorous process for instrument development.

Human-computer trust specific to IDA use was defined within the trust framework developed in this thesis, as an instance of micro-trust comprised of both cognitive and affective components. As such, it was then more precisely defined as, *the extent to which a user is confident in, and willing to act on the basis of, the recommendations, actions, and decisions of an artificially intelligent decision system* (adapted from McAllister 1995, p.25). Previous research had failed to both specifically define the human-computer trust construct and to design a measure specifically for human-computer trust. Results from previous research were thus, inconsistent and at times contradictory.

The current research was based on the theories of social attribution and ethopoeia. The supposition that individuals attribute human characteristics to machines, just as they attribute other humans with certain characteristics based on their experiences with them, was an essential underlying perspective from which this investigation proceeded. Therefore, a secondary aim of this study was to demonstrate that human-computer trust could be measured from this perspective. As such the human-computer trust measured here is the perceived trust that the user has in a particular system. This perspective is also consistent with the trust framework proposed in this study.

This research provides an important practical foundation for the development and implementation of IDA as well as for the continuance of human-computer trust research. This study has produced a consistent definition for the construct, it has demonstrated a sound method for the development of psychometric instruments generally, and it has developed a valid and reliable instrument, based on the model, with which human-computer trust can be further investigated.

5.2 Summary of the stages of instrument development and testing

As previously discussed in Chapter Three, there are several criterion of quality with which psychometric instrument development practitioners are concerned. These are the equivalence reliability, the content validity, and the construct validity of the instrument. This study was designed to address these criteria in a logical, rigorous progression through the various stages of research. These criterion are therefore addressed here in the order in which they have been considered during the development process of the new HCT instrument.

5.2.1 Stage one: the nominal group

The first type of validity to be considered was content validity. Content validity is the measure of how representative the scales are of the construct they are intended to measure (Neuman 1994).

Content validity is improved by attempting to define constructs precisely and by sampling from all aspects of the definition for representative indicators. The nominal group helped to ensure that major aspects of the overall human-computer trust construct were addressed by identifying and defining a set of essential sub-constructs, or dimensions, of the primary variable, human-computer trust. In addition, the inclusion of sub-constructs and potential indicators from previous research assured that as many aspects of each sub-construct were covered as was possible.

5.2.2 Stage two: the sorting rounds

Content validity was also addressed, in part, in the second stage of this research by using the judging panels to sort items. Items that were judged not to fit the constructs, that is, items that did not appear to be valid for the construct for which they were designed, were discarded and constructs which failed to be discriminated by the judges were also dropped. This ensured that only the sub-constructs and indicators that were judged to be the most representative of their construct were retained.

The judging panels also addressed the face validity of the constructs and their indicators. Face validity is the simplest and least reliable of the validities. It simply refers to whether or not, in the judgement of others, the measures appear to be what they are intended to be. None-the-less, it can be a useful means by which initial constructs and items can be selected.

The iteration of the instrument through the judging process also provided the means with which the inter-rater reliability of the scales could be tested. This process primarily allowed the instrument to be refined and reduced to a minimal set of constructs and indicators. Since the inter-rater reliability was calculated on the overall results for the group of scales it was also a preliminary indication of the equivalence reliability, or internal consistency, of the instrument as a whole.

5.2.3 Stage three: the pilot study

The pilot study was intended to assist in the preparation of the survey as a whole. Participants were asked to comment on the format of the survey, any difficulties they had in answering any of the items, and any difficulties they had in understanding the instructions. From the pilot study it was also possible to make an estimation of the time needed to complete the survey.

The results from the pilot study were not able to be used to determine either the reliability or validity of the scales, as the sample size was too small. However, it was important to consider the overall usability of the instrument. Survey response rates are not generally high and a survey that is ill formatted and not understandable is even less likely to attract respondents.

5.2.4 Stage four: the main study

The main study was a field study and the data collected enabled quantitative measurement of both equivalence reliability and construct validity of the instrument and its individual scales.

The results for the internal consistency, calculated as Cronbach's alpha, for each of the five scales was particularly good. It can be seen that the equivalence reliability of the instrument after the final stage, calculated as Cronbach's alpha, 0.94, was higher than that calculated as Cohen's kappa at the end of satge three in the final sorting round, 0.83.

The final type of validity to be considered in this research was construct validity. Construct validity is concerned with whether or not the items converge, that is items measuring the same construct converge with each other and also whether or not items are discriminant, that is, items for one construct can be differentiated from items measuring different constructs. The judging panels of stage three were a preliminary means of improving the construct validity of the scales. In the final stage the construct validities of the scales were able to be quantitatively determined through the use of principal components analysis.

Each of the sub-constructs in the final set was expected to be a factor underlying the primary construct of human-computer trust. It was found that the indicators for the sub-constructs were highly correlated with one overall factor, HCT. Table 4.6 shows how all items loading significantly on the primary factor. It was also found that the two types of perceived trust, affect-based and cognitive-based trust existed as expected (Table 4.7).

When the five factor model was examined (Table 4.10) it was found that most of items loaded on the constructs for which they were designed. This finding supports both the convergent and discriminant validity of those items and scales. Eight items were observed to have unexpected factor loadings and with the exception of one *faith* item, these items belonged to only two of the five sub-constructs.

The two sub-constructs, *perceived technical competence* and *perceived reliability* to which most of these non-discriminant, non-convergent items belonged were given further consideration with regard to their definitions and the way in which their respective items loaded on other constructs. It was proposed that the definition for *perceived technical competence* might require some refinement and that the items for this construct required rewording or replacement to improve the convergent validity of these indicators. The *perceived reliability* construct appeared to be more stable and highly correlated with *personal attachment* and *faith*. The definition for this construct was seen to be satisfactory but alterations to some of the items in terms of their wording have been suggested.

The first overarching factor found in this study accounts for **43%** of the variance in the data (Table 4.11). Selecting factors with eigenvalues greater than one, which is the default analysis criteria used in SPSS8© indicated that only four factors were needed to describe HCT which together accounted for **60.3%** of the variance in the data. However, since the selection of the number of factors to retain is an arbitrary one (Dunteman 1989), in the case of this study it was reasonable to retain the first five factors, in accordance with the proposed HCT model, which together explained approximately **69%** of the variance in the data.

5.3 Conclusions about the research findings

Table 5.1 Conclusions about the expected research outcomes

Expected Relationship	Definition	Supported
User's overall level of trust in the IDA		
E 1	The user's *cognition-based trust* is a component of the user's *overall perceived trust* in the system.	Yes
E 2	The user's *affect-based trust* is a component of the user's *overall perceived trust* in the system.	Yes
User's cognitive-based trust in the IDA		
E 3	The user's perception of the *technical competence* of the system is a component of the user's *cognition-based trust* in the system.	Yes
E 4	The user's perception of the *reliability* of the system is a component of the user's *cognition-based trust* in the system.	No
E 5	The user's perception of the *understandability* of the system is a component of the user's *cognition-based trust* in the system.	Yes
User's affect-based trust in the IDA		
E 6	The user's *personal attachment* to the system is a component of the user's *affect-based trust* in the system.	Yes
E 7	The user's *faith* in the system is a component of the user's *affect-based trust* in the system.	Yes

The results of the principal components analysis strongly support the first two expected relationships, E1 and E2, which state that the overall perceived trust that a user has in an IDA is comprised of both cognitive and affective components.

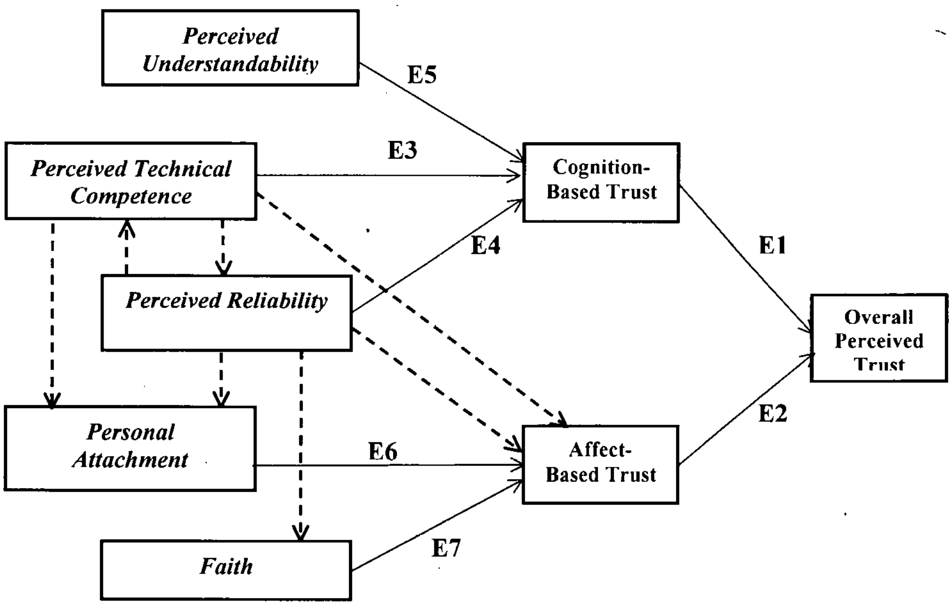
Also supported are the expected relationships that *cognition-based trust* is comprised of both the *perceived understandability* of the system and the *perceived technical competence* of the system, E3 and E5 . The expected relationship, E4, that the *perceived reliability* of the system is a component of the *cognition-based trust* is not supported in these findings.

The expectation that *affect-based trust* is comprised of both *faith* and *personal attachment* is strongly supported by the findings from this analysis. An unexpected result, for which there is no expected relationship, is that the *perceived reliability* of the system was observed to be a component of *affect-based trust.*

Perceived technical competence appears to be related to both *affect-based* and *cognitive-based trust.* This finding supports E3 in part, but also introduces a possible new relationship between *perceived technical competence* and *affect-based trust.* What remains unclear is whether or not the *perceived technical competence* and *perceived reliability* variables are related to affect-based trust directly or whether they relate to *affect-based trust* through their relationships with the *personal attachment* variable. The relationships among these variables are likely to be far more complex than proposed initially in the current research. Further investigation of the structure of HCT is, unfortunately, beyond the scope of this study.

There is insufficient evidence from this analysis upon which to conclusively alter the structure of the model of HCT as proposed initially in this study, due to the exploratory nature of the investigation. However, it is possible to depict a structure, Figure 5.1, based on the results of the principal components analysis. The detailed structure of human-computer trust is, as yet uncertain, and this model should not be seen as a final solution.

Figure 5.1 Possible Variable Relationships Indicated By TDS Results



5.3.1 Conclusions about the sub-construct scales

The *perceived understandability, personal attachment* and *faith* scales demonstrated good reliabilities and their respective items were found to have both convergent and discriminant validities. These scales, therefore, are satisfactory as they stand and do not require further consideration at this time. The *perceived technical competence* and *perceived reliability* scales did not fare quite as well.

The *perceived reliability* scale demonstrated sound internal consistency, but its items did not show good convergent validity on their own construct. These items actually converged more consistently on the *personal attachment* construct as seen in Table 4.8. The internal consistency of a scale derived from both the *perceived reliability* and the *personal attachment* items was particularly high at 0.92. This and other sub-scale combinations, for example scales that comprise the expected and observed component structure of *affect-based* and *cognitive-based trust*, are included in Appendix E.

While a user's perception of the reliability of the system and their personal attachment to the system appear to be quite distinct concepts these results would suggest that there may be a close relationship between them. If a strong positive relationship is found then it may be sufficient to drop the reliability construct and to simply measure personal attachment because the reliability items could then be seen to be redundant. This decision would be consistent with the result of the four principal components (Table 4.5) suggested by the default analysis criteria of SPSS8©.

At this point, however, it is believed that the *perceived reliability* items are important to the content validity of the overall scale and should, therefore, be retained. The conclusion about this scale, then, is to try to improve the convergent validity of its items by clarifying the wording of the items, as suggested in Chapter Four, and to re-test the instrument.

The *perceived technical competence* scale had the lowest internal consistency of any other scale in this study, although at 0.74 it might still be considered adequate. These items did not converge well on their own construct nor did they converge well on any other construct. From the two factor model this construct appears to be related to both cognitive and affect-based trust. There were clearly some problems with the interpretation of its items, T4 in particular could be seen to be the least discriminant of any of the items in this instrument, as discussed in Chapter Four. It is difficult from the limited amount of information available from one set of data to draw any hard conclusions about how to deal with this construct. The best suggestion that can be made at this point is that the items for this construct should be reworded and replaced as previously suggested in chapter four and then the instrument should be re-tested.

5.3.2 Conclusions about the rigorous approach to instrument development

The approach taken to the development of the new HCT instrument in this study has proven to be one which resulted in a set of scales that show high internal consistency and construct validity. The results from stage one suggest that beginning scale development on the ground by soliciting the opinions of a representative sample of the target group, in this case user's of IDA, provides a solid foundation from which to proceed. This ensures that the constructs chosen for investigation are indeed those which are relevant to the study.

The second stage of the research was also invaluable in ensuring that the scales developed are reliable and valid. The output of the various rounds of sorting required some interpretation on the part of the researcher. This was a concern at the time because there appeared to be a danger of introducing bias during this stage particularly in the two cases where the judges did not have the constructs. With none of these judges creating five categories, and few of them naming their categories with names similar to those expected, decisions had to be made about what constructs these judges were actually defining. The strategy adopted in these cases was to interpret the categories based on their best fit to those expected. As well as this, the overall strategy in this stage of the research was to allow the placement data to determine the fate of the items, that is whether they would be changed or dropped from the scales. Furthermore, the inter-rater reliability (Cohen's kappa) is a measure of equivalence reliability which would be expected to be comparable to the internal consistency of the scale (Cronbach's alpha) in the final stage which is also a measure of equivalence reliability.

Stage three, the pilot study, was useful in assessing the usability of the survey as a self-report mechanism. Although this stage is usually incorporated with the first quantitative data collection, it was a valuable exercise to assess the survey format for its ease of use and adequacy in terms of information provided and instructions prior to using it in the main study for the collection of quantitative data.

The field study, stage four, was of particular interest because most instrument testing and the majority of human-computer trust research to date has been performed under experimental conditions. Although it was difficult to get adequate numbers of responses to the survey from the selected groups, the data that was collected was perhaps more valuable because the participants were in fact regular users of IDA. As such, they had had sufficient time and experience to develop their opinions about the system. Most respondents had at least one year's experience using the system. This condition could not be easily replicated in an experimental design.

Because there were also several small groups of participants, in terms of the companies that they worked for, spread across several states, it is possible to be optimistic that the results of this study, that is the validity of the instrument, will be found to be applicable to various groups of IDA users. There were also sub-groups based on the type of system being used and the type of job being performed. An analysis of the between group differences using the Mann-Whitney U test for comparison of sample means in these groups showed that there were no significant between group differences.

5.4 Implications of the research for theory and practice

The theory of ethopoeia as described by (Nass et al. 1993) suggests that rational humans using computers attribute human characteristics to the machines as if they were in fact sentient beings. The results from Nass et al. (1993) show that these attributions are not made indirectly as a consequence of attributions about the creator of the computer nor are they indicators of dysfunctional behavior in the users themselves. The machine is related to in a way which is similar to the way in which humans relate to each other.

In this study, respondents were able to answer questions relating to both their perceptions about the IDA they were using and their feelings about the machine. This observation supports the theory of ethopoeia.

Social attribution theory is concerned with the way in which individuals make attributions about the behavior of others. Behavior can be explained as being due to either internal or external causes. Behavior attributed to internal causes relates an individual's behavior to a stable characteristic of the individual while behavior attributed to external causes relates an individual's behavior to the situation in which that individual is observed.

These attributions are based on both the level of information that the attributor has in a given circumstance (Kelley 1973). The level of information is generally restricted by the nature of the context within which the behavior of the other is observed. It is unusual for attributors to have complete information from which to make an attribution. Without complete information attribution errors occur. An attribution error is a psychological shortcut that allows explanations to be given without the need to seek out all the relevant information (Bernstein et al. 1991). The fundamental attribution error, discussed in Chapter Two, is the tendency to attribute behavior to internal causes.

In the context of IDA use it would also be unusual for the user to have all the information that they needed to make accurate attributions. Thus, it is possible that IDA users also have a tendency to attribute the system's behavior to *internal* *characteristics*. The potential for IDA users to make inaccurate attributions because of a lack of information has several implications for HCT research.

With respect to the current study, it may help to explain why the affective components of HCT were seen to be stronger than the cognitive ones. A user's *personal attachment* to the system and their *faith* in the system are not characteristics of the system attributed to it by the user. These constructs are specific feelings the user has about the system. By definition, then, these affective components of HCT are "beyond attributional activity" (Lerch et al. 1993, p. 7). Thus, it appears to have been easier for users to judge how they *felt* about using a particular system than it was for them to make attributions about the system's behavior. This phenomenon resulted in more consistent responses to the affective items than to the cognitive items in this study.

The implication of the fundamental attribution error on perceived human-computer trust, and system use, manifest trust, is particularly important to consider. If a user inaccurately attributes the system with characteristics it does not have, which in turn impact on the level of trust that the user has in the system, then, it is likely that their level of trust in the system will not be congruent with the actual trustworthiness of the system. They would therefore be said to ***mistrust*** the system. This mistrust may be positively or negatively biased and as Muir (1987) has accurately pointed out mistrust may result in ***misuse*** of the system.

Muir's (1987, 1994, 1996) general hypothesis was that trust could be calibrated in an effort to improve the use of IDA. Calibration suggests that the user's level of trust in a particular system might be changed to be more congruent with the actual trustworthiness of the system.

Trust researchers have found that once a level of trust has been reached it is difficult to change (Lerch et al. 1997, Rempel et al. 1985). Calibration of a user's trust in a system, then, would appear to be a difficult task to accomplish once the user has reached a particular level of trust. However, if it were possible to provide the user with the appropriate evidence from which they could make attributions without cognitive biases, such as the fundamental attribution error, then it may be possible to calibrate their trust appropriately and avoid misuse of the system.

This has broad implications for system's designers and developers, who might be able to positively exploit attributional bias to better calibrate the users' trust in their systems or to design systems that provide the necessary information to the user from which they can make less biased attributions. For example, Lerch et al. (1997) provided users with information that the source of the knowledge used in the IDA was a human expert, for whom they also had a description detailing his experience, and found that the user's confidence in the IDA increased. Of course, manipulating the users in this way is only satisfactory if the system is in fact trustworthy.

5.5 Limitations of the research

Some limitations of this study result from difficulty of finding participants for the various stages of the study.

The initial stage involved the use of the Nominal Group Technique to identify factors which should be investigated. The group in this study was comprised of only four people. It is recommended that these groups be comprised of eight to twelve people. Ideally, then there would have been twice as many participants in this phase and they would have been people who use Intelligent Systems on a regular basis. As regular user's of IDA were not available to this researcher, people who had had some experience with IDA were asked to participate. The set of constructs from this nominal group however, proved to be very similar to those already under consideration from previous research. This suggests that the nominal group technique is robust to group size and therefore a very useful tool in this type of investigation.

The second phase of the research also suffered from a less than ideal, in terms of research, pool of participants from which to choose. Ideally the various groups would have been comprised again of expert user's of IDA. As experts were not available, the groups were comprised of staff and research students from Central Queensland University, some of whom had relatively little experience using computers. This was a concern during this phase of the study. However, the quality of the results from this phase proved that the process of sorting items into concepts can be adequately performed by people outside the field of research to which the concepts belong. In fact as has been mentioned in Chapter Three the group which consisted of expert computer users and practitioners had the most difficulty in sorting and grouping the items.

The third phase of the research, the pilot study, was designed to test the format of the survey rather than the validity of the constructs. As such there was no empirical analysis performed on the responses to the HCT items. The purpose of this stage was to make the survey as a whole as easily read and understood as possible. The fourteen students who participated in this stage provided adequate response in terms of any difficulties they had with the survey. There were few changes which were made to the survey as a result of this stage. It is believed however, that this was an important step in the overall design process for the instrument.

5.5.1 Statistical conclusion validity

The fourth stage again proved a problem in terms of sampling numbers. The Taxi-Dispatch systems, while providing a pool of some 900 users from which approximately 400 were asked to participate resulted in only 75 useable responses. Generally, it is expected that there will be a return rate of some 25% from survey research. This study failed to reach this expected response rate. This appears to be mainly due to the organizational environment and not directly related to the survey itself.

For example, one group who received the surveys just at the time when they were fund-raising for a charity used the $1 coins attached to the surveys as their donations. This added incentive, while outside the control of the researcher, resulted in a 64% response rate from this group. This would suggest that providing some real incentive to respondents could have greatly improved the response rate.

It would have been satisfying to have at least achieved the 25% expected response rate from the overall sample of respondents. This would have resulted in 100 cases which is just 25 short of the optimum minimum number of sample cases of 125. It is suggested that to ensure sampling adequacy there be at least five cases for each item in the instrument. With twenty-five items the desirable number of cases was 125.

The results from the KMO test for sampling adequacy and Bartlett's test of spherecity as reported in Chapter Four indicated that the sample, such as it was, was adequate for a principal components analysis.

5.5.2 Internal Validity

Threats to internal validity in this study were few due to the exploratory nature of the study. Internal validity deals with the possibility that the effects found through analysis were due to a set of unhypothesized or unmeasured variables (Straub 1989). Since this study as a whole was designed to develop a psychometric instrument from the ground up, its primary purpose was to ensure that the variable of interest, human-computer trust was being measured. The main study was based on a principal components analysis of data collected with the carefully designed instrument. This resulted in principal components which could be said to be the variables that were expected to be found. It is unlikely that these factors were other than those proposed or that anything other than human-computer trust was measured.

For example, it might be suggested that since the dominant factor appeared to be the Personal Attachment factor, the instrument could be seen to be measuring another version of the popular "User-friendliness" construct investigated by so many information systems researchers. Interpreting this over-arching factor as user-friendliness however is not a sufficient explanation of the findings. Since the scales for the sub-constructs, which were highly correlated with the overall factor, included items dealing with the correctness of the system, the reliability of the system, the user's faith in the system and the user's understanding of the system this over-arching predominant factor must be seen to be more than simply user-friendliness.

It is reasonable to conclude that the construct of perceived human-computer trust is being measured with this instrument and that the internal validity of this study was satisfactory.

5.5.3 External validity

A test of the instrument's external reliability is beyond the scope of this study but it is hoped that this instrument or its derivatives will be shown by future research to be reliable across various user groups and various IDA.

Given that the predictive validity of this instrument can be considered to be quite high. It is reasonable to expect that the instrument would perform equally as well were other systems to be investigated. The prospective HCT researcher should be cautioned, however, that with trust relationships being sensitive to situational influences if the IDA application they are interested in differs considerably from the one discussed here, it might behoove the researcher to reassess the instrument by following the development process suggested here.

Users of two different TDS systems were in fact sampled in this investigation. The two systems included in the study, were the Expertech system and the Raywood system. These systems are very similar in design and implementation with only minor operational differences. The most significant operational difference is that the Expertech system employs a naming convention for each area of the town the while Raywood system employs a numerical coding system. The former is potentially the easiest to learn to use since the naming convention is based on the names of the suburbs in the town. The latter requires that the users of the system learn the set of numbers and the corresponding suburbs to which they refer.

There were also reported differences between the vendors of these systems which may have affected the original purchase choice by the various boards of directors. It is unlikely however that these vendor differences would have affected the users of the systems generally, as only those few holding positions on the boards of directors had input into the original decision about which system to purchase.

Of the 75 respondents 12 were using the Raywood system. All others were using the Expertech TDS. Based on the Mann-Whitney U test for independent means reported in Chapter Four there was no significant difference between the users' of these two systems.

The issue of context specificity is of vital concern in instrument development in terms of the external validity, or representative reliability, of the measurement. In constructing scales of this nature items may become too specific for one particular group of respondents and therefore not be valid for other groups.

This is a particularly important issue in this research since IDA come in many shapes and forms and have, by their very nature, high specificity for their implementation. As discussed above, some care must be taken to diminish the effects of context specificity by carefully wording the items. Simple, concise statements that avoid the use of context specific terminology or events, like those in the *personal attachment* scale are likely to perform equally as well in a variety of applications. On the other hand, one runs the risk of becoming so general as to be irrelevant. It is thus a fine balance which is primarily determined by the intended use of the instrument under development. This is precisely why, even well tested instruments require validation before being adopted or adapted into new research applications.

5.6 Opportunities for Future Research

This study has laid the foundation from which much future research can proceed. The two main areas of interest are the investigation of the structure of HCT via a confirmatory factor analysis and the investigation of the factors involved in the development of trust in various contexts. While some work has been performed on the latter problem, the reliability of the results from these studies remains uncertain because researchers have tended to use untested, modified psychometric instruments developed specifically for interpersonal trust rather than ones developed specifically for HCT. The instrument developed here will allow work of this nature to become more consistent and reliable.

The structure of HCT requires a more detailed investigation through confirmatory factor analysis on the instrument developed here and a path analysis so that the relationships between the variables can be better understood.

The investigation of the development of trust is of particular interest. User and system characteristics influencing the degree of trust could be explored. Will (1992, 1994) and Lee and Moray (1994) have already suggested that user characteristics such as self-confidence affect the degree of trust of the users. Lerch et al. (1993) have also suggested that the user's perceived locus of control effects their attributions about the system. It has been suggested here that attributional theory may help to explain how users decide how much trust they have in a particular system. The application of this theory to human-computer relationships, generally, and human-computer trust, in particular, also requires further investigation and development.

A third area of interest is the nature of the transition from perceived trust to manifest trust. Kee and Knox (1970) suggested that there is a threshold, specific to the individual and the context, at which perceived trust manifests itself in action. The new instrument provides the ability to measure HCT reliably and manifest trust can easily be determined by observation. It should now be possible to investigate the factors that influence this transition within the context of computer-aided decision-making. A better understanding of this transition through such an investigation could provide valuable and useful information to IDA researchers and developers alike.

Finally, areas of current interest and some considerable importance are web-based applications and trust in the context of electronic commerce. Investigations into trust in these areas may require the development of a specific set of scales for this type of trust. It has been noted in this study that the components of trust are dependent on the situational context in which they develop. The current HCT scale has been developed to measure human-computer trust in IDA. The dimensions of HCT may be quite different to those involved in trust within the context of web-based applications and e-commerce in particular.

For example, the computer in the latter context may be seen to mediate the trust between the user and the vendor rather than being the direct object of the user's trust. Thus, some web based applications may involve computer-mediated trust between individuals as well as, or instead of, human-computer trust. This investigation might best begin by following the development process demonstrated in the current study to establish whether the essential factors involved differ from those found to be involved in human-computer trust.

Appendix A.

Trust Scales from previous HCT research

1. Lerch et al. (1993) - Human-Machine (Expert System ES) Trust
   (a) ES are unpredictable from one problem to the next P (dropped)
   (b) ES behave in a very consistent manner P
   (c) ES give same advice for the same situation over time P
   (d) ES provide good advice across different situations D
   (e) ES are dependable in important decisions D
   (f) Even when ES give unusual advice, users are confident that the advice is correct D
   (g) ES have proven to trustworthy D (dropped)
   (h) ES make users feel secure in new situations F (dropped)
   (i) Users with little expertise trust the advice form ES F
   (j) Users know ES were developed to help them make good decisions. F

KEY: R-eliability, D-ependability, F-aith

2. Lerch et al. (1993) - Attributions about specific ES and study characteristics
   (k) Problems have too many factors S
   (l) Problems are tricky S
   (m) Problems are easy S
   (n) Large amounts of business and economics knowledge K
   (o) Good knowledge about country risk analysis K
   (p) Large amounts of knowledge K
   (q) Use of all available knowledge R
   (r) Identification of important and relevant factors R
   (s) Use of thorough and consistent analysis R
   (t) Good reasoning capabilities R

KEY: S-ituation, K-nowledge, R-easoning

3. Lee and Moray (1994) - Human-Machine Trust.
   a) Specific
      (i) How high was your self-confidence in controlling the feedstock pump?
      (ii) How much did you trust the automatic controller of the feedstock pump?
      (iii) How high was your self-confidence in controlling the steam pump?
      (iv) How much did you trust the automatic controller of the steam pump?
      (v) How high was your self-confidence in controlling the steam heater?
      (vi) How much did you trust the automatic controller of the steam heater?
   a) General
      (vii) Trust in the local bus service to get you to the store on time./Self-confidence in your ability to get to the store on time.

(viii) Trust in your calculator or computer to produce the right answer./Your self-confidence in your ability to arrive at the correct answer doing the calculations manually.

(ix) Trust in the heating system where you live to keep you comfortable./Your self-confidence in your ability to turn the heater on and off manually to keep you comfortable.

(x) Trust in you watch to tell the correct time./Your self-confidence in your ability to estimate the correct time.

(Self-confidence here being what I am calling judgment confidence)

4. Muir and Moray (1996) - Human-machine Trust

The "pump" referred to below is an automated control feedpump

a) To what extent does the pump perform its function properly? (Competence)

V.C. -To what extent does it produce the requested flow rates? (Competence)

b) To what extent can the pump's behavior be predicted from moment to moment? (Predictability)

c) To what extent can you count on the pump to do its job? (Dependability)

d) To what extent does the pump perform the task it was designed to do in the system?(Responsibility)

V.C. -To what extent does it maintain system volume? (Responsibility)

e) To what extent does the pump respond similarly to similar circumstances at different points in time? (Reliability over time)

f) Your degree of faith that the pump will be able to cope with other system states in the future.

g) Your degree of trust in the pump to respond accurately.

h) Your degree of trust in the pump's display.

i) Your overall degree of trust in the pump.

j) VC - Verbal Clarification

Appendix B.

Nominal Group Materials

B.1. INFORMATION SHEET

Networked Information and Electronic Commerce project (NIEC)

Part of the MECIS program

(red Meat industry use of Electronic Communication and Information Systems, funded by the Australian Red Meat Research Corporation 1997-2001)

FOR PARTICIPANTS IN THE PROJECT:

We are investigating the use of electronic communication for information provision and electronic commerce in the Australian red meat industry. Areas we are looking at include:

Access to telecommunications in rural areas

The need for training in the use of electronic mail and the Internet

How useful information can be supplied electronically

How the internet can be used for marketing and electronic commerce

We will be gathering data by talking to people, by distributing survey forms and be focus group discussions.

We are aware that we need to respect the confidentiality of people involved in the project. In addition, as the project team involves people from Central Queensland University (CQU), we will abide by the guidelines for conducting research established by the Human Ethics Research Review Panel at that university.

Confidentiality

We will treat all information supplied in a confidential manner. No information will be attributed to any individual without their permission in writing. We will also exercise care where data that has been gathered from a particular group or organization is to be published. If it is possible to identify a group or organization from whom data has been collected, we will obtain their permission in writing before the data concerning their group is published.

Consent

By completing and returning a "Consent Form" you are indicating your consent to take part in the project. You are free not to take part, and may withdraw at any time.

Reports

We will publish results from the project at conferences and in reports to the Meat Research Corporation. With all focus groups, participants will be given the opportunity to indicate whether they would like a copy of a report arising form the activity to be forwarded to them.

Contact Details

Ms. Maria Madsen, School of Mathematical and Decision Sciences, CQU, Gladstone Campus, Bryan Jordan Drive, Gladstone, 4680. Phone: 4970 7246

Dr. Shirley Gregor, School of Computing and Information Systems, CQU, Rockhampton, 4702. Phone: 4930 9682

B.2. CONSENT and FEEDBACK FORM

Networked Information and Electronic Commerce project (NIEC)

FOR PARTICIPANTS IN A FOCUS GROUP

Before taking part in the focus group you are requested to sign this form, indicating that you give consent to be interviewed and for the taping and subsequent transcription of the interview.

During the focus group you will be asked questions about trust in computers. The focus group will last for approximately one hour and will be tape-recorded, transcribed and analysed. The tapes and notes of the focus group will not be shared with anyone outside the research team and no responses will be identifiable with any participants in any subsequent reports or publications.

Please read and sign the following before the focus group commences.

I understand that I can withdraw from the focus group at any time and that I will not be placed at risk as a consequence of my participation in this research.

I therefore agree to participate in the focus group and for the proceedings to be tape-recorded. Furthermore, I understand that at the completion of this research the tapes will be stored in a locked room in the Faculty of Informatics and Communications, Central Queensland University, Gladstone Campus. I understand that if the results are to be published my name will not be associated with the report.

I understand that I am free to refuse to answer any specific question and may withdraw from the focus group at any time, without penalty.

I have been given an information sheet with details of the project. I have also been given the opportunity to ask questions and received satisfactory answers about the research project.

THIS IS TO CERTIFY THAT I HEREBY AGREE TO PARTICIPATE AS A VOLUNTEER IN THE ABOVE NAMED RESEARCH.

--	--
(Participant's name – please print)	(Participant's signature)
--	--
(Researcher's signature)	(Date)

TEAR-OFF AND RETURN THIS SLIP IF YOU WANT FEEDBACK

To: Maria Madsen, Faculty of Informatics and Communications, Central Queensland University, Gladstone Campus, Gladstone 4680.

Please send me a report with results from the study in which I am participating.

Name: ______________________________

Address: ______________________________

B.3. INSTRUCTIONS

Defining Human-Computer Trust:

Human-Computer Trust (HCT) is the extent to which a user is confident in, and willing to act on the basis of, the communications, actions, and decisions of an intelligent computer system.

(adapted from McAllister, 1995, p.25)

The Focus Group Question:

What factors influence your trust in a computer system?

Example Scenarios:

1. The first example is based on the person's *perceptions* of another.

You have heard that a particular doctor is a very skilled surgeon. In fact your best friend was very pleased with the work that this doctor performed on her. Because of this knowledge about the doctor's skill and ability you decide that you would trust this doctor to perform an operation for you – you have confidence in his/her abilities. In this case knowledge about the doctor's ability leads you to perceive him/her as **competent** and to be willing to allow him/her to perform surgery for you.

The factor here is the need for you to know about the abilities and skill levels of the person to be trusted. The name of the factor is **competence.**

2. The second example is based on the person's *feelings* about another.

You now need an operation so you go to see the doctor from the first example. When you meet this doctor there is something about them that you do not like. He/she doesn't seem really interested in your problems, but he/she has asked if you are covered by medical insurance. You don't like this doctor at all and you decide that no matter how good they are as a surgeon you won't let him/her operate on you after all. In this case your knowledge of the doctor's competence has been overshadowed by the bad feeling you got when you met him/her. Because of your **dis-like** of this person you have lost confidence in them and become unwilling to have them operate on you.

The factor here is the feeling of dislike that you have about the person you need to trust. The name of the factor in this case would be **liking**.

Summary

You experience trust in another to varying degrees in varying situations depending on many factors.

For this group activity I want you to focus on a computer system, instead of a person, as the *other* that is to be trusted.

Write down as many of the factors that influence the way you, as a user, trust that computer system, as possible.

Name each of the factors. This was done in the preceding examples which gave *competence* and *liking* as two factors involved in trusting a physician.

Appendix C

Sorting Round Materials

C.1. INFORMATION SHEET

Networked Information and Electronic Commerce project (NIEC)

Part of the MECIS program

(red Meat industry use of Electronic Communication and Information Systems, funded by the Australian Red Meat Research Corporation 1997-2001)

FOR PARTICIPANTS IN THE PROJECT:

We are investigating the use of electronic communication for information provision and electronic commerce in the Australian red meat industry. Areas we are looking at include:

- Access to telecommunications in rural areas
- The need for training in the use of electronic mail and the Internet
- How useful information can be supplied electronically
- How the internet can be used for marketing and electronic commerce

We will be gathering data by talking to people, by distributing survey forms and by judging panel discussions.

We are aware that we need to respect the confidentiality of people involved in the project. In addition, as the project team involves people from Central Queensland University (CQU), we will abide by the guidelines for conducting research established by the Human Ethics Research Review Panel at that university.

Confidentiality

We will treat all information supplied in a confidential manner. No information will be attributed to any individual without their permission in writing. We will also exercise care where data that has been gathered from a particular group or organization is to be published. If it is possible to identify a group or organization from whom data has been collected, we will obtain their permission in writing before the data concerning their group is published.

Consent

By completing and returning a "Consent Form" you are indicating your consent to take part in the project. You are free not to take part, and may withdraw at any time.

Reports

We will publish results from the project at conferences and in reports to the Meat Research Corporation. With all judging panels, participants will be given the opportunity to indicate whether they would like a copy of a report arising form the activity to be forwarded to them.

Contact Details

Ms. Maria Madsen, School of Mathematics and Decisional Sciences, CQU, Gladstone Campus, Bryan Jordan Drive, Gladstone, 4680. Phone: 4970 7246

Dr. Shirley Gregor, School of Computing and Information Systems, CQU, Rockhampton, 4702. Phone: 4930 9682

C.2. CONSENT and FEEDBACK FORM

FOR PARTICIPANTS IN A JUDGING PANEL

Before taking part in the judging panel you are requested to sign this form, indicating that you give consent to be interviewed and for the subsequent transcription of the interview.

During the judging panel you will be asked to sort questions about trust in computers. The judging panel will last for approximately one and a half hours and the results will be analyzed. The notes of the judging panel will not be shared with anyone outside the research team and no responses will be identifiable with any participants in any subsequent reports or publications.

Please read and sign the following before the judging panel commences.

I understand that I can withdraw from the judging panel at any time and that I will not be placed at risk as a consequence of my participation in this research.

I therefore agree to participate in the judging panel and for the proceedings to be recorded. Furthermore, I understand that at the completion of this research the notes will be stored in a locked room in the Faculty of Informatics and Communications, Central Queensland University, Gladstone Campus. I understand that if the results are to be published my name will not be associated with the report.

I understand that I am free to refuse to answer any specific question and may withdraw from the judging panel at any time, without penalty.

I have been given an information sheet with details of the project. I have also been given the opportunity to ask questions and received satisfactory answers about the research project.

THIS IS TO CERTIFY THAT I HEREBY AGREE TO PARTICIPATE AS A VOLUNTEER IN THE ABOVE NAMED RESEARCH.

--	--
(Participant's name – please print)	(Participant's signature)
--	--
(Researcher's signature)	(Date)

TEAR-OFF AND RETURN THIS SLIP IF YOU WANT FEEDBACK

To: Maria Madsen, Faculty of Informatics and Communications, Central Queensland University, Gladstone Campus, Gladstone 4680.

Please send me a report with results from the study in which I am participating.

Name: ______________________________________

Address: ______________________________________

C.3. Card Sorting Instructions

This card sorting exercise is a manual test of construct validity and item discriminant validity. It is similar in nature to a factor analysis only performed manually. The purpose of this exercise is to reduce the items and refine the constructs before they are used in the pilot study. This prior refinement of the questions improves the quality of the data that is collected when the questionnaire is used.

Each judge is presented with a large pool of items. Four groups of judges will be asked to participate in the sorting phases. At each phase the group of judges will be presented with the refined set of items from previous phases. Two groups will have categories into which the items on the cards are to be sorted. The other two groups will be asked to sort the cards into categories of their own creation.

Instructions for those groups with unknown categories:

Sort the items on the cards into groups based on the similarity of the concept they appear to measure. As you are doing this try to name and define the concept to which you think these items relate. Any items which don't seem to fit into any category or which are not clearly understandable should be placed in a discard group.

Instructions for those groups with known categories:

Sort the items on the cards into the categories provided based on the similarity among the concepts to which the items relate. Any items which don't fit into any category or are, in your opinion, ambiguous, should be placed into the discard category.

Practice Sorting:

A practice sorting exercise will be performed by each judging panel and judges will be encouraged to ask questions and clarify their understanding of the process before the real sorting will begin.

C.4. Trial Sort Items for Familiarization Exercises

Taken from: Moore & Benbasat, 1992. "Adoption of Information Technology Innovation", *Information Systems Research,* vol. 2, no. 3.

Voluntariness – the degree to which use of innovation is perceived as being voluntary, or of free will, p.195.

1. My superiors expect me to use a personal workstation
2. My use of personal workstation is voluntary (as opposed to required by my superiors or job description).
3. My boss does not require me to use a personal workstation.
4. Although it might be helpful, using a personal workstation is certainly not compulsory in my job.

Relative Advantage – the degree to which an innovation is perceived as being better than its predecessor, p. 195.

1. Using a personal workstation enables me to accomplish tasks more quickly.
2. Using a personal workstation improves the quality of work I do.
3. Using a personal workstation makes it easier to do my job.
4. The disadvantages of my using a personal workstation far outweigh the advantages.
5. Overall I find using a personal workstation to be advantageous in my job.
6. Using a personal workstation increases my productivity.

Compatibility – the degree to which the innovation is perceived as consistent with the existing values, past experiences, and needs of the potential adopter, p. 199.

1. Using a personal workstation is compatible with all aspects of my work.
2. Using a personal workstation is completely compatible with my current situation.
3. I think that using a personal workstation fits well with the way I like to work.
4. Using a personal workstation fits into my work style

Ambiguous Items – taken from other constructs.

1. I have seen what others are doing with their personal workstation.
2. I know where I can go to satisfactorily try out various uses of a personal workstation.
3. I believe I could communicate with others the consequences of using a personal workstation.

C.5. Sorting Round One Item Placements

The number in the placement cell refers to the number of the construct to which the judge matched that item and corresponds to the numbers of the construct definitions as listed here. D refers to items that were discarded by the judges as not belonging to any construct or being ambiguous.

1. Reliability of the system, in the usual sense of repeated, consistent functioning.

ITEM	Judge 1	Judge 2	Judge 3	Judge 4
i) The .... uses thorough analysis methods.	6	D	2	5
ii) The .... uses consistent analysis procedures.	2	1	1	1
iii) The ....'s behavior can be predicted form moment to moment.	2	4	9	2
iv) The .... responds the same under the same conditions at different times.	1	5	1	1
v) I can rely on the .... to behave in certain ways.	1	1	1	1
vi) I expect the .... to be correct.	1	6	9	9
vii) I find that the .... is dependable especially when it comes to important, strategic decisions.	1	2	9	1
viii) I would like to avoid using the .... because it is unpredictable.	D	1	9	1

2. Robustness of the system, meaning demonstrated or promised ability to perform under a variety of circumstances.

ITEM	Judge 1	Judge 2	Judge 3	Judge 4
i) If I have to make an important decision I have never encountered before, I know the .... will provide accurate and appropriate support.	1	6	1	6
ii) I have to keep alert or the .... might not use appropriate methods and come to an incorrect decision.	D	9	1	D
iii) I can rely on the .... to perform up to the standard promised by the developers.	1	2	9	1
iv) I will use the .... for unfamiliar problems.	9	9	2	6
v) The .... provides good advice across different situations.	2	D	2	2
vi) I am willing to let the .... make decisions in situations that others might find risky.	6	6	9	9
vii) I know the .... will be able to support me in future tasks.	9	9	9	9

3. Familiarity, that is the system employs procedures, terms, and cultural norms which are familiar, friendly and natural to the trusting person.

ITEM	Judge 1	Judge 2	Judge 3	Judge 4
i) I am comfortable using the	3	3	3	3
ii) The ....'s interactive display guided me in a way which was natural to me.	3	3	5	3
iii) I can interact easily with the	3	3	4	8
iv) I am familiar with the patterns of behavior the .... has established.	3	3	5	3
v) The ....'s interactive display is concise.	D	1	5	4
vi) The .... provides feedback that makes sense to me.	3	3	4	3
vii) The .... is easy to use.	3	8	4	3
viii) The .... responds the way I want it to.	1	3	1	1

4. Understandability in the sense that the human supervisor or observer can form a mental model and predict future system behavior.

ITEM	Judge 1	Judge 2	Judge 3	Judge 4
i) I believe that the .... will do things that I don't understand.	9	D	4	4
ii) I feel the .... does everything it should for the decision problem.	1	1	2	2
iii) I know how the .... will behave the next time I use it.	1	4	1	4
iv) I understand how the .... will help me make a decision.	4	5	5	4
v) The .... will responds in ways that makes sense to me.	3	3	4	4

5. Explication of intention, meaning the system explicitly displays or says that it will act in a particular way (as contrasted to its future action having to be predicted from a model).

ITEM	Judge 1	Judge 2	Judge 3	Judge 4
i) If I needed help, I believe that the .... can provide both explanations and assistance.	4	D	4	9
ii) The .... tells me what it is going to do next.	5	5	5	5
iii) The .... will respond in the way that it says it will.	4	5	9	6
iv) I know what is going to happen next because the .... tells me.	5	3	5	5
v) I know what will happen if something I've done is not correct because the .... tells me.	4	9	5	9

6. Technical Competence of the .... meaning that the system is perceived to perform the tasks accurately and correctly based on the information that is input.

ITEM	Judge 1	Judge 2	Judge 3	Judge 4
i) The .... uses appropriate methods to reach decisions.	2	6	6	6
ii) I can count on the .... to make use of the information I give it.	9	6	9	1
iii) The .... uses the information I provide properly.	4	6	1	6
iv) If I disagreed with the ....'s solution I would believe the	9	D	9	9
v) If the .... tells me something is so, then it must be true.	9	9	9	9
vi) The .... has good quality knowledge about the problem area.	6	2	2	1
vii) The .... has good reasoning capabilities.	6	D	2	6
viii) The .... has large amounts of problem specific knowledge.	6	D	6	6
ix) The .... has large amounts of knowledge generally.	2	D	2	D
x) The .... has real expertise in solving problems.	2	D	2	1
xi) The .... is able to identify important and relevant factors in the decision process.	1	4	6	9
xii) The .... makes use of all the available knowledge.	6	6	2	2
xiii) I can count on the .... to do its job.	1	1	1	9
xiv) The .... always produces the requested result	1	1	2	1
xv) The .... addresses well the problem, for which it was designed.	6	2	6	D
xvi) The .... performs its function properly.	1	1	1	1

7. Integrity of the system in the sense that the system is able to recover from technical failures or user errors without loss of data.

ITEM	Judge 1	Judge 2	Judge 3	Judge 4
i) I believe the .... will tell me if I make a mistake.	9	D	5	4
ii) If I have to restart the .... after a technical problem for example, power failures, I am sure	7	7	7	1
iii) I am able to make corrections to the information I have entered, as I need.	D	7	7	7
iv) The .... automatically backs up my work so that if something goes wrong I can start again	7	7	7	6
v) If I make a mistake the .... will help me to correct it.	9	7	9	9

8. Personal Attachment to the .... comprised of: liking meaning that the user finds using the .... agreeable and it suits their taste and loving meaning that the user has a strong preference for the ...., is partial to using it and has an attachment to it.

ITEM	Judge 1	Judge 2	Judge 3	Judge 4
i) I like using the .... for decision making.	8	8	8	8
ii) The .... is designed to my liking.	8	3	8	8
iii) I find the .... agreeable to use in decision making.	8	D	8	8
iv) I find the .... is suitable to my style of decision making.	4	D	8	6
v) If given a choice, I am inclined to use the .... for decision making.	6	D	8	9
vi) I regard the .... favorably.	8	1	8	8
vii) It pleases me to use the .... for decision making.	8	D	8	8
viii) I would feel a sense of loss if the .... was taken away and I couldn't use it anymore.	8	8	8	8
ix) I prefer to make decisions with the .... than without it.	8	1	9	8
x) I am partial to using the .... for decision making.	8	D	8	8
xi) I love using the .... in decision making.	8	3	8	8
xii) If given a choice, I would definitely use the .... for decision making.	D	8	8	8
xiii) I have made a considerable emotional investment in learning to use the	D	D	8	8
xiv) I feel an attachment to the	8	8	8	8

9. Faith meaning that the user has faith in the future ability of the system to perform even in situations in which it is untried.

ITEM	Judge 1	Judge 2	Judge 3	Judge 4
i) I have a high degree of faith that the .... will be able to cope with other problems in the	9	9	9	9
ii) Even if I have no reason to expect the .... will be able to solve a difficult problem I still feel	9	9	9	9
iii) I know that the .... will make good decisions.	6	8	9	1
iv) Advice from the .... can be trusted even when I don't know for certain that it is correct.	9	9	9	9
v) Even when the .... gives unusual advice I am confident that the advice is correct.	9	9	9	9
vi) Even when the .... gives explanations which sound rather unlikely, I am confident that it	9	4	9	9
vii) Even when I don't know how the .... will react, I feel comfortable using it for a new	9	D	1	1

Appendix D

Pilot Study Materials

D.1. INFORMATION SHEET

FOR PARTICIPANTS IN THE PROJECT:

I am investigating attitudes that people have to using computerized decision support systems. Areas I am looking at include:

Identifying and classifying different types of human-computer trust

The factors that engender trust in computerized decision support systems

People's perceptions of computerized decision support systems.

I will be gathering data by talking to people, by distributing survey forms and by focus groups discussions.

I am aware that I need to respect the confidentiality of people involved in the project. In addition, as the project team involves people from Central Queensland University, I will abide by the guidelines for conducting research established by the Human Ethics Research Review Panel at the university.

Confidentiality

I will treat all information supplied in a confidential manner. No information will be attributed to any individual without their permission in writing. I will also exercise care where data that has been gathered from a particular group or organization is to be published. If it is possible to identify a group or organization from whom data has been collected, I will obtain their permission in writing before the data concerning their group is published.

Consent

By completing and returning the survey you are indicating your consent to take part in the project. You are free not to take part, and may withdraw at any time.

Reports

I will publish results from the project at conferences and in reports to the Faculty of Informatics and Communications at Central Queensland University. With all surveys, participants will be given the opportunity to indicate whether they would like a copy of a report arising from the activity to be forwarded to them.

Contact Details

Ms. Maria Madsen, School of Mathematical and Decision Sciences, CQU, Gladstone campus, Bryan Jordan Drive, Gladstone, 4680. Phone: 4970 7246.

Dr. Shirley Gregor, School of Computing and Information Systems, CQU, Rockhampton, 4702. Phone: 4930 9682

D.2. CONSENT and FEEDBACK FORM

Before responding to the following survey, you are requested to read this form.

After reading this form, you can give consent to participate voluntarily in the research and for your responses to be analyzed and reported by answering the questions on the survey.

In the survey you will be asked questions about your opinions and you will be asked to provide some personal details such as your age. Both your opinions and personal details are important in the analysis of the results. However, you will not be asked to disclose your name. Participant responses to the survey will not be shared with anyone outside the research team and no responses will be identifiable with any participants in any subsequent reports or publications. The survey will take approximately one half hour to complete.

Please read the following before you begin the survey.

I understand that I can discontinue my response to the survey at any time and that I will not be placed at risk as a consequence of my participation in this research.

I therefore agree to participate in the survey.

Furthermore, I understand that at the completion of this research the survey forms will be stored in a locked room in the Faculty of Informatics and Communications, Central Queensland University, Gladstone Campus. I understand that if the results are to be published my name or any other identifying features of my response will not be associated with the report.

I understand that I am free to refuse to answer any specific question and may withdraw from the survey at any time, without penalty.

I have been given an information sheet with details of the project.

BY RESPONDING TO THE ATTACHED SURVEY, I THEREBY AGREE TO PARTICIPATE AS A VOLUNTEER IN THE ABOVE NAMED RESEARCH.

TEAR-OFF AND RETURN THIS SLIP IF YOU WANT FEEDBACK

PLEASE SEND SEPARATELY TO THE SURVEY FORM

- -

To: Maria Madsen, Faculty of Informatics and Communications, Central Queensland University, Gladstone Campus, Gladstone 4680.

Please send me a report with results from the study in which I am participating.

Name: ______________________________

Address: ______________________________

____

____

D.3. Test Piece for Pilot Study

D.3.1 Word document with grammatical errors

Rustic Steele

[1]A yarn, a cool beer and an inspirational memory. It was raining – drought breaking rain. Seven years of drought and the day I drove south from Toowoomba through Steele Rudd Country, the only [2]thing dry was the humour of Arthur Hoey Davis.

Davis's stories are about the trials and tribulations of laconic settlers Dad and Dave Rudd [3]which Davis long ago etched into Australian folklore under the pen name of Steele Rudd.

Steele Rudd books, such as *On Our Selection,* are based on life in a tiny slab [4]hut on the Davis property. Take a detour off the New England Highway and at, East Greenmount, you will find Steele Rudd's hut – a small, dark and damp place. Life here must have been anything but pleasant.

Then, just as now, the people of the Darling [5]Downs who live in the black soil hamlets share something more than the backdrop of some good yarns mateship.

[6]Further down the track you will find a small village called Nobby, which is endowed with the art of living and the gift of friendship. Eugene and Vicky Hollis-Neath run Rudd's [7]pub, crammed with memorabilia depicting the life of Nobby's famous author, here you can yarn with the locals in the pub in which, some say, Davis penned his characters. Alternatively, you can stay a while in the pub's own bed and [8]breakfast cottage.

(Adapted from: Gilchrist, David, 1998. Rustic Steele, The Sunday Mail, September 27, 1998. p. 110.)

Note: Grammatical errors added for the purposes of this research.

D.3.1 Grammatical errors found by Grammar Checker

The style of grammar was set as Formal for this experiment. Following are the error statements provided by Word 97© corresponding to the numbered, underlined fragments in the essay piece above. The grammar checker made some suggestions and the Help facility provided examples of the error and suggested changes to the structure of the phrase or word.

Participants were instructed to read the suggestions and to seek help with each error and then to decide if they would change that particular error based on those suggestions.

1. Fragment - no suggestions

FRAGMENT: If the marked words are an incomplete thought, consider developing this thought into a complete sentence be adding a subject or a verb or combining this text with another sentence.

- Instead of: Meteors the entire night.
- Consider: We watched meteors the entire night.

- Instead of: A rose by any other name.
- Consider: A rose by any other name still smells sweet.

2. Suggestion: Thing dry was/things dry were

SUBJECT-VERB AGREEMENT: The verb of a sentence must agree with the subject in number and in person.

- Instead of: What was Stephen and Laura like as schoolchildren?
- Consider: What were Stephen and Laura like as schoolchildren?

- Instead of: Tom watch the snowy egret stab at the fish.
- Consider: Tom watches the snowy egret stab at the fish.

3. Suggestion: "That" or "Which"

If the marked group of words is essential to the meaning of your sentence, "that" to introduce the group of words. Do not use a comma. If these words are not essential to the meaning of your sentence, use "which" and separate the words with a comma.

- Instead of: Did you learn the dance, that is from Guatemala?
- Consider: Did you learn the dance, which is from Guatemala?
- Or Consider: Did you learn the dance that is from Guatemala?

- Instead of: We want to buy the photo which Harry took.
- Consider: We want to buy the photo, which Harry took.
- Or consider: We want to buy the photo that Harry took.

4. Passive Voice - no suggestions

For a livelier or more persuasive sentence, consider rewriting your sentence using an active verb (the subject performing the action; as in "The ball hit Catherine") rather than a passive verb (The subject receives the action, as in "Catherine was hit by the ball"). If you rewrite with an active verb, consider what the appropriate subject is - "they", "we", or a more specific noun or pronoun.

- Instead of: Juanita was delighted by Michelle
- Consider: The boss gave Eric more work.

- Instead of: The garbage needs to be taken out.
- Consider: You need to take the garbage out.

5. Capitalization (downs)

Capitalize the marked word if it is part of a title, the name of a place, or the name of a person. Do not capitalize minor words such as "of", "a", and "the" unless these words officially begin with name of the place or the title, as in "The Hague" or "the New Yorker."

- Instead of: The mean dog that mr. Crotchety owns can almost jump his fence.
- Consider: The mean dog that Mr. Crotchety owns can almost jump his fence.

- Instead of: She had a good time visiting the Hague.
- Consider: She had a good time visiting The Hague.

6. (Further,) Comma Use

To make you sentence easier to read and to signal a pause, consider using a comma to set off words or phrases (especially introductory words or phrases).

- Instead of: Unfortunately it rained the day of the picnic.
- Consider: Unfortunately, it rained the day of the picnic.

- Instead of: Once he got home he began to calm down.
- Consider: Once he got home, he began to calm down.

7. (Pub) Comma Use

If you have placed the subject of your sentence directly next to the verb, it is incorrect to use a comma to separate the subject and verb.

- Instead of: the dog, ate my homework again.
- Consider: The dog ate my homework again.

- Instead of: His excuses, were not very original.
- Consider: His excuses were not very original.

8. Verb Confusion (no suggestions)

For correct usage, you may need to reword your sentence by adding a preposition directly after the marked verb or by substituting a more appropriate verb. It is incorrect to put a direct object after the marked verb.

- Instead of; the new witness emerged the truth.
- Consider: The new witness emerged with the truth.
- Or Consider: The new witness told the truth.

D.4. PILOT STUDY SURVEY INSTRUCTIONS

Should you choose to participate, please read these instructions carefully and answer all the questions.

Questions which are difficult to understand can be marked with a question mark but should still be answered to the best of your ability. Please feel free to make comments anywhere on the survey. This is not a test of your skill it is an investigation of your opinions and any opinions you have about the survey will be valuable. There is space provided for your comments at the end of the survey and I would encourage you to use this to record any comments you may have either in regard to the survey itself or to aspects of your use of the decision support system that you believe were not covered in the survey.

All the questions have been carefully prepared and are meaningful in terms of the analysis. Each question has an associated numerical scale from 1 to 7 with 1 meaning that you *strongly disagree* with the statement and 7 meaning that you *strongly agree* with it. The numbers in between signify varying degrees from strongly disagree to strongly agree. You are asked to choose the number that best describes the way in which each statement applies to you. Your cooperation in this is much appreciated.

In order to make this survey as general as possible your particular decision support system will be referred to in the questions as "*the system*". For the purpose of analysis, however, you will first be asked to identify the system, that you will be using as your frame of reference in your responses.

System identification

1. Name and describe the system that you are using as your frame of reference for this survey.

For the remainder of this survey the decision support system you have identified will be referred to as, "the system"

For this pilot study the system is the **grammar checker** in Microsoft Word '97.

2. Describe the type of work that you do using the system named above:

For the remainder of this survey the work that you do with the system will be referred to as, the "problem" or "decision"

For this pilot study the problem is **correcting grammar**.

D.5. THE SURVEY QUESTIONS

Your opinions about the system

		Strongly Disagree						Strongly Agree
3.	I would feel a sense of loss if the system was unavailable and I could no longer use it.	1	2	3	4	5	6	7
4.	The system uses appropriate methods to reach decisions.	1	2	3	4	5	6	7
5.	I believe advice from the system even when I don't know for certain that it is correct.	1	2	3	4	5	6	7
6.	The system always provides the advice I require to make my decision.	1	2	3	4	5	6	7
7.	I feel a sense of attachment to using the system.	1	2	3	4	5	6	7
8.	When I am uncertain about a decision I believe the system rather than myself.	1	2	3	4	5	6	7
9.	I find the system suitable to my style of decision making.	1	2	3	4	5	6	7
10.	I know what will happen the next time I use the system because I understand how it behaves.	1	2	3	4	5	6	7

		Strongly Disagree						Strongly Agree
11.	I make my final decision based solely on the advice from the system.	1	2	3	4	5	6	7
12.	The system has sound knowledge about this type of problem built into it.	1	2	3	4	5	6	7
13.	The system performs reliably.	1	2	3	4	5	6	7
14.	The system responds the same way under the same conditions at different times.	1	2	3	4	5	6	7
15.	I understand how the system will assist me with decisions I have to make.	1	2	3	4	5	6	7
16.	I can rely on the system to function properly.	1	2	3	4	5	6	7
17.	Although I may not know exactly how the system works, I know how to use it to make decisions about the problem.	1	2	3	4	5	6	7
18.	If I am not sure about a decision, I have faith that the system will provide the best solution.	1	2	3	4	5	6	7
19.	The advice the system produces is as good as that which a highly competent person could produce.	1	2	3	4	5	6	7

		Strongly Disagree						Strongly Agree
20.	It is easy to follow what the system does.	1	2	3	4	5	6	7
21.	I like using the system when I'm working on this type of problem.	1	2	3	4	5	6	7
22.	The system correctly uses the information I enter.	1	2	3	4	5	6	7
23.	I make my final decision based solely on my own knowledge of the problem regardless of what the system says.	1	2	3	4	5	6	7
24.	I have a personal preference for making decisions with the system.	1	2	3	4	5	6	7
25.	When the system gives unusual advice I am confident that the advice is correct.	1	2	3	4	5	6	7
26.	Even if I have no reason to expect the system will be able to solve a difficult problem, I still feel certain that it will.	1	2	3	4	5	6	7
27.	I make my decision based on both my own knowledge about the problem and the advice from the system.	1	2	3	4	5	6	7
28.	The system analyzes problems consistently.	1	2	3	4	5	6	7

		Strongly Disagree						Strongly Agree
29.	The system makes use of all the knowledge and information available to it to produce its solution to the problem.	1	2	3	4	5	6	7
30.	I recognize what I should do to get the advice I need from the system the next time I use it.	1	2	3	4	5	6	7
31.	Before today, I have always used the grammar checker when using a Word Processor.	1	2	3	4	5	6	7
32.	Before today, I have always used the explanation facility when using the grammar checker.	1	2	3	4	5	6	7
33.	I find the explanation facility of the grammar checker useful.	1	2	3	4	5	6	7
34.	I find the explanation facility of the grammar checker understandable.	1	2	3	4	5	6	7

Demographic data

35. Age

❑ 15-24 years ❑ 35-44 years ❑ 55-64 years

❑ 25-34 years ❑ 45-54 years ❑ 65-75 years

36. Gender

❑ Male ❑ Female

37. For how many years have you been using computers?

❑ 1-4 years ❑ 5-9 years

❑ 10-14 years ❑ 15-20 years

Other please specify:

____

38. As an indication of your proficiency in English grammar please give your level of achievement in Grade 12 English or equivalent.

____

____

____

39. Are there any grammatical errors that you think the grammar checker missed?

❑ Yes ❑ No

If you answered "yes", please specify the errors

____

____

____

____

40. What persuaded you to *accept* the grammar checker's advice when you made the suggested changes?

____

____

____

41. What persuaded you to *reject* the grammar checker's advice when you ignored the suggested changes?

____

____

____

42. Comments

THANK YOU FOR YOUR PARTICIPATION AND RESPONSE.

Appendix E

Field Study Materials

E.1. INFORMATION FOR PARTICIPANTS IN THE PROJECT

I am investigating attitudes that people have to using computerized decision support systems. Areas I am looking at include:

Identifying and classifying different types of human-computer trust

The factors that lead to trust in computerized decision support systems

People's perceptions of computerized decision support systems.

I will be gathering data by talking to people, by distributing survey forms and by focus groups discussions.

I am aware that I need to respect the confidentiality of people involved in the project. In addition, as the project team involves people from Central Queensland University, I will abide by the guidelines for conducting research established by the Human Ethics Research Review Panel at the university.

Confidentiality

I will treat all information supplied in a confidential manner. No information will be attributed to any individual without their permission in writing. I will also exercise care where data that has been gathered from a particular group or organization is to be published. If it is possible to identify a group or organization from whom data has been collected, I will obtain their permission in writing before the data concerning their group is published.

Consent

By completing and returning the survey you are indicating your consent to take part in the project. You are free not to take part, and may withdraw at any time.

Reports

I will publish results from the project at conferences and in reports to the Faculty of Informatics and Communications at Central Queensland University. With all surveys, participants will be given the opportunity to indicate whether they would like a copy of a report arising from the activity to be forwarded to them.

Contact Details

Ms. Maria Madsen, School of Mathematical and Decision Sciences, CQU, Gladstone campus, Bryan Jordan Drive, Gladstone, 4680. Phone: 4970 7246.

Dr. Shirley Gregor, School of Computing and Information Systems, CQU, Rockhampton, 4702.

Phone: 4930 9682.

E.2. CONSENT

Before responding to the survey, you are asked to read the information about confidentiality and voluntary consent which follows.

After reading this, you can give consent to participate voluntarily in the research and for your responses to be analyzed and reported by answering the questions in the survey.

In the survey you will be asked questions about your opinions and you will be asked to provide some personal details such as your age. Both your opinions and personal details are important in the analysis of the results. However, you will not be asked to disclose your name. Participant responses to the survey will not be shared with anyone outside the research team and no responses will be identifiable with any participants in any subsequent reports or publications. The survey will take approximately one half hour to complete.

I understand that I can discontinue my response to the survey at any time and that I will not be placed at risk as a consequence of my participation in this research.

I therefore agree to participate in the survey.

Furthermore, I understand that at the completion of this research the survey forms will be stored in a locked room in the Faculty of Informatics and Communications, Central Queensland University, Gladstone Campus. I understand that if the results are to be published my name or any other identifying features of my response will not be associated with the report.

I understand that I am free to refuse to answer any specific question and may withdraw from the survey at any time, without penalty.

I have been given information with details of the project.

By responding to the attached survey, I thereby agree to participate as a volunteer in the above named research.

E.3. INSTRUCTIONS

Thank you for choosing to participate. Please read these instructions carefully and answer all the questions.

Questions which are difficult to understand can be marked with a question mark but should still be answered to the best of your ability. Please feel free to make comments anywhere on the survey. This is not a test of your skill. It is an investigation of your opinions. There is space provided for your comments at the end of the survey and I would encourage you to use this to record any comments you may have either in regard to the survey itself or to aspects of your use of the decision support system that you believe were not covered in the survey.

All the questions have been carefully prepared and are meaningful in terms of the analysis. Each question has an associated numerical scale from 1 to 7 with 1 meaning that you *strongly agree* with the statement and 7 meaning that you *strongly disagree* with it. The numbers in between signify varying degrees from strongly agree to strongly disagree. You are asked to choose the number that best describes the way in which each statement applies to you. Your cooperation in this is much appreciated.

In order to make this survey as general as possible your particular decision support system will be referred to in the questions as "*the system*". For the purpose of analysis, however, you will first be asked to identify the decision support system that you will be using as your frame of reference in your responses.

System identification: Computerized Taxi Dispatch System

43. Name and describe the computer system that you are currently using. The rest of your responses should be made with this system in mind.

____

____

____

For the remainder of this survey the decision support system you have identified above will be referred to simply as, "the system"

44. Describe the type of work that is done by the system named above.

____

____

____

For the remainder of this survey the work you have described will be referred to as, the "problem" or "decision".

E.4. THE SURVEY and FEEDBACK FORM

Your opinions about your system and the work you do with it.

		Strongly Agree						**Strongly Disagree**
45.	I would feel a sense of loss if the system was unavailable and I could no longer use it.	1	2	3	4	5	6	7
46.	The system uses appropriate methods to reach decisions.	1	2	3	4	5	6	7
47.	I believe the system even when I don't know for certain that it is correct.	1	2	3	4	5	6	7
48.	The system always provides the information I need to do my job.	1	2	3	4	5	6	7
49.	I feel a sense of attachment to using the system.	1	2	3	4	5	6	7
50.	When I am uncertain about a decision the system has made I believe the system rather than myself.	1	2	3	4	5	6	7
51.	I find the system suitable to my style of working.	1	2	3	4	5	6	7
52.	I know what will happen the next time I use the system because I understand how it behaves.	1	2	3	4	5	6	7
53.	I make my final decision based solely on the advice from the system.	1	2	3	4	5	6	7
54.	The system has sound knowledge about our rules built into it.	1	2	3	4	5	6	7
55.	The system performs reliably.	1	2	3	4	5	6	7
56.	The system responds the same way under the same conditions at different times.	1	2	3	4	5	6	7
57.	I understand how the system helps me in my job.	1	2	3	4	5	6	7
58.	I can rely on the system to function properly.	1	2	3	4	5	6	7
59.	Although I may not know exactly how the system works, I know how to use it to do my work.	1	2	3	4	5	6	7

		Strongly Agree						Strongly Disagree
60.	If I am not sure about a decision, I have faith that the system will provide the best solution.	1	2	3	4	5	6	7
61.	The decision the system makes is as good as that which a highly competent person could make.	1	2	3	4	5	6	7
62.	It is easy to follow what the system does.	1	2	3	4	5	6	7
63.	I like using the system when I'm working.	1	2	3	4	5	6	7
64.	The system correctly uses the information I enter.	1	2	3	4	5	6	7
65.	I make my final decision based solely on my own knowledge of the problem regardless of what the system says.	1	2	3	4	5	6	7
66.	I have a personal preference for working with the system.	1	2	3	4	5	6	7
67.	When the system makes an unusual decision I am confident that it is correct.	1	2	3	4	5	6	7
68.	Even if I have no reason to expect the system will be able to resolve a difficult problem, I still feel certain that it will.	1	2	3	4	5	6	7
69.	I make my decision based on both my own knowledge about the problem and the advice from the system.	1	2	3	4	5	6	7
70.	The system analyzes problems consistently.	1	2	3	4	5	6	7
71.	The system makes use of all the knowledge and information available to it to produce its solution to the problem.	1	2	3	4	5	6	7
72.	I recognize what to do to get the advice I need from the system the next time I use it.	1	2	3	4	5	6	7

Demographic data

73. Job Description: Please check more than one box if applicable.

❑ Operator ❑ Driver ❑ Owner ❑ Director

74. Please provide your age in years: ______________________

75. Gender ❑ Male ❑ Female

76. For how many years have you been using computers generally ?

❑ 1-4 years ❑ 5-9 years ❑ 10-14 years ❑ 15-20 years

Other please specify: ______________________________

77. For how many years have you been using the system identified in Question 1?

❑ 1-4 years ❑ 5-9 years ❑ 10-14 years ❑ 15-20 years

Other please specify: ______________________________

78. For how many years have you been doing the type of work identified in Question 2 without a computer system?

❑ 1-4 years ❑ 5-9 years ❑ 10-14 years ❑ 15-20 years

Other please specify: ______________________________

79. If possible, describe what persuades you to *accept* the system's decision ?

____
____
____

80. If possible, describe what persuades you to *reject* the system's decision ?

____
____
____

81. Are there any additional functions that you would like the system to perform ?

❑ Yes ❑ No

If you answered "yes", please specify the functions you would like to have available.

____
____
____
____

82. Personal comments about the survey.

____
____
____
____
____
____
____

4.1 Personal comments about the system.

____
____
____
____
____
____
____
____
____
____
____
____
____
____
____

THANK YOU FOR YOUR PARTICIPATION AND RESPONSE.

FEEDBACK REQUEST

If you would like a copy of the report which results from this survey please fill in the form below with your mailing details.

To ensure confidentiality this feedback request will be detached from the survey before analysis.

To: Maria Madsen, Faculty of Informatics and Communications, Central Queensland University, Gladstone Campus, Bryan Jordan Drive, Gladstone, 4680.

Please send me a report with results from the study in which I am participating.

Name: ______________________________

Address: ______________________________

____

____

E.5. Analysis of Scale Reliabilities as Cronbach's Alpha

E.5.1 Overall HCT Scale

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Alpha if Item Deleted
F1_1	75.1156	698.7940	.5748	.6321	.9400
F2_1	74.7707	702.3781	.5404	.7664	.9405
F3_1	75.4907	699.7959	.6558	.6733	.9386
F4_1	75.2641	705.8828	.6961	.7331	.9383
F5_1	75.0507	700.2874	.7120	.7776	.9380
P1_1	75.3707	692.1143	.6618	.6434	.9386
P2_1	75.5574	691.7729	.6755	.6464	.9384
P3_1	76.1156	688.8369	.7789	.7539	.9370
P4_1	76.6774	695.9648	.7534	.8642	.9374
P5_1	76.1707	693.3891	.7301	.8122	.9376
R1_1	76.2241	705.9045	.6058	.5871	.9393
R2_1	75.9574	696.3255	.7547	.7868	.9374
R3_1	76.0424	703.5919	.6348	.6434	.9389
R4_1	75.8907	700.5736	.7241	.7228	.9379
R5_1	75.7507	703.2560	.6105	.7209	.9393
T1_1	75.9534	703.3267	.6293	.6526	.9390
T2_1	76.1041	705.7978	.5799	.5998	.9397
T3_1	75.6774	696.8557	.6140	.7354	.9393
T4_1	76.9669	731.9650	.3954	.5638	.9417
T5_1	76.6696	727.8289	.4637	.6427	.9409
U1_1	76.8507	729.1600	.4157	.6749	.9415
U2_1	76.8241	708.1785	.6487	.7645	.9388
U3_1	77.3441	741.1170	.3339	.6549	.9421
U4_1	77.1274	725.2749	.5300	.6230	.9403
U5_1	77.0507	731.2549	.4279	.5819	.9413

Reliability Coefficients 25 items N of Cases = 75.0

Alpha = .9416 **Standardized item alpha = .9411**

E.5.2 Observed Affect-Based HCT Scale

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Alpha if Item Deleted
F1_1	49.1768	340.4322	.6202	.5375	.9360
F2_1	48.8319	339.3180	.6359	.6663	.9355
F3_1	49.5519	343.8233	.6688	.6002	.9343
F4_1	49.3253	345.3948	.7639	.7090	.9323
F5_1	49.1119	342.6135	.7554	.6793	.9322
P1_1	49.4319	338.9570	.6646	.5833	.9346
P2_1	49.6186	336.5881	.7105	.5976	.9332
P3_1	50.1768	337.6700	.7698	.7115	.9317
P4_1	50.7386	343.3478	.7323	.8086	.9328
P5_1	50.2319	340.3251	.7287	.7665	.9327
R1_1	50.2853	347.4898	.6286	.5129	.9353
R2_1	50.0186	343.8514	.7292	.7359	.9329
R3_1	50.1036	349.0502	.6066	.5273	.9359
R4_1	49.9519	344.4398	.7400	.6932	.9327
R5_1	49.8119	347.5753	.6016	.5752	.9361

Reliability Coefficients 15 items N of Cases = 75.0

Alpha = .9380 Standardized item alpha = .9398

E.5.3 Expected Affect-Based HCT Scale

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Alpha if Item Deleted
F1_1	32.7885	159.6560	.6241	.4857	.9158
F2_1	32.4436	157.3672	.6729	.6335	.9128
F3_1	33.1636	163.5440	.6412	.5187	.9140
F4_1	32.9369	163.5103	.7685	.6849	.9081
F5_1	32.7236	162.3018	.7402	.6058	.9090
P1_1	33.0436	158.9538	.6641	.5264	.9131
P2_1	33.2303	157.7454	.7025	.5469	.9107
P3_1	33.7885	159.7146	.7349	.6528	.9088
P4_1	34.3503	161.9350	.7387	.7690	.9090
P5_1	33.8436	159.5732	.7405	.7501	.9085

Reliability Coefficients 10 items N of Cases = 75.0

Alpha = .9191 Standardized item alpha = .9220

E.5.4 Observed Cognitive-Based HCT Scale

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Alpha if Item Deleted
T1_1	22.6415	87.4753	.4366	.4303	.8467
T2_1	22.7921	81.7804	.6025	.4828	.8303
T3_1	22.3654	82.8131	.5029	.5107	.8426
T4_1	23.6550	89.3891	.4873	.4061	.8403
T5_1	23.3577	87.8744	.5674	.4898	.8339
U1_1	23.5388	87.8003	.5263	.4887	.8371
U2_1	23.5121	82.0563	.7121	.6406	.8200
U3_1	24.0321	90.9430	.5240	.5950	.8381
U4_1	23.8155	86.6966	.6603	.4836	.8271
U5_1	23.7388	88.3116	.5641	.4235	.8343

Reliability Coefficients 10 items N of Cases = 75.0

Alpha = .8491 Standardized item alpha = .8559

E.5.5 Expected Cognitive-Based HCT Scale

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Alpha if Item Deleted
R1_1	39.3004	212.9278	.5126	.3934	.8921
R2_1	39.0338	204.7648	.7279	.7459	.8831
R3_1	39.1188	207.0694	.6392	.5910	.8867
R4_1	38.9671	209.2629	.6460	.6550	.8866
R5_1	38.8271	209.3823	.5615	.6188	.8901
T1_1	39.0298	210.6331	.5552	.5457	.8903
T2_1	39.1804	205.7839	.6317	.5480	.8870
T3_1	38.7538	205.8033	.5676	.6007	.8904
T4_1	40.0433	220.2018	.4577	.4250	.8936
T5_1	39.7460	216.6366	.5631	.5527	.8900
U1_1	39.9271	218.5133	.4787	.5269	.8929
U2_1	39.9004	208.4333	.6834	.6693	.8852
U3_1	40.4204	224.4028	.4336	.6379	.8943
U4_1	40.2038	216.0931	.6160	.5608	.8884
U5_1	40.1271	219.1430	.5127	.5226	.8917

Reliability Coefficients 15 items N of Cases = 75.0

Alpha = .8962 Standardized item alpha = .8971

E.5.6 Sub-Scale Combinations

E.5.6.1 Personal Attachment and Perceived Reliability

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Alpha if Item Deleted
P1_1	28.8701	147.5637	.6711	.5519	.9117
P2_1	29.0568	146.4669	.7083	.5840	.9093
P3_1	29.6150	145.5970	.8124	.6965	.9031
P4_1	30.1768	149.9236	.7616	.8028	.9064
P5_1	29.6701	148.7780	.7332	.7280	.9077
R1_1	29.7235	155.5962	.5791	.4102	.9163
R2_1	29.4568	151.3569	.7291	.7107	.9082
R3_1	29.5418	153.9769	.6228	.5038	.9139
R4_1	29.3901	151.3792	.7508	.6668	.9072
R5_1	29.2501	153.5608	.6023	.5020	.9152

Reliability Coefficients 10 items

Alpha = .9182 Standardized item alpha = .9194

E.5.6.2 Personal Attachment and Perceived Technical Competence

Scale	Scale Mean if Item Deleted	Corrected Variance if Item Deleted	Item-Total Correlation	Squared Multiple Correlation	Alpha if Item Deleted
T1_1	27.9467	127.1688	.6330	.4924	.8721
T2_1	28.0973	128.6604	.5676	.4678	.8770
T3_1	27.6706	128.1838	.5178	.3885	.8816
T4_1	28.9602	138.8127	.4222	.3846	.8854
T5_1	28.6629	138.5543	.4464	.5192	.8840
P1_1	27.3640	121.1800	.6927	.5341	.8675
P2_1	27.5506	122.9748	.6584	.5168	.8702
P3_1	28.1088	122.3981	.7530	.6822	.8632
P4_1	28.6706	124.5239	.7532	.7849	.8639
P5_1	28.1640	123.6205	.7208	.7140	.8656

Reliability Coefficients 10 items N of Cases = 75.0

Alpha = .8846 Standardized item alpha = .8835

E.5.6.3 Faith and Perceived Technical Competence

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Alpha if Item Deleted
T1_1	32.1467	116.5630	.5861	.4233	.8516
T2_1	32.2973	117.6221	.5323	.4583	.8561
T3_1	31.8706	110.0842	.6694	.5006	.8442
T4_1	33.1602	132.2116	.2288	.3029	.8757
T5_1	32.8629	126.6576	.4178	.4843	.8636
F1_1	31.3088	111.7879	.5996	.5016	.8510
F2_1	30.9640	111.2314	.6138	.6432	.8496
F3_1	31.6840	111.4413	.7210	.5878	.8402
F4_1	31.4573	118.1091	.6418	.5643	.8482
F5_1	31.2440	112.3434	.7682	.6454	.8375

Reliability Coefficients 10 items

Alpha = .8651 Standardized item alpha = .8628

E.5.6.4 Faith and Perceived Understandability

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Alpha if Item Deleted
U1_1	29.2184	95.3242	.4075	.4833	.8338
U2_1	29.1918	89.1367	.6017	.5992	.8164
U3_1	29.7118	98.6739	.3788	.5434	.8354
U4_1	29.4951	95.5451	.4689	.4310	.8287
U5_1	29.4184	95.8897	.4337	.4288	.8314
F1_1	27.4833	83.5495	.5831	.5297	.8187
F2_1	27.1384	84.7056	.5484	.6655	.8229
F3_1	27.8584	86.1079	.6089	.5303	.8149
F4_1	27.6318	88.1502	.6591	.6042	.8113
F5_1	27.4184	87.3557	.6322	.6303	.8130

Reliability Coefficients 10 items

Alpha = .8379 Standardized item alpha = .8393

Bibliography

Ajzen, I. (1991). "The theory of planned behavior." *Organizational Behavior And Human Decision Processes, 50*, 179-211.

Anderson, C.A., Lepper, M.R., & Ross, L. (1980). "Perseverance of social theories: The role of explanation in the persistence of discredited information." *Journal of Personality and Social Psychology, 39*(6), 1037-1049.

Anderson, L.A., & Dedrick, R.F. (1990). "Development of the trust in physician scale: A measure to assess interpersonal trust in patient-physician relationships." *Psychological Report, 67*, 1091-1100.

Anderson, U., & Wright, W.F. (1988). "Expertise and the explanation effect." *Organizational Behavior and Human Decision Processes, 42*, 250-269.

Andrews, L.W., & Gutkin, T.B. (1991). "The effects of human versus computer authorship on consumers' perceptions of psychological reports." *Computers in Human Behavior, 7*, 311-317.

Bailey, K.D. (1978). *Methods of social research.* London: Collier Macmillan Publishers.

Barber, B. (1983). *The logic and limits of trust.* New Brunswick: Rutgers University Press.

Barfield, W., & Robless, R. (1989). "The effects of two- or three-dimensional graphics on the problem-solving performance of experienced and novice decision makers." *Behaviour and Information Technology, 8*(5), 369-385.

Begg, I.M., Anas, A., & Farinacci, S. (1992). "Dissociation of processes in belief: source recollection, statement familiarity, and the illusion of truth." *Journal of Experimental Psychology: General, 121*(4), 446-458.

Benbasat, I., Goldstein, D.K., & Mead, M. (1987). "The case research strategy in studies of information systems." *MIS Quarterly*(September), 369-385.

Benbasat, I., & Todd, P. (1996). "The effects of decision support and task contigencies on model formulation: A cognitive perspective." *Decision Support Systems, 17*, 241-252.

Bernstein, D.A., Roy, E.J., Srull, T.K., & Wickens, C.D. (1991). *Psychology.* Boston: Houghton Mifflin Company.

Birnbaum, M., & Stegner, S.E. (1979). "Source credibility in social judgement: bias, expertise, and the judge's point of view." *Journal of Personality and Social Psychology, 37*(1), 48-74.

Boose, J.H. (1989). "A survey of knowledge acquisition techniques and tools." *Knowledge Acquisition, 1*, 3-37.

Botterbusch, K.A. (1986). *A comparison of computerized job matching systems.* Menomonie Wisconsin: Materials Development Center, Stout Vocational Rehabilitation Institute, School of Education and Human Services, University of Wsconsin - Stout.

Brickman, P., Becker, L.J., & Castle, S. (1979). "Making trust easier and harder through two forms of sequential interaction." *Journal of Personality and Social Psychology, 37*(4), 515-521.

Briscoe, M.E., Woodyard, H.D., & Shaw, M.E. (1968). "Personality impression change as a function of the favorableness of first impressions." , 343-357.

Burgess, R.L., & Huston, T.L. (1979). *Social Exchange in Developing Relationships*. New York: Academic Press.

Carroll, J.M., & Campbell, R.L. (1989). "Artifacts as psychological theories: the case of human-computer interaction." *Behaviour and Information Technology, 8*(4), 247-256.

Carterinicchio, R.P. (1979). "Testing plausible path models of interpersonal trust in patient-physician treatment relationship." *Social Science and Medicine, 13 A*, 81-99.

Cecez-Kecmanovic, D. (1994). "Engineering type information systems research: A discussion on its position and quality". In *Australasian Conference on Information Interchange*, Melbourne: Unpublished.

Chen , J.J., & Ko, M. (1994). "The disability index analysis system via an ergonomics, expert systems, and multiple attribute decision-making process." *International Journal of Industrial Ergonomics, 13*, 317-355.

Chen, H. (1995). "Machine learning for information retrieval: neural networks, symbolic learning, and genetic algorithms." *Journal of the American Society for Information Science, 46*(3), 194-216.

Clancey, W.J. (1983). "The epistemology of a rule-based expert system - a framework for explanation." *Artificial Intelligence, 20*, 215-251.

Coakes, S.J., & Steed, L.G. (1996). *SPSS for windows: Analysis without anguish.* Milton: John Wiley & Sons.

Cohen, P., Cohen, J., Teresi, J., Marchi, M., & Velez, C.N. (1990). "Problems in the measurement of latent variables in structural equations and causal models." *Applied Psychology Measurement, 14*(2), 183-196.

Connors, M.M., Harrison, A.A., & Summit, J. (1994). "Crew systems: Integrating human and technical subsystems for the exploration of space." *Behavioral Sciences, 39*, 183-212.

Cook, T.D., & Campbell, D.T. (1979). *Quasi-Experimentation Design & Analysis Issues for Field Settings*. Boston: Houghton Mifflin Company.

Cooke, N.J. (1994). "Varieties of knowledge elicitation techniques." *International Journal of Human-Computer Studies, 41*(6), 801-849.

Cronbach, L.J. (1949). *Essentials of Psychological Testing*. New York: Harper and Row Publishers.

Cronbach, L.J., & Meehl, P.E. (1955). "Construct validity in psychological tests." *Psychological Bulletin, 52*(4), 281-302.

Cuthbert, S.L. (1980). "Interpersonal trust: A comparison of attitudinal and situational factors." *Human Relations, 33*(11), 805 -812.

Dassonville, I., Jolly, D., & Desodt, A.M. (1994). "Modelisation of trust between man and machine in a teleoperation system". In *European robotics and intelligent systems conference*, Malaga Spain: EURISCON - Conference Proceedings 1994.

Davis, F.D. (1989). "Perceived usefulness, perceived ease of use, and user acceptance of information technology." *MIS Quarterly*(September), 319-339.

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), 982-1003.

DeGreef, H.P., & Neerincx, M.A. (1995). "Cognitive support: designing aiding to supplement human knowledge." *International Journal of Human-Computer Studies, 42*, 531-571.

Delbecq, A.L., Van De Ven, A.H., & Gustafson, D.H. (1975). *Group techniques for program planning*. USA: Scott, Foresman and Company.

Delbridge, A., Bernard, J., Blair, D., Butler, S., Peters, P., & Tardif, R., Eds. (1990). *The Macquarie Dictionary*: The Macquarie Library Pty. Ltd.

DeLone, W.H., & McLean, E.R. (1992). "Information systems success: the quest for the dependent variable." *Information Systems Research, 3*(1), 60-95.

Dhaliwal, J.S., & Benbasat, I. (1996). "The use and effects of knowledge-based system explanations: theoretical foundations and a framework for empirical evaluation." *Information Systems Research, 7*(3), 342-362.

Dieng, R., Corby, O., & Lapalut, S. (1995). "Acquisition and exploitation of gradual knowledge." *International Journal of Human-Computer Studies, 42*, 465-499.

Doll, W.J., & Xia, W. (1997). "Confirmatory factor analysis of the end-user computing satisfaction instrument: a replication." *Journal of End User Computing, 9*(2), 24-31.

Drilscoll, J.W. (1978). "Trust and participation in Organisational Descision Making as Predictors of Satisfaction." *Academy of Management, 21*(1), 44-56.

Driscoll, R., Davis, K.E., & Lipetz, M.E. (1972). "Parental interference and romantic love: the Romeo and Juliet effect." *Journal of Personality and Social Psychology, 24*(1), 1-10.

Dunteman, G.H. (1989). *Principal Components Analysis*. Newbury Park: Sage Publications.

Edmondson, W.G. (1989). "Asynchronous parallelism in human behaviour: a cognitive science perspective on human-computer interaction." *Behaviour and Information Technology, 8*(1), 3-12.

Einhorn, H.J., & Hogarth, R.M. (1978). "Confidence in judgment: Persistence of the illusion of validity." , 395-417.

Elig, T.W., & Frieze, I.H. (1979). "Measuring causal attributions for success and failure." *Journal of Personality and Social Psychology, 37*, 621-634.

Everett, A.M. (1994). Unpublished Thesis: University of Nebraska, .

Forsythe, D.E., & Buchanan, B.G. (1989). "Knowledge acquisition for expert systems: some pitfalls and suggestions." *IEEE Transactions on Systems, Man, and Cybernetics, 19*(3), 435-442.

Fox, J., & Jackson-Smale, A. (1994). "Why should you trust a decision support system? Towards a quality and safety culture for knowledge based systems." *Knowledge and Decisions in Health Telematics*, 81-89.

Freeman, S.A., Jones, D.D., & Field, W.E. (1994). "A case study in knowledge system acceptance". In *Computers in Agriculture 1994 5th International Conference*. Watson, D.G., Zazueta, F.s. , & Harrison, T.V. (Eds), Orlando: American Society of Agricultural Engineers.

Frost, T., Stimpson, D.V., & Maughan, M.R. (1978). "Some correlations of trust." *Psychology, 99*, 103 - 108.

Fu, L., & Fu, L. (1990). "Mapping rule-based systems into neural architecture." *Knowledge-Based Systems, 3*(1), 48-56.

Gallant, S.I. (1988). "Connectionist expert systems." *Communications of the ACM, 31*(1), 152-169.

Gill, T.G. (1995). "Early expert systems: Where are they now?" *MIS Quarterly*(March), 51-79.

Gregor, S.D. (1996). Explanations from knowledge-based systems for human learning and problem solving. Brisbane: University of Queensland, Unpublished Thesis.

Hammond, K.R., Stewart, T.R., Brehmer, B., & Steinmann, D.O. (1975). *Social judgment theory*. In Human Judgment and Decision Processes. Kaplan, M.F. , & Schwartz, S. (Eds.). New York: Academic Press. *272-307*.

Harrison, S.R., & Tamaschke, H.U. (1984). *Applied Statistical Analysis*. Sydney: Prentice Hall of Australia.

Henderson, J.C. (1987). "Finding synergy between decision support systems and expert systems research." *Decision Sciences, 18*, 333-349.

Henerson, M.E., Morris, L.L., & Fitz-Gibbon, C.T. (1978). *How to measure attitudes*. Los Angeles: Sage Publications.

Hosmer, L.T. (1995). "Trust: the connecting link between organizational theory and philosophical ethics." *Academy of Management Review, 20*(2), 379-403.

Imber, S. (1973). "Relationship of trust to academic performance." *Journal of Personality and Social Psychology, 28*(1), 145-150.

Johns, J.L. (1996). "A concept analysis of trust." *Journal of Advanced Nursing, 24*, 76-83.

Johnson-George, C., & Swap, W.C. (1982). "Measurement of specific interpersonal trust: construction and validation of a scale to assess trust in a specific other." *Journal of Personality and Social Psychology, 43*(6), 1306-1317.

Johnstone, A., Berry, U., & Nguyen, T. (1994). "There was a long pause: influencing turn-taking behaviour in human-human and human-computer spoken dialogues." *International Journal of Human-Computer Studies, 41*, 383-411.

Kahneman, D., & Tversky, A. (1979). "Prospect Theory: An Analysis Of Decision Under Risk." *Econometrica, 47*(2), 263 - 291.

Kaplan, B., & Duchon, D. (1988). "Combining qualitative and quantitative methods in information systems research: a case study." *MIS Quarterly*(December), 571-585.

Kee, H.W., & Knox, R.E. (1970). "Conceptual and methodological considerations in the study of trust and suspicion." *Conflict Resolution, 14*(3), 2 - 12.

Keen, C. (1996). "A survey of course work masters in information systems in Australia". In *The Australian Debate on Information Systems Curriculum.* Arnott, D., Dampney, K. , & Scollary, A., (Eds). Melbourne: Monash University.

Keen, P.G.W. (1991). "Relevance and rigour in information systems research: Improving quality, confidence, cohesion and impact". In *International Federation for Information Processing Information Systems Research: Contemporary Approach and Emergent Traditions.* Nissen, H.E., Klein, H.K., & Hirhscheim, R., (Eds). Elsevier Science Publishers.

Kelley (1973). "The process of causal attribution." *American Psychologist, 28*, 107-128.

Kennedy, A., Wilkes, A., Elder, L., & Murray, W.a.S. (1988). "Dialogue with machines." *Elsevier Science Publishers*, 37-72.

Kim, J.O., & Mueller, C.W. (1978). *Introduction to factor analysis: What it is and how to do it.* Newbury Park: Sage Publications.

Klein, M.R., & Methlie, L.B. (1995). *Knowledge-based Decision support system with applications in business.* West Sussex: John Wiley & Sons.

Kolodner, J.L. (1993). *Case-Based Reasoning.* San Mateo: Morgan Kaufmann.

Komorita, S.S., Hilty, J.A., & Parks, C.D. (1991). "Reciprocity and Cooperation in Social Dilemmas." *Journal of Conflict Resolution, 35*(3), 494 -518.

Koonce, L. (1992). "Explanation and counterexplanation during audit analytical review." *The Accounting Review, 67*(1), 59-76.

Kottemann, J.E., Davis, F.D., & Remus, W.E. (1994). "Computer-assisted decision making: Performance, beliefs, and the illusion of control." *Organizational Behavior and Human Decision Processes, 57*, 26-37.

Kraemer, K.L., & King, J.L. (1988). "Computer-based systems for cooperative work and group decision making." *ACM Computing Surveys, 20*(2), 115-146.

Krippendorff, K. (1980). *Content analysis: A introduction to its Methodology.* Beverly hills, London: Sage Publications.

LaPadula, L.J. (1994). "A rule-set approach to formal modeling of a trusted computer system." *Computing Systems, 7*(1), 113 - 167.

Larzelere, R.E., & Huston, T.L. (1980). "The dyadic trust scale: toward understanding interpersonal trust in close relationships." *Journal of Marriage and the Family*(August), 595-604.

Lee, J.D., & Moray, N. (1994). "Trust, self-confidence, and operators' adaptation to automation." *Int. J. Human-Computer Studies, 40*, 153-184.

Lerch, F.J., Prietula, M.J., & Kim, J. (1993). "Measuring trust in machine advice." *School of Business Adminsitration*. Ypsilanti: Eastern Michigan University, 30.

Lerch, F.J., Prietula, M.J., & Kulik, C.T. (1997). "The turing effect: The nature of trust in expert system advice". In *Expertise in Context*. P .J. Feltovich, K.M.F., R. R. Hottman, (Eds). Menlo Park CA: The MIT Press, 417 - 447.

Levin, M.E.R.I.P. (1968). "Impression formation as a function of source credibility and order of presentation of contradictory information." *Journal of Personality and Social Psychology, 10*(2), 167-174.

Lewis, J.D., & Weigert, A. (1983). *Trust as a Social reality.* The Univeristy of North Carolina Press.

Lietz, P. (1996). "Learning and writing difficulties at the tertiary level: their impact on first year results." *Studies in Educational Evaluation, 22*(1), 41-57.

Lietz, P. (1998). Personal communication.

Mayer, R.C., Davis, J.H., & Schoorman, F.D. (1995). "An integrative model of organizational trust." *Academy of Management Review, 20*(3), 709-734.

McAllister, D.J. (1995). "Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations." *Journal of The Acadamy of Management, 38*(1), 24 - 59.

McBroom, L.W., & Seaman, J. (1988). "Evaluation of computerized job matching system's responsiveness to persons who are visually impaired." *Vocational Evaluation and Work Adjustment Bulletin*(Winter), 139-147.

McDermid, J.A. (1991). "Achieving and demonstrating trustworthiness or Asking the wrong question yields the wrong answer.". In *MILCOMP Europe '91 military computer systems and software*. London: Microwave Exhibitions and Publishers.

Meyers, M.D. (1994). "Quality and qualitative research in information systems". In *Australasian Conference on Information Interchange* (Ed).^(Eds)., Melbourne: Unpublished.

Miller, D.L. (1991). Path analysis. *Handbook of Research Design and Social Measurement*. Newbury Park, California: Sage Publications, 286-292.

Milne, G.R., & Boza, M.-E. (1998). *Trust and concern in consumer's perceptions of marketing information management practices.* Cambridge: Marketing Science Institute.

Misztal, B.A. (1998). *Trust in Modern Societies*. Cambridge: Polity Press.

Moore, G.C., & Benbasat, I. (1991). "Development of an instrument to measure the perceptions of adopting an information technology innovation." *Information Systems Research, 2*(3), 192-222.

Morgan, G., & Smircich, L. (1980). "The case for qualitative research." *Academy of Management Review, 5*(4), 491-500.

Morgan, R.M., & Hunt, S.D. (1994). "The commitment-trust theory of relationship marketing." *Journal of Marketing, 58*(July 1994), 20-38.

Mrozek, A. (1992). "A new method for discovering rules from examples in expert systems." *International. Journal of Man-Machine Studies, 36*, 127-143.

Muir, B.M. (1987). "Trust between humans and machines, and the design of decision aids." *International. Journal of Man Machine Studies, 27*, 527-539.

Muir, B.M. (1994). "Trust in automation: Part 1. Theoretical issues in the study of trust and human intervention in automated systems." *Ergonomics, 37*(11), 1905-1922.

Muir, B.M., & Moray, N. (1996). "Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation." *Ergonomics, 39*(3), 429-460.

Nass, C., Steuer, J., Henriksen, L., & Dryer, D.C. (1993). "Machines, social attributions, and ethopoeia: performance assessments of computers subsequent to "self-" or "other-" evaluations." *International Journal of Human-Computer Studies, 40*, 543-559.

Neuman, L.W. (1994). *Social Research Methods: Qualitative and Quantitative Approaches*. Needham Heights: Allyn and Bacon.

Novosel, M., & Kohn, R. (1979). "Testing plausible path models of interpersonal trust in patient-physician treatment relationships." *Medical psychology and medical sociology, 13A*, 81 -99.

Oliver, R.D., & Stephane, L. C., (1995). "Acquisition and exploitation of gradual knowledge." *International Journal of Human - Computer Studies, 42*, 465-499. .

Parks, C.D., & Hulbert, L.G. (1995). "High and low trusters' responses to fear in a payoff matrix." *Journal of Conflict Resolution, 39*(4), 718-730.

Pervan, G.P. (1994). "Ensuring quality in the conduct and presentation of research in information systems". In *Australasian Conference on Information Interchange* (Ed).^(Eds). , Melbourne: Unpublished.

Pincus, K.V. (1991). "Audit judgment confidence." *Behavioral Research in Accounting, 3*, 39-65.

Power, D.J., Meyeraan, S.L., & Aldag, R.J. (1994). "Impacts of problem structure and computerized decision aids on decision attitudes and behaviors." *Information and Management, 26*, 281-294.

Ravetz, J. (1993). "Expert system development in ill-structured domains: the application of artificial intelligence technology to diagnosis and assessment in the human services." *Computers in Human Services, 9*(3-4), 351-360.

Rees, P.L. (1993). "User participation in expert systems." *Industrial Management and Data Systems, 93*(6), 3-7.

Rempel, J.K., Holmes, J.G., & Zanna, M.P. (1985). "Trust in close relationships." *Journal of Personality and Social Psychology, 49*(1), 95-112.

Robey, D. (1979). "User attitudes and management information system use." *Academy of Management Journal, 22*(3), 527-538.

Rosenbaum, M.E., & Levin, I.P. (1968). "Impression formation as a function of source credibility and order of presentation of contradictory information." *Journal of Personality and Social Psychology, 10*(2), 167-174.

Rosenbaum, M.E., & Levin, I.P. (1969). "Impression formation as a function of source credibility and the polarity of information." *Journal of Personality and Social Psychology, 12*(1), 34-37.

Ross, W., & LaCroix, J. (1996). "Multiple meanings of trust in negotiation theory and research: A literature review and integrative model." *International Journal of Conflict Management, 7*(4), 314 - 348.

Ross, W.H., & Wieland, C. (1996). "Effects of interpersonal trust and time pressure on managerial mediation strategy in a simulated organizational dispute." *Journal of Applied Psychology, 81*(3), 228-248.

Rotenberg, K.J. (1995). "The socialisation of trust: parents' and children's interpersonal trust." *International Journal of Behavioral Development, 18*(4), 713-726.

Rotenberg, K.J., & Morgan, C.J. (1995). "Development of a scale to measure individual differences in children's trust-value basis of friendship." *Genetic Psychology, 156*(4), 489 -498.

Rotter, J.B. (1968). "A new scale for the measurement of interpersonal trust." *Journal of Personality, 35*, 651-665.

Rotter, J.B. (1980). "Interpersonal trust, trustworthiness, and gullibility." *American Psychologist, 35*(1), 1-7.

Russell, D. (1982). "The casual dimension scale: a measure of how individuals perceive causes." *Journal of Personality and Social Psychology, 42*(6), 1137-1145.

Sabatelli, R.M., Buck, R., & Dreyer, A. (1983). "Locus of control, interpersonal trust, and nonverbal communication accuracy." *Journal of personality and social psychology, 44*(2), 399 - 409.

Schindler, P.L., & Thomas, C.C. (1993). "The structure of interpersonal trust in the workplace." *Psychological Reports, 73*, 563-573.

Segars, A.H. (1997). "Assessing the unidimensionality of measurement: a paradigm and illustration within the context of Information Systems research." *Omega, International Journal of Management Science, 25*(1), 107-121.

Sellin, N., & Keeves, J.P. (1997). Path analysis with latent variables. *Educational Research, Methodology & Measurement: An International Handbook.* Keeves, J.P. Oxford: Elsevier Science, 633-640.

Sethi, V., & King, W.R. (1994). "Development of measures to assess the extent to which Information Technology application provides competitive advantage." *Management Science, 40*(12), 1601-1627.

Sharma, R.S., & Conrath, D.W. (1992). "Evaluating expert systems: the socio-technical dimensions of quality." *Expert Systems, 9*(3), 125-138.

Shaw, R.B. (1997). *Trust in the Balance.* San Francisco: Jossy-Bass Publishers.

Sheridan, T.B. (1988). "Trustworthiness of command and control systems". In *Man-Machine Systems.* Oulu: IFAC.

Sitkin, S.B. (1992). "Reconceptualizing the determinatns of risk behaviour." *Acadamy of Management Review, 17*(1), 9 -38.

Sprague Jr., R.H., & Watson, H.J. (1996). *Decision Support for Management.* Upper Saddle River: Prentice Hall.

Straub, D., W. (1994). "The effect of Culture on IT diffusion: E-mail nad FAX in Japan and the U.S." *Information Systems Research, 5*(1), 23-47.

Straub, D., Limayem, M., & Karahanna-Evaristo, E. (1995). "Measuring System Usage: Implications for IS testing theory." *Management Science, 41*(8), .

Straub, D.W. (1989). "Validating instruments in MIS research." *MIS Quarterly*(June), 146-169.

Sussman, M., & Robertson, D.U. (1986). "The validity of validity: an analysis of validation studies." *Journal of Applied Psychology, 71*(3), 461-468.

Swanson, E.B. (1987). "Information channel disposition and use." *Decision Sciences, 18*, 131-145.

Tabacknick, B.G., & Fidell, L.S. (1989). *Using Multivariate Statistics.* New York: Harper Collins Publishers.

Taylor, S., & Todd, P.A. (1995). "Understanding information technology usage: a test of competing models." *Information Systems Research, 6*(2), 144-176.

Todd, P.A., & Benbasat, I. (1992). "The use of information in decision making: An experimental investigation of the impact of computer based decision aids." *Mis Quarterly*(September), 372-393.

Todd, P.A., & Benbasat, I. (1994). "The influence of decision aids on choice strategies under conditions of high cognitive load." *IEEE Transactions on Systems, Man, and Cybernetics, 24*(4), 537-547.

Torkzadeh, G., & Doll, W.J. (1994). "The test-retest reliability of user involvement instruments." *Information & Management, 26*, 21-31.

Trafimow, D., & Sniezek, J.A. (1994). "Perceived expertise and its effect on confidence." *Organizational Behavior and Human Decision Processes, 57*, 290-302.

Turban, E. (1995). *Decision Support and Expert Systems Management Support Systems*. Englwood cliffs: Prentice Hall.

Turban, E., & Aronson, J.E. (1998). *Decision Support and Intelligent Systems.* Upper Saddle River: Prentice Hall.

Vinze, A.S. (1992). "Empirical verification of effectiveness for a knowledge-based system." *International Journal of Man-Machine Studies, 37*, 309-334.

Waern, Y. (1996). "People's perception of human and computer advice." *Computers in Human Behavior, 12*(1), 17-27.

Watson, H.J., & Frolick, M.N. (1993). "Determining information requirements for an EIS." *MIS Quarterly*(September), 255-269.

Welbank, M. (1990). "An overview of knowledge acquisition methods." *Interacting with Computers, 2*(1), 83-91.

Wickens, C.D. (1994). "Designing for situation awareness and trust in automation". In *Integrated Systems Engineering*. Baden-Baden, Germany: IFAC.

Will, R.P. (1991). "True and false dependence on technology: evaluation with an expert system." *Computers in Human Behavior, 7*, 171-183.

Will, R.P. (1992). "Individual differences in the performance and use of an expert system." *International. Journal of Man-Machine Studies, 37*, 173-190.

Wong, B.K., Bodnovich, T.A., & Selvi, Y. (1997). "Neural network applications in business: a review and analysis of the literature (1988-95)." *Decision Support Systems, 19*, 301-320.

Wong, D.W., Chan, F., A, D., GayChalong, & Hattori, H.K. (1992). "Toward the development of a second generation computerized job-matching system for persons with disabilities: a conceptual framework." *Journal of Rehabilitation* (January/February/March), 70-77.

Woods, D.D., Roth, E.M., & Bennett, K.B. (1989). Explorations in joint human-machine cognitive systems. *Cognition, Computing, and Cooperation.* Robertson, S.P., Zachary, W. , & Black, J.B. (Eds). Norwood: Ablex Publishing Corporation, 123-158.

Wrightsman, L.S. (1966). "Personality and attitudinal correlates of trusting and trustworthy behaviors in a two-person game." *Journal of Personality and Social Psychology, 4*(3), 328-332.

Yamagishi, T. (1986). "The provision of sanctioning system as a public good." *Journal of Personality and Social Psychology, 51*(1), 110-116.

Ye, L.R., & Johnson, P.E. (1995). "The impact of explanation facilities on user acceptance of expert systems advise." *MIS Quarterly*(June), 157-172.

Yin, R.K. (1993). *Applications of case study research.* Newbury Park, California: Sage Publications Ltd.

Yoon, Y., Guimaraes, T., & O'Neal, Q. (1995). "Exploring the factors associated with expert systems success." *MIS Quarterly*(March), 83-106.

Young, C.A. (1996). *Validity Issues in Measuring Psychological Constructs.* http://trochim.human.cornell.edu/tutorial/young/eiweb2.htm.

Yuan, Y., & Kochen, M. (1987). "Matching support systems: information systems for supporting matching processes." *Human Systems Management, 7*, 129-141.

Zahedi, F. (1990). "A method for quantitative evaluation of expert systems." *European Journal of Operational Research, 48*, 136-147.

Zand, D.E. (1972). "Trust and managerial problem solving." *Administrative Science Quarterly, 17*, 229 -239.

Zurada, J.M. (1992). *Introduction to Artificial Neural Systems.* St. Paul: West Publishing Company.

Central Queensland
UNIVERSITY

LIBRARY DECLARATION

(This declaration shall be appended to the first of the requisite copies of the thesis submitted for examination to the Research Services Office, Central Queensland University, Rockhampton, Queensland, 4702).

MARIA MADSEN — MASTER OF BUSINESS (RESEARCH)

Author: — *Degree:*

Title of Thesis: The Development of a Psychometric Instrument for Human-Computer Trust: An Investigation of Trust Within The Context of Computer-Aided Decision-Making.

I acknowledge that, following successful examination, the first copy of the thesis described above, shall be deposited in the University Library and made available in accordance with the University's regulations governing the submission and examination of research higher degree theses.

☑ *I agree that the thesis shall be made freely available for the purpose of research or private study.

OR

☐ *I request that access to the thesis, or parts thereof, be restricted for a specified period, and that release of the thesis during that period for the purpose of research or private study be contingent upon the approval of the Director of the Division of Library, Information and Media Services (or nominee). I understand, in those cases where I am not reasonably available, such approval may be granted without my having been consulted.
(An author electing this option must complete the reverse side of this declaration).

* Tick the elected option.

Signature Redacted

SIGNATURE (AUTHOR) — 3/4/2000 DATE

REQUEST TO RESTRICT ACCESS TO HIGHER DEGREE THESIS

(I) INDICATE THE REASON(S) FOR SEEKING TO RESTRICT ACCESS TO THE THESIS OR PARTS THEREOF.
(Sufficient details should be provided in order that a decision may be made in the author's absence).

(II) IS IT INTENDED TO RESTRICT ACCESS TO THE WHOLE THESIS? YES/NO
Where the restriction is intended to apply to parts of the thesis, please nominate the relevant section. (If practicable, such sections should form a separate confidential appendix).

(III) NOMINATE THE DESIRED PERIOD OF RESTRICTED ACCESS
(Restricted access shall normally be granted for a period of twelve months. Any extension beyond that period shall be at the discretion of the Research Board).

SIGNATURE (AUTHOR) DATE

(IV) APPROVAL TO RESTRICT ACCESS TO HIGHER DEGREE THESIS
Permission is hereby granted to restrict access to the thesis or parts thereof, as detailed above, for a period of ....

SIGNATURE (CHAIR, RESEARCH COMMITTEE) DATE