Measurement of Factors Determining Relevance Judgements

By

Amir Ghaebi

A Doctoral Thesis

Submitted in partial fulfilment of the requirements

for the award of

Doctor of Philosophy of Loughborough University

January 2003

In the Name of God, The Most Gracious, The Most Merciful

Abstract

This study has focused on cognitive aspects of the human processes involved in relevance judgements. Several criteria and a great number of measures have been proposed and used for relevance assessment. However, there is a lack of agreement as to which are the best measures and to what extent they are affected by variability of relevance judgements. The purpose of this research was to identify those cognitive factors, which primarily contribute to relevance judgement. In this study, 16 criteria that influence cognitive relevance were identified and used. The study addressed three questions: 1) What cognitive factors affect relevance judgement? 2) What is the importance of each factor on relevance judgement? 3) Do the factors remain stable over time?

Both qualitative and quantitative techniques were used for data collection and analysis. Structural Equation Modelling (SEM) as a multivariate technique was used to develop a statistical model of cognitive relevance. It seems, it is the first time that this technique has been applied to measure cognitive factors of relevance. Saracevic's (1996) stratified model, as a cognitive IR model was adopted to provide a necessary framework to incorporate relevance cognitive theory and a user approach in measuring relevancy.

A series of experiments were conducted; real users with real information needs who are research students in science and engineering of Loughborough University and the University of Nottingham made the assessments. The sample comprised 30 Ph.D. research students (19 in Loughborough University and 11 from the University of Nottingham); seven were female and 23 were male. Twenty-one of the participants were studying science and nine were studying engineering. The experiment involved the participants making relevance measurements three times in three-month intervals between 2001 and 2002. The objective of the study was to determine to what extent a relevance judgement is related to various factors such as Aboutness, Quality of Information, Characteristics of Information and Information Novelty. Structural equation analysis identified the cognitive criteria and factors that were rated as most important by participants. These criteria can be categorised into four classes: Aboutness (about the topic, informative, suitability, background, bibliography and controversial), Quality of Information (validate viewpoint, consistent, understandable, and described methods), Characteristics of Information (know the author, know the journal and authors eminence) and Information Novelty (new, original and unique or only sources). The results show how much of the variance is explained by all of the variables put together. The conceptual framework assumed that Aboutness, Quality of Information, Characteristics of Information and Information Novelty are four independent latent factors that would represent relevance judgement. The results of

structural equation modelling (SEM) indicate that a four-factor structure is necessary to measure cognitive relevance judgement. The path coefficients of factors show that the contribution of all factors in three studies are important and, in fact, are statistically significant. The results also support the hypothesis regarding the stability of the factors of relevance judgement. This study discusses possible implications of these findings for the debate over the use of factors for the interactive Information Retrieval Systems (IRS) developer. This study can be considered as a starting point that may initiate several future studies to focus on the specific areas of interest suggested in the findings.

Key words: Interactive Information Retrieval Systems (IIRS), relevance measurement, cognitive relevance, Structural Equation Modelling (SEM).

Acknowledgements

I wish to express my sincere appreciation to my supervisor Professor Charles Oppenheim whose guidance, assistance and professional support helped me during the time of research and with writing this thesis. I wish to thank my director of research, Professor Cliff Mcknight, for his valuable advice.

For my financial support, I wish to thank the Ministry of Science, Research & Technology of Islamic Republic of Iran. I also thank the Department of Information Science at Loughborough University for their continual support and for providing the necessary funding I required to recruit the participants. I would also like to thank Shirley Horner, Departmental Computer Officer, for her supports during my study. I also wish to thank all the experts in IR and PhD students who participated in this study.

Finally, I would like to give my special thanks to my wife Marzieh for her long support.

Amir Ghaebi

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Relevance is central to information retrieval research. However, a great deal remains to be explored in this area. Despite its obvious importance, cognitive relevance has not been widely investigated. The present study is an attempt to partially fill this gap by giving a preliminary insight into and measure of factors that affect cognitive relevance. This chapter provides an overview of the thesis by describing the purpose and scope of the research. The chapter specifies the background to the study, statement of the problem, the broad research aims, and states the significance of the research. Finally, the chapter outlines the structure of the thesis.

1-1. Context of the Study

Many researchers believe that relevance is one of the most fundamental concepts in information retrieval theory. Froelich (1994) is one of many authors who have noted that relevance is important for evaluating an information retrieval system. The concept of relevance has been much debated and developed since it was first adopted as the basis of measures of the effectiveness of an Information Retrieval System (IRS), and it continues to be used in evaluations today. The nature of relevance and the role of users in relevance judgements are clearly questions that will not go away in the near future (Tague-Sutcliffe 1996). Recent studies on relevance view IR as a cognitive interaction between human and computer. Harter (1996) for example, placed strong emphasis on the cognitive nature of relevance. Relevance involves an interactive, dynamic relationship by inference, with intentions toward a context (Saracevic 1996). Research on relevance judgements has developed in the context of information retrieval. The notion of relevance was initially viewed as an objectified constant embedded in the mechanism of document retrieval, functioning as a measurement criterion for system performance. The advancement of information retrieval research saw the emergence of a user-oriented approach, in contrast to the conventional system-oriented approach. Ingwersen (1996) stated the cognitive paradigm argues that the essence of the

information retrieval process is the interplay among various cognitive elements and structures. This study has focused on cognitive aspects of the human processes involved in relevance judgements.

1-2. Statement of the Problem

Relevance is complex and inseparable from the context in which judgements of relevance are made (Saracevic 1975; Schamber 1994). Earlier research demonstrated the difficulties with the system-oriented approach and that many aspects of human cognitive processes contribute to relevance judgement (Saracevic 1975; Schamber 1994; Barry 1994; and Harter 1996). Sperber and Wilson (1995) claimed that relevance is signalled by the cognitive improvement of an individual in the context of document retrieval. They also pointed out that relevance is reflected by cognitive changes in the mental state of the user. There are a variety of factors suggested by scholars that contribute to relevance judgement (e.g., Barry 1993; Barry 1994; Bruce 1994; Cool et al. 1993; Howard 1994; Park 1995; Regazzi 1988; Schamber 1991; Wang and Soergel 1998). Bateman (1998) proposed a model consisting of six constructs that classified the various factors. Tang and Solomon (2001) noted that in spite of attempts to classify relevance criteria, there is no consensus on the categorisation and labelling the classes of criteria. Spink, Greisdorf and Bateman (1998) stated that cognitive correspondence, informativeness, novelty and information quality are the factors by which cognitive relevance is inferred. Tang and Solomon (2001) in their laboratory experiment tabulated fifteen criteria and grouped them into three categories: topicality, information quality and cognitive state. Maglaughlin and Sonnenwald (2002) investigated the use of criteria to assess relevant document and identified twenty-nine criteria grouped into six categories. Although some studies (e.g. Barry 1994; Bateman 1998; Cool et al. 1993; Maglaughlin and Sonnenwald 2002; Schamber and Bateman 1996; Tang, Shaw and Vevea 1999; Tang and Solomon 2001; Wang and Soergel 1998) have investigated criteria used to determine relevant documents, the question remains: what is the degree of impact of each criterion or factors? This study deals with those factors which affect cognitive relevance. In this study, sixteen criteria that influence cognitive relevance were

identified and statistical analyses were used to identify the most important of these factors.

1-3. Objectives of the Study and Research Questions

The primary purpose of this research is to identify the set of cognitive criteria that might influence relevance of documents in searching. The second purpose of the study is to measure the degree of effect of each these cognitive factors As a result, the importance of each criterion of cognitive relevance can be recognised and Relevance judgements are regressed against several potential variables. In this study, we examined to what extent relevance judgement related to cognitive factors (Aboutness, Quality of information, Characteristics of Information and Information Novelty). The third purpose of the study is to examine whether cognitive factors of relevance judgements remain stable over time and thus arguably represent constant factors of cognitive relevance judgements.

The present study addressed three research questions. First, what cognitive factors affect relevance judgement? Second, what is the importance of each cognitive factor on relevance judgement? And finally, do these factors remain stable over time? Answers to these questions will further our understanding of the conceptual development of theories of relevance and suggest design possibilities for information retrieval systems.

1-4. The Structure of the Thesis

The thesis has seven chapters. Chapter 2 presents a literature review relating to cognitive relevance. The chapter describes three main cognitive models of relevance judgement and dimensions of relevance. It also discusses the theoretical foundations of relevance and cognitive factors that may affect relevance judgements. Chapters 3 explains the methodology of the study and outlines methods employed to select the sample, the

design adopted for the research, and the tools used to collect data. It then discusses statistical techniques used for data analysis. Chapter 4 describes the first stage of the empirical study for developing a theoretical framework. It reports on the measurement item generation, a pilot study and a study of experts' attitude measurement. In Chapter 5 a quantitative analysis of the scale items is reported. It reports the results of analysis of reliability and validity of scales for Study 1, Study 2 and Study 3. Chapter 6 develops a structural equation modelling analysis to examine relationships among relevance criteria (independent variables) and their latent factors (independent latent factors). A further analysis of all studies including modelling for longitudinal study is presented in this chapter. Chapter 7 discusses the findings of the research. The conclusions drawn from this research are also detailed in this chapter. Finally, this chapter summarises the implications of this study and discusses the potential for future research.

Chapter 2: Review of the Literature

2-1. Background

The paradigms for traditional IR evaluation were established by the Cranfield experiments (Cleverdon, Mills and Keen 1966) and were continued in many experiments on information systems in 1960s and 1970s. The demand for a hybrid approach to the evaluation of information retrieval systems that combines the system-centred and the cognitive user approaches is increasing (Borlund 2000).

Su (1992) reported that there is no clear guidance about how to evaluate interactive IR performance despite the efforts of designers and developers of IR systems since the 1950s, and of researchers and practitioners since the 1960s. The development of IR in the 1960s gave rise to the need for measures of the effectiveness of these systems. Since the main objective of an IR system is to retrieve information relevant to user queries (Saracevic 1975), relevance has been adopted as the appropriate fundamental measure of effectiveness, and it continues to be used in evaluation today.

Evaluation of IR systems through the 1970s concentrated on testing the components or sub-processes of IR systems in terms of recall and precision (the best known relevance-based measures) in controlled environments (Saracevic 1975).

Van Rijsbergen (1979) noted that many of the tasks, which certainly have an information storage and retrieval dimension, such as cataloguing and general administration, have successfully been addressed by new technology. However, the problem of effective retrieval remains largely unsolved.

A number of comprehensive critical literature reviews have appeared on various aspects of information retrieval. Belkin and Vickery (1985) stated that research has not yet provided a satisfactory solution to the problem of interfacing between end-user and large-scale databases. Saracevic et al. (1988) commented that, despite a relatively large amount of literature about the subject, research in IR was in its infancy.

Saracevic (1996) believed that research on interactive aspects of IR had not reached maturity, even it was little by little emerging out of infancy. Although a variety of research articles and reports reflecting theoretical, experimental, or observational studies regarding relevance have been published in the information retrieval literature, a research approach still needs to be developed to bridge the gap in this field.

The notions of relevance and user's relevance judgements are critical to the theory and research of information retrieval. The basic objective of IR is often stated as the retrieval of relevant items (texts, images, and sounds) matched to a user query, which in turn represents user's information need. The concept of relevance has been widely debated in information science literature. This study, however, emphasises the measurement of relevance.

This chapter reviews the role of relevance in evaluation of information retrieval systems and follows with a synthesised overview of the theory of relevance. The fourth section deals with the issue of relevance models and follows with a summary of literature of relevance criteria.

2-2. Information Retrieval Systems (IRS)

An early and straightforward definition of an IRS was given by Lancaster (1968): ' An IRS does not inform (i.e. change the knowledge of) the user on the subject of his inquiry. It merely informs on the existence (or non-existence) and whereabouts of documents relating to his request.'

Kowalski (1997) described an IRS as a system that is capable of storage, retrieval, and maintenance of information. Information in this context can be composed of text (including numeric and date data), images, audio, video and other multimedia objects.

Chowdhury (1999) stated that an IRS composed a set of interacting components, each of which is designed to serve a specific function for a specific purpose. All these components are interrelated to achieve a goal, that is, to retrieve information in a narrower sense, and to increase the level of knowledge of the users in a broader sense.

Information retrieval systems are established to answer questions by users. The general objective of an IRS is to minimise the "overhead" of a user locating needed information. This "overhead" can be expressed as the time a user spends in all from initial query formulation to reading an item containing the needed information. An IRS serves as a bridge between the world of creators or generators of information, and the users of that information. Lancaster (1968) listed some of the major functions of an IRS as follows:

- To identify the information (sources) relevant to the areas of interest of the target users' community
- To analyse the contents of the sources (documents)
- To represent the content of the analysed sources in a way that will be suitable for matching users' queries
- To analyse users' queries and to represent them in a form that will be suitable for matching with the database
- To match the search statement with the stored database
- To retrieve the information that is relevant
- To make necessary adjustments in the system based on feedback from the users.

An IRS is designed to facilitate the process of searching a collection of documents with the goal of identifying documents that relate to a particular topic (Hert 1997). According to Ellis (1992), information retrieval has been dominated by two major paradigms: the physical (system-oriented) and the cognitive (user-oriented). The goal of both approaches is to improve IRSs. Relevance and relevance judgement are two user oriented approaches studies which if fully understood, might lead to the improved performance of IRSs.

2-3. Evaluation of IRS

Evaluation techniques are critical to research in all areas of science and engineering. Interest in the evaluation techniques for IRS has significantly increased with the commercial use of information retrieval technologies.

In recent years, the evaluation of IRS and techniques for indexing, sorting, searching and retrieving information have become increasingly important (Salton and McGill 1983). This growth in interest is due to the growing number of retrieval systems, starting with online systems 30 years ago (Salton 1992), and developing greatly in the form of web search engines in the last few years.

Generally, one of the main tasks of evaluating IRSs is to obtain information about the satisfaction of the user's task in a specific work environment.

Evaluation of IRS is concerned with how well the system is satisfying, not just individual cases, but collectively, all actual and potential users. The purpose of evaluation is ultimately to lead improvements in the information retrieval process itself. Evaluation of IRS not only improves existing systems, but also provides the ground for new systems to be developed with high reliability and effectiveness. There are many reasons to evaluate the effectiveness of an IRS (Kowalski 1997):

- To aid in the selection of a system to be bought
- To monitor and evaluate system effectiveness
- To evaluate query generation process for improvements
- To provide inputs to cost-benefit analysis of an information system
- To determine the effects of changes made to an existing information system.

To evaluate how effective the system is, some writers believe that the original user must be involved in the relevance judgements. Others believe that at least some aspects of a system can be evaluated without relevance judgements from the user (Tague-Sutcliffe 1996).

Traditional IR experiments have been carried out for almost forty years ranging from Cranfield to TREC (Harman, 1995) studies. The first major information retrieval evaluation was carried out when Cyril Celeverdon's Cranfield projects created what remains the primary evaluation paradigm (Belkin and Vickery 1985). In this case, questions were based on individual papers to be in a test collection, and did not address real users' information needs. For each question, there was known to be at least one relevant document, i.e., the document used to create the question (Tague-Sutcliffe 1996).

Interest in the evaluation techniques for IRSs has significantly increased with the commercial use of information retrieval technologies in the Internet. Until 1993, the evaluations were done primarily by academicians using a few small, well known corpora of test documents or even smaller test databases created within academia. The evaluations focused primarily on the effectiveness of search algorithms (Hert 1997). The creation of the annual Text Retrieval Evaluation Conference (TREC) has since become the standard process of evaluating information systems.

The TREC is a series of annual conferences that focus on text retrieval from large collections of full text documents. Participants test their retrieval systems on a large test collection. Search results from different systems are discussed in the conference. Conferences have been held every year, starting from 1992. The conference provides a standard database consisting of gigabytes of test data, search statements and the expected results from the searches to academic researchers and commercial companies for testing of their systems. This has placed a standard baseline into comparisons of algorithms. Although there is now a standard database, there is still debate on the accuracy and utility of the results from use of the test corpus. The central task in each TREC has been to test the ability of retrieval systems to produce a ranked list of documents in response to each of 50 test questions (Voorhees 2000). TREC is not based on real users or real information needs. According to Harman (1995), however, external human assessors made the final relevance judgements in TREC. Ingwersen (2001) pointed out the TREC and other interactive investigations and experiments demonstrate the problematic issues concerned with the concept of relevance and the evaluation methods generally applied in IR.

As noted earlier, researchers have identified two major approaches: system –oriented approach and user-oriented approach. Most authors (Ellis 1992, Schamber 1994, Hert 1997) are agreed on that major division. The system-oriented approach perspective focuses on topical relevance and concerns finding documents that address a concept-based information need. Recall and precision are typically used as measures of effectiveness with this view of relevance. This approach refers to a correspondence between user's query terms and the terms that are indexed or stored in an IRS. In the system-oriented approach, relevance is the relationship between the stated request and the response to the request. However, the users' perception of how those items relate to

his or her information need is not considered. , In TREC, relevance can be regarded as a binary, topical, and stable manifestation (Ingwersen 2001).

The user-oriented perspective on relevance is somewhat broader, seeking to characterise the relationship between information and the user's problem situation and attempting to account for the various aspects of human cognitive processes used in making relevance judgements. User-oriented approach includes subjective, situational and cognitive aspects of relevance. In this case, the user should make a judgement on the items retrieved. A user-oriented approach of relevance was chosen for this study.

2-4. Relevance

Relevance is a fundamental concept for documentation, information science and information retrieval (Mizzaro 1997). Saracevic (1975) claimed that relevance was the reason for the birth of information science. Froelich (1994) noted that relevance is important for building and evaluating an information retrieval system. The concept has been much debated and developed since it was first adopted as the basis of measures of the effectiveness of IRS, and it continues to be used in evaluations today. Although relevance has been a constant in terms of its continued use as the chosen measure, numerous different definitions of the concept have been proposed.

The issue of definition is a vital one, since the choice of definition will radically affect any evaluation of system performance (Walker and Janes 1999). It is therefore essential, when conducting an evaluation, to be clear about the sense in which relevance is being used. For example, where comparisons of systems are made on the relevance of their output, it is vital for the same interpretation of relevance to have been applied to each system.

The definition of relevance has occupied many researchers. Many possible alternate terms, expressions, and ideas have been offered over the years: "topicality" (documents that are on the same topic as the search request), "satisfaction" (documents that the user or someone else says satisfy the request), "utility" (documents that are useful to the

user), "pertinence" (related to topicality and "aboutness") and so on (Walker and Janes 1999).

Schamber, Eisenberg and Nilan (1990), in an article reviewing more than three decades of work on relevance and related issues, came to the following conclusions about the nature of relevance and its role in information behaviour:

> "Relevance is a multidimensional cognitive concept whose meaning is largely dependent on users' perceptions of information and their own information need situations; relevance is a dynamic concept that depends on users' judgements of the quality of the relationship between information and information needs at a certain point in time; and relevance is a complex but systematic and measurable concept if approached conceptually and operationally from the user's perspective".

Some words such as: "multidimensional," "dynamic," "complex," "systematic," "measurable." and also the heavy emphasis on the user are noticeable in this view.

Schamber, Eisenberg and Nilan (1990) further outline the philosophical underpinnings of the dynamic, situational approach.

> "The dynamic, situational approach we suggest views the user – regardless of system – as the central and active determinant of the dimensions of relevance. We believe that relevance is a multidimensional concept; that it is dependent on both internal (cognitive) and external (situational) factors; that it is based on a dynamic human judgement process; and that it is a complex but systematic and measurable phenomenon".

2-4-1. Three Views of Relevance

The numerous different approaches that have been taken in defining relevance can be grouped into three main categories: the system view; the judged view; and the user view.

The System View. This view of relevance focuses on the system element in the information retrieval process. It is often referred to in the literature as 'topicality'. Essentially, system relevance refers to direct matches between terms in queries and terms in system documents. Barry (1994) summed up the system view as assuming that:

> "... relevance is solely a property of the internal mechanism of the system and that relevance is the result of a match between the subject terms of a query and the subject terms assigned to documents. This view implies that all documents correctly retrieved by the system will be, by definition, relevant to the user."

The Judged View. This view of relevance focuses on the user. It involves a subjective human judgement of the match between a query and a retrieved document based upon the user's request for information.

Essentially, judged relevance refers to the relationship between the information request and retrieved documents, as evaluated by a judge. Judged relevance is determined on the basis of a relationship between documents and an expressed information need. This is not necessarily the same, as actual information need. As Saracevic (1975) commented:

> "...experience has taught us that at times, often unintentionally, a question does not exactly coincide with what a questioner had on his/her mind...it may be difficult to verbalise a question even if it is on one's mind."

The User View. The user view of relevance is the most subjective of the three approaches. It refers to the user's judgement of the ability of a document to satisfy his/her information need. In the other words, user relevance refers to the user's decision to accept or reject the retrieved information and that includes any factor, however subjective, that the user takes into account when making the decision. Schamber (1994) refers to this type of relevance as situation relevance, describing it as the relationship between information and the user's information problem situation (Schamber 1994). Saracevic (1975) refers to it as pertinence rather than relevance, and Cooper (1971) terms it utility.

The user view acknowledges that the user is seeking to satisfy an actual information need, rather than some inadequate understanding or expression of that need. User relevance has been applied as an evaluative measure of IRS output in a number of recent research projects. Schamber (1994) noted in her summary of eight relevance studies which were conducted between 1988-1994, that each study had applied user relevance in the evaluation of search results (Schamber 1994).

However, one consequence of the shift from system relevance to user relevance is that systems are being evaluated for their ability to perform a function for which they were not necessarily designed. It is therefore necessary to establish the criteria, however subjective, that users employ in deciding on relevance and to seek to incorporate these criteria into the IR system process. Barry (1994) acknowledged that it would be impossible to incorporate every factor that influenced user relevance judgements.

Kowalski (1997) claimed that it is necessary to define the context under which the concept is used. From a human judgement standpoint, relevance depends upon a specific user's judgement (subjective), relates to a user's requirements (situational), depends on human perception and behaviour (cognitive), changes over time (temporal) and also it is observable at a point in time (measurable).

2-4-2. Dimensions of Relevance

Saracevic (1996) reconsidered relevance on the basis of recent developments and distinguished four frameworks for relevance: systems framework, communication framework, situational framework and psychological framework. As a cognitive notion, relevance involves an interactive, dynamic establishment of a relation by inference, with intentions toward a context (Saracevic 1996). During this interaction, new dimensions of relevance must be added. Saracevic (1996) also proposed the interactive framework that incorporates elements of all previous four frameworks and sustains a more complete view of relevance. Time is an important dimension that has recently been proposed by some scholars such as Spink, Greisdorf and Bateman (1998) and Cosijn and Ingwersen (2000).

2-4-3. Attributes of Relevance

Relevance has certain attributes, and relevance manifests itself in different ways. Saracevic (1996) in his article distinguishes the following attributes of relevance:

Relation. It is clear that relevance always implies a relation. This relation is between some entity and information object, which is simply defined as "texts".

Intention. The relation in expression of relevance involves intentions such as objectives, roles, expectations (motivation).

Context. Intention always comes from a context, and is always directed toward that context.

Inference. Assessment of the effectiveness of a given relation.

Interaction. Inference is accomplished through a dynamic process of interaction, and change of cognition causes the change of interpretations.

Relevance is intuitively well understood by users, particularly in all uses of information. Relevance has to combine some intuitive attributes such as: it is based in cognition, it is expressed in a context; it deals with effectiveness and people use it dynamically, so it involves interaction. Saracevic (1996) noted that relevance involves an interactive, dynamic establishment of a relation by inference, with intentions toward a context. Relevance may be defined as a criterion reflecting the effectiveness of exchange of information between users and systems in communicative relation, all within a context.

2-5. Relevance Measurement

Measurement of relevance has been widely debated in information science literature since the 1960's. The nature of relevance and the role of users in relevance judgement are clearly questions that will not go away in the near future (Tague-Sutcliffe 1996). Although the measurement and assessment of relevance has been a vital concern for IRS designers and developers and also for researchers and practitioners, there is no clear guidance yet about how to evaluate interactive IRS performance in terms of relevance (Su 1992).

Harter (1996) noted that the effect of variations in relevance assessment on the evaluation model on which retrieval performance is assessed - that is, on the measurement instrument - is almost completely unstudied. Relevance assessments are significantly affected by many factors that have been studied experimentally. In addition, judgements are affected by many characteristics of retrieved records and users, and also by other factors that are not fully understood. Several criteria and a great number of measures have been proposed and used for relevance assessment. However, there is a lack of agreement as to which are the best evaluation measures and to what extent they are affected by variability of relevance judgements.

2-6. Models

Despite the fact that the term "model" is widely used in systems research literature, Greenberg and Crissey (1976) stated it does not have a single, clear, unequivocal meaning. Widman, Loparo and Nielsen (1989) define a model in the following words:

> "A model is a representation or abstraction of an actual object or situation. It shows the relationships (direct or indirect) and interrelationships of an action and reaction in terms of a cause and effect. The model, to be complete, must be representative of those aspects of reality that are being investigated" (P. 28).

Rothenberg (1986) stated a model allows us to use something that is simpler, safer, or cheaper than reality for some purposes. A model is an abstraction of reality in the sense that it cannot represent all aspects of reality.

Burch and Strater (1974) categorised models into five classes. These five classes distinguish models according to function, structure, time reference, uncertainty reference and generality. Table 2-1 shows the classification scheme devised by them.

Table 2-1. The Classification of Models

Function	
1. Descriptive	Descriptive models simply provide a "picture" of a situation and do not predict or recommend.
2. Predictive	Predictive models indicate that "if this occurs, then that will follow. They relate dependent and independent variables and permit trying out "what if" questions.
3. Normative	Normative models are those that provide the "best" answer to problem. They provide recommended courses of action.
Structure	
1. Iconic	Iconic models retain some of the physical characteristics of the things they represent.
2. Analogue	Analogue models are those for which there is a substitution of components or processes to provide a parallel with what is being modelled.
3. Symbolic	Symbolic models use symbols to describe the real world.
Time Reference	
1. Static	Static models do not account for change over time.
2. Dynamic	Dynamic models have time as an independent variable.
Uncertainty Reference	
1. Deterministic	For a specific set of input value, there is a uniquely determined output that represents the solution of a model under conditions of certainty.
2. Probabilistic	Probabilistic models involve probability distributions for inputs (or processes) and provide a range of values of at least one output variable (having a probability associated with each value). These models assist with decisions made under conditions of risk.
3. Game	Game theory models attempt to develop optimum solutions in the face of complete ignorance or uncertainty. Games against nature and games of competition are sub-classifications.
Generality	
1. General	General models for business are models that have applications in several functional areas of business.
2. Specialised	Specialised models are those that have application to a single functional area of business

Source: Burch and Strater (1974)

Models can be characterised in many alternative ways, but most suggest categorisation specific to the application areas under consideration. Models may be characterised in terms of their form or structure, their relationship to reality, their purpose, the way they interact with their users (Greenberg and Crissey 1976), the way they are used, their assumptions about certainty of their data (i.e., deterministic vs. probabilistic models), their treatment of time (static vs. dynamic), the kinds of questions they can answer, the kinds of answers they give and so on (Rothenberg 1986). The message, which emerges from these efforts, is that no one model is complete and perfect for all purposes (Quade 1985).

There is a question that what determines which type of model should be used. It is impossible to evaluate or use a model without understanding its purpose. In addition, it is almost impossible to prevent using a model for purposes for which it may be highly inappropriate. The following section concentrates on one particular form of model - a cognitive model- not because it is the best, but because it is related to the context of this study.

2-6-1. Information Retrieval Models

Research in IR interaction emphasises searching as an interactive task and user's interaction with IR systems. A number of interactive IR models have been developed over the years to represent the nature of interactions between user and IRS.

In the traditional model, system and user are represented as two elements of IR and the IR processes intend to match the user's information needs with the data stored. Traditional model of IR is based on the notion that IR deals with two separate tracks system and user. Although traditional model of IR implies interaction, it does not address the interactive processes directly. Moreover traditional models of IR do not account for dynamic nature of IR process and ignore the complexity of interaction. From the perspective of user oriented approach and in order to address the inadequacies of traditional model, a number of researches such as Belkin, Ingwersen, and Saracevic directly address the problems of interaction.

In this section and next sections three models that attempt to describe dynamic nature of IR interaction are explored. These three models are: Ingwersen's (1996) cognitive model, Belkin's (1996) episode model and Saracevic (1996) stratified interaction model. These models reflect the interactive nature of IRSs and explain the reasoning processes behind user's interaction with IRSs at the cognitive level.

Ingwersen's cognitive model concentrates on identifying processes of cognition, which may occur in all the information processing elements. He shows that within each area of his model, the functions of the information user, the document author, the intermediary, the interface and the IR system are the result of an explicit or implicit cognitive model (Ingwersen 1996). Ingwersen's cognitive model is shown in Figure 2-1.

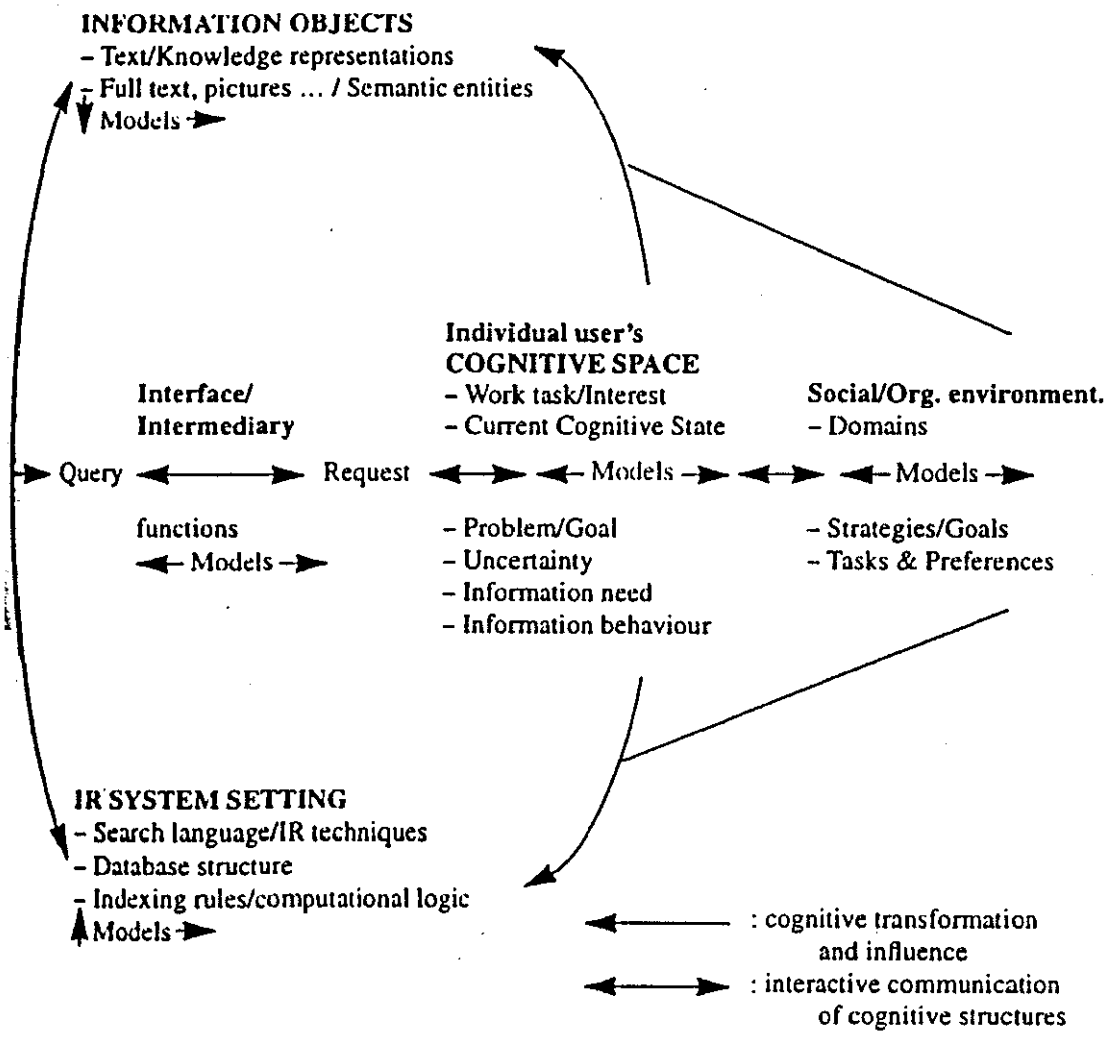


Figure 2-1. Ingwersen's Cognitive Model (Ingwersen 1996)

Wilson (1999) noted that the "focus on cognitive structures and the idea of poly-representation", is an important strength of Ingwersen's model. Saracevic (1996) stated that the "weakness is in that it does not provide for testability... and even less for application to evaluation of IR systems"

Belkin's (1996) episode model focuses on the actions carried out in an information search, from scanning to searching, within a framework of three other dimensions: 'goal of interaction'; 'mode of retrieval' and 'source considered'. The model shows how many of same events in IR interaction repeat themselves for this, the repeating frames display the cyclic nature of interaction over time. Wilson (1999) noted that the focus of Belkin's model is on the design of IR systems. Saracevic (1996) stated "the strength of this model is that it directly addresses interaction, and goes on to specify that there are a number of types of interaction". Belkin's model could be testable and probably could be used in evaluation of IR systems and interactions. However, Saracevic (1996) noted that one of the weaknesses of the model might be the limitation of practical application. A general characterisation of information-seeking goals, a related cognitive task analysis and empirical observations of typical information-seeking situations are some of requirements of Belkin's episode model. These requirements can be regarded as limitation of practical application. Moreover, it is not clear at all that a user actually act according to script (a plan for dialogue between the user and system). Belkin's episode model is shown in Figure 2-2.

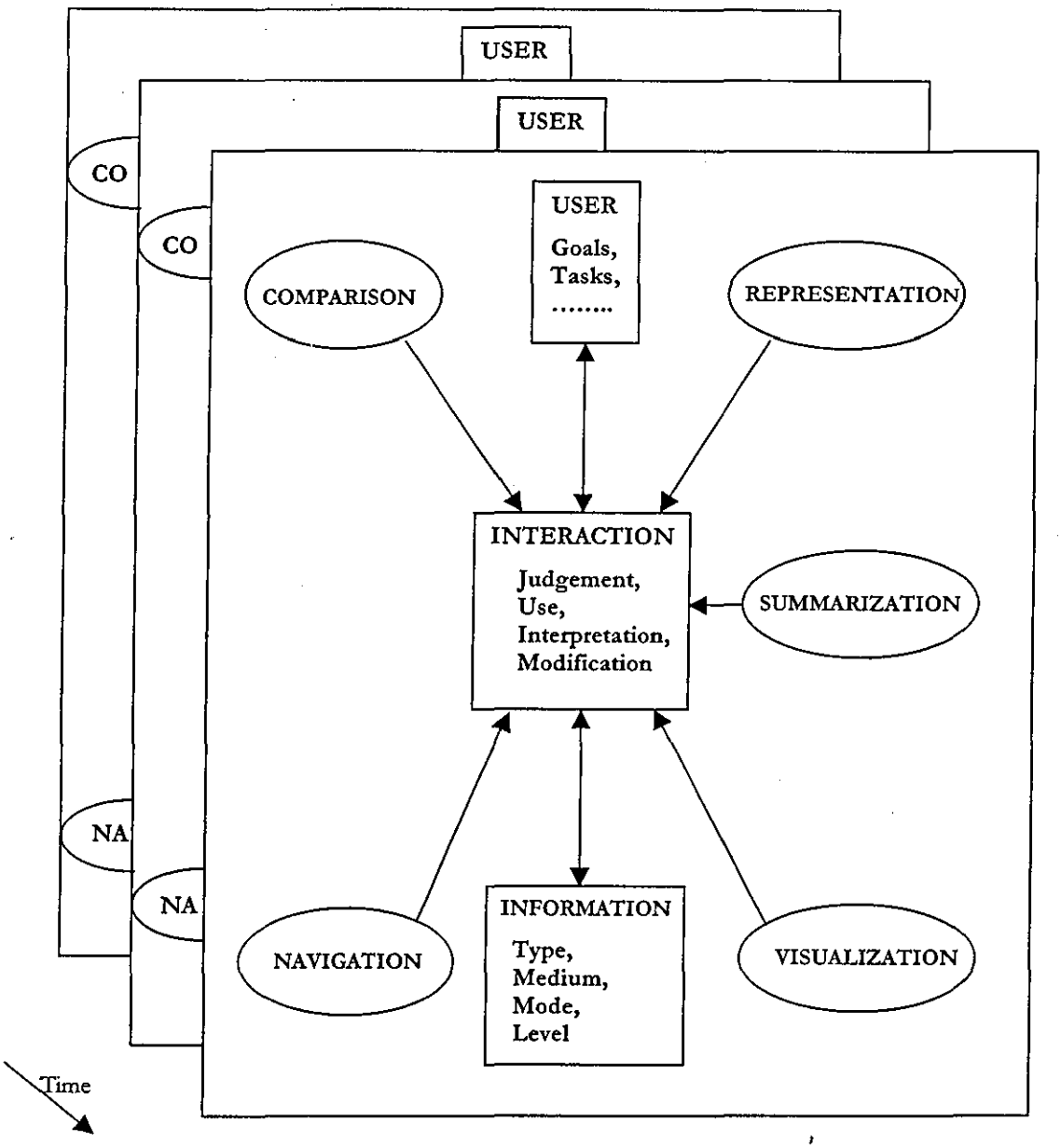


Figure 2-2. Belkin's Episode Model (Belkin et al. 1996)

2-6-2. Cognitive Science Theory

Cognitive science is a science of mind, of intelligence, of thought, a science concerned with knowledge and its uses (Norman 1981). It focuses on understanding knowledge and cognitive processes. The critical aspect of cognitive science is the search for

understanding of cognition, be it real or abstract, human or machine. The main goal of cognitive science is to understand the principles of intelligent, cognitive behaviour.

The importance of the cognitive view in information retrieval, and the power and adequacy of using cognitive instruments in the information retrieval field are largely recognised (Mizzaro 1997). Bruce (1994) stated that while information science is concerned with a complete view of information behaviour, the interaction of an information user with the information system is key. This interaction has upon the schema[1] that individuals apply to relevance estimation is significant. Bruce (1994) also noted that the notion of individualised schema or knowledge structures[2], which mediate information-processing behaviour, is central to the cognitive viewpoint. Ellis (1996) stated that the basic features of cognitive approach to information retrieval system design have been described. The construction of a model of the user in the system and derivation of this model from cognitive characteristics of the user are two important features of cognitive approach.

2-6-3. Cognitive Relevance

Cognitive approach in IR research started at the end of the 1970s according to Ingwersen (2001). He further stated that the problematic issues concerned with the concept of relevance have been demonstrated in TREC and other investigations and experiments. Theories of relevance are moving from a systems approach to a user- and socially-oriented approach (Cosijn and Ingwersen 2000). Relevance in IR does not have a single definition, and it can be viewed as a key point of a system of relevance. Saracevic (1996) expressed different manifestations of relevance (See Table 2-2). These manifestations are discussed in more detail later in this section.

---

1 In psychology, a pattern imposed on complex reality or experience to assist in explaining it, mediate perception, or guide response (Bruce 1994).

2 Knowledge structures are mechanisms, which allow individuals to connect and relate concepts (Bruce1994).

Table 2-2. Manifestations of Relevance (Saracevic 1996)

Relevance	Describe a relation between
System/Algorithmic relevance	Query and Information objects (texts)
Topical relevance	Subject/topic expressed in a query and Subject/topic covered by Information objects
Cognitive relevance or pertinence	State of knowledge and cognitive information need of the users and information objects
Situational relevance or utility	Situation, task or problem at hand and Information objects
Motivational / affective relevance	Intends, goals and motivations of the user and information object

Cosijn and Ingwersen (2000) plotted these manifestations of relevance against the attributes of relevance (both as defined by Saracevic, 1996), and then discussed a new dimension of relevance by considering connections between them. The connection between attribute of relevance and manifestations of relevance is shown in Table .

Cosijn and Ingwersen (2000) have placed affective relevance, not as an attribute nor as a manifestation, but as a dimension in line with time. Affective relevance is highly individual and personal, and therefore very subjective. They argue that affective relevance influences all other subjective types of relevance. Affective relevance, under various labels, has been studied for quite some time (Schamber 1994). Barry (1994) defined a similar type of relevance labelled as "criteria pertaining to the user's beliefs and preferences". In her investigation, she identified subjective accuracy/ validity and effectiveness. Subjective validity is defined as the extent to which the information in the document supports the user's point of view, and affection refers to emotional responses to any aspect of the document. Cosijn and Ingwersen (2000) also regard socio-cognitive relevance as a subjective type of relevance determined by an individual in interaction with others within a community.

Table 2-3. Attributes and Manifestations of Relevance (Cosijn and Ingwersen 2000)

Attributes of relevance	Manifestations of relevance				
			⇔ Affective Relevance ⇔		
	Algorithmic	Topical	Cognitive/pertinence	Situational/utility	Socio-cognitive
Relation	Query ⇒ Information objects(future-based)	Subject/topic expressed in query ⇒ Information objects	State of knowledge/cognitive information need ⇒ information objects	Situation work task or problem at hand as perceived ⇒ Information objects	Situation, task or problem at hand as perceived in socio-cultural context ⇒ Information objects
Intention	(a) System dependent (b) Intent/motivation behind algorithm	(a) User/assessor expectations. (b) Intent/motivation behind query	Highly personal and subjective, related to information need, intentions and motivations	Highly personal and subjective or even emotional. Related to goals, intentions and motivations	Personal, subjective/org. strategy. Related to user's experience, traditions, scientific paradigms
Context	Tuning search engine performance (e.g. TREC)	All types of subjective relevance are, by definition, context-dependent (user's/assessor's context)			
Inference	Weighting and ranking functions	Interpretation of aboutness and subject matter at semantic level	Subjective and individualised process or cognitive/pragmatic interpretation, selection and filtering	User's ability to utilise information objects in a way meaningful to user	Users' (or group's) ability to utilise information objects, meaningful to environment
Interaction	Automatic relevance feedback or query modification	Relevance judgements are content dependent	Relevance judgements are content, feature, form and presentation dependent	Including interaction *with* environment	Including interaction *within* environment

Mizzaro (1998) offered what is arguably the most framework for relevance in IR. His framework has four dimensions:

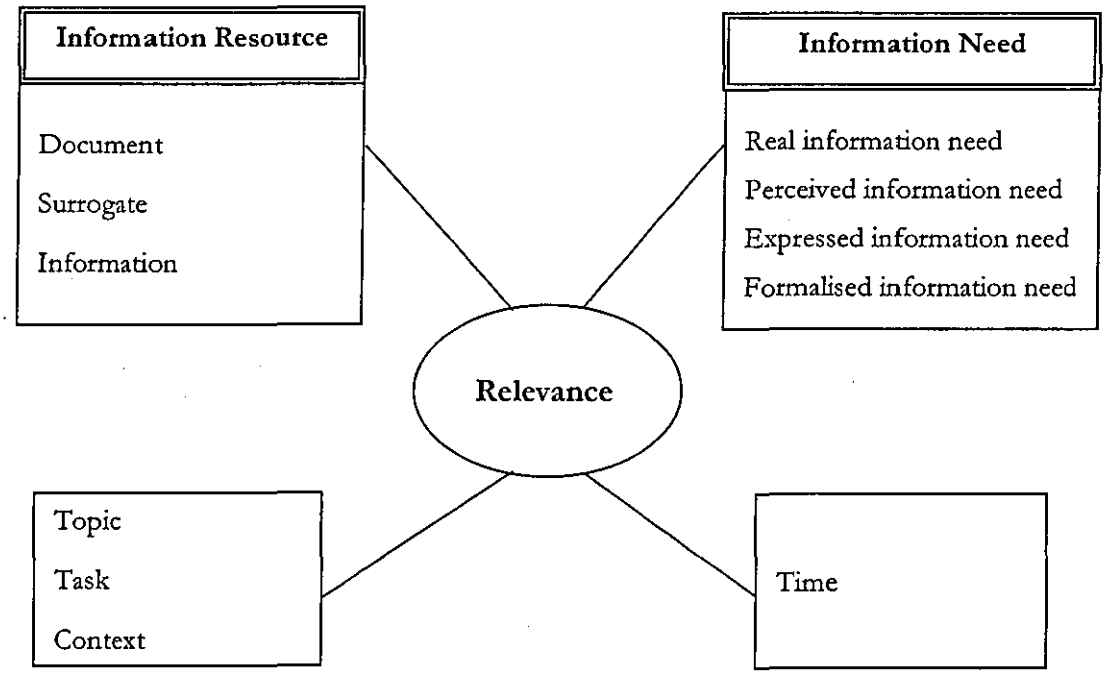


Figure 2-3. Dimensions of Relevance (from Mizzaro 1998)

Some of his ideas are described below (Draper 1998):

Real Information Need (RIN): The need external to the user, perhaps not fully understood by them.

Perceived Information Need (PIN): The mental representation of the query in the user's mind.

Expressed Information Need (EIN): the need expressed in natural language.

Formalised Information Need (FIN): the expressed need formalised in a machine language.

Topic: the central and final goal, the information sought.

Task: the activity for which the information need will be used (in effect, a higher level goal than the information need).

Context: information about the subject domain used during the activity, e.g., meta-information about the topic.

Time: the relevance of an item changes over time: it can fall if the item or a similar one has already been found; it can rise, if information was found about the "context" that now lets the user recognise its importance for the first time. Thus, the relevance of the same document for the same information need can, and often does, change during the course of a search session (Mizzaro 1998).

Mizzaro's framework is a comprehensive one. However, there are some issues that the framework may not cover properly. This framework supposes that relevance is an abstract quantity, based on binary judgement and that has some real absolute value. However, in this study, relevance regarded as multidimensional.

Mizzaro's approach is to have an ordered dimension of information need. It seems the order is that of time and cause (Real Information Need comes first, and then is made from it), but sometimes the request may come first and the Perceived Information Need be created from it. Overall, then, Mizzaro's approach does not cover relevance sufficiently comprehensively.

2-6-4. Cognitive Models of Relevance

The purpose of specifying a retrieval strategy is to retrieve as many as possible of the relevant documents, whilst at the same time retrieving as few of the non-relevant as possible. Intellectually, it is possible for a human to establish the relevance of a document to a query (Van Rijsbergen 1979). However, to study relevance in detail we need to construct a model of it. Van Rijsbergen (1979) noted that most research in information retrieval could be argued to have been concerned with different aspects of such a model.

In recent years, there has been a growing consensus that the entire human-computer system ought to be viewed as an adaptive, cognitive system, in order for effective interfaces and even whole systems to be designed. The cognitive viewpoint implies that

the processing of information is always mediated by some kind of a model of the world (Froelich 1994).

The traditional model for experimental evaluation of IRSs is typified by the second series of studies conducted in Cranfield (Cleverdon, Mills, and Keen 1966). Ellis (1996) traces physical and cognitive paradigms in information science back to the Cranfield tests, and argues because relevance is a relative concept, entirely dependent on the context in which a judgement is made, the possible approaches to models and theories involving relevance are extraordinarily diverse.

There are many ideas of meaning attached to 'mental' or 'cognitive' models. However, there is an agreement that cognitive models can be regarded as images that the components of a system, whether the components are people or machines. It is possible to identify a number of types of cognitive models: Borgman (1986) suggested that mental models in the man-machine interface context refer to the user's model of a system, whereas conceptual models are those which are presented to the user, usually by a systems designer, a trainer, etc. There are user models, which refer to the computer's model of the user and should probably also include the computer's model of the user's model of the system (Daniels 1986).

Recently, Saracevic (1996) proposed a stratified IR interaction model representing IR interaction as the interplay between the user level (cognitive, affective and situational) and the computer level (engineering, processing and content) through an interface level at a surface level (Figure 2-4) (Spink, Greisdorf and Bateman 1998).

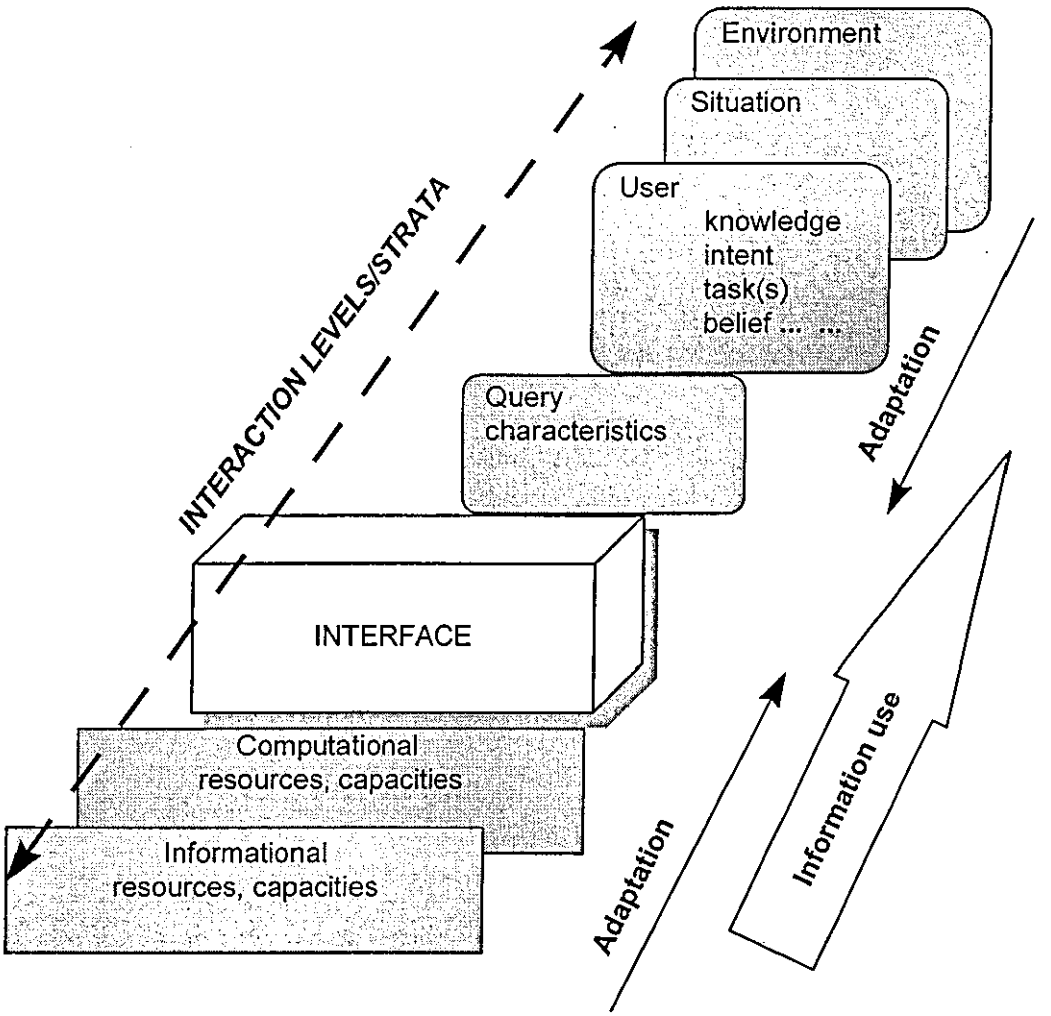


Figure 2-4. Elements in the Saracevic's Model of IR Interaction (Saracevic 1996)

The model borrows from the concepts clarified in human computer interaction research and from notions incorporated in the stratificational theory in linguistics and communication (Saracevic 1996). Saracevic (1996) concluded that the traditional model does not reflect interaction, and that over the three or more decades of research into improvement in IR techniques and in IR evaluation, interaction was ignored.

Spink, Greisdorf and Bateman (1998) proposed a useful concept of relevance as a relationship and an effect on the movement of a user through the repeatable stages of their information-seeking process. They also suggest a three-dimensional model for a user's relevance judgement based on Saracevic's cognitive model. Spink (1999) claimed this model could be a framework for the development of theoretical and empirical research to examine relevance judgements within users' information seeking processes, and to broaden relevance research to include the concurrent exploration of relevance judgement level, region and time.

The first dimension, relevance judgement level, was developed from Saracevic (1996), who suggested that "as a cognitive notion relevance involves an interactive, dynamic establishment of a relation by inference, with intentions toward a context." Within the cognitive level of IR interaction, Spink (1999) and Saracevic (1996) proposed the following manifestations or levels of relevance:

Systems or algorithmic relevance: the relation between a query and information objects (texts) in the file of a system as retrieved, or as failed to be retrieved, by a given procedure or algorithm.

Topical or subject relevance: the relation or non-relation between the subject or topic expressed in a query, and the topic or subject covered by retrieved texts, or more broadly, by texts in the systems file, or even in existence. "Aboutness" is the criterion by which topicality is inferred.

Cognitive relevance or pertinence: the relation or non-relation between the state of knowledge and cognitive information need of the user, and texts retrieved, or in the file of the system, or even in existence. Cognitive correspondence, informativeness, novelty, information quality, and the like are criteria by which cognitive relevance or pertinence is inferred.

Situational relevance or utility: the relation or non-relation between the situation, task, or problem at hand, and texts retrieved by a system or in the files of a system, or even in existence. Usefulness in decision-making, appropriateness of the information in the

resolution of the problem, reduction of uncertainty, and the like are criteria by which situational relevance is inferred.

Motivational or affective relevance: the relation or non-relation between the intents, goals, and motivations of a user and the texts retrieved by a system or in the files of a system, or even in existence. Satisfaction, success, accomplishment, and the like are criteria for inferring motivational relevance.

Each user criterion for relevance judgement could be situated within one of Saracevic's levels of relevance. Spink (1999) commented that some relevance levels (e.g., cognitive relevance) may be hard to measure and some levels may interact with each other (e.g., cognitive, situational and affective) and may therefore be difficult to measure separately.

In the model, the second dimension is depicted within the four regions: (1) highly relevant; (2) partially relevant; (3) partially non relevant; and (4) not relevant. The criteria are situated within one of those four relevance regions.

The plane of judgement exists within an interaction along a plane of time. For example, users make judgements during an evolving information-seeking process or during successive searches. Time may be plotted from the initiation of a user's information need, including the measures associated with the attributes of searches and judgements, in a 3D model. The plane of time consists of two dimensions: Information-seeking Time; and Interaction Time (Spink 1999).

Saracevic's stratified model represents IR interaction as the interplay between user levels and computer level through an interface level. Each level involves different elements and and/or specific processes. This study will focus on cognitive aspects of the human processes involved in relevance judgements.

At the cognitive level, a user interacts with texts[1] in the information resources, considering them as cognitive structures. Therefore, interaction is between cognitive structures, above and beyond the system. Users interpret and cognitively judge the texts obtained, and may absorb the information in them cognitively.

In this study, the relation or non-relation between the state of knowledge and cognitive information need of the user, and texts retrieved, as a form of cognitive relevance will be considered.

2-7. Factors and Criteria of Relevance

In this study, the understanding of relevance judgement is based on the theoretical framework provided by Saracevic (1996), Schamber and Bateman (1999) and Tang and Soloman (2001). Saracevic (1996)'s approach follows previous research exploring users' relevance judgements that have produced many studies examining user criteria for relevant items retrieved from an IRS. Each user criterion for a relevance judgement could be identified within one level of relevance that Saracevic proposed. Many factors have been suggested as affecting relevance judgement. However, this study only deals with those factors, which affect cognitive relevance. For this purpose, previous research can be considered.

There is not a unified definition of relevance. Froelich (1994) stated that the absence of a single definition of relevance does not mean that information scientists cannot determine the diverse criteria that people bring to systems by which to judge its output.

Park (1995) grouped relevance criteria into three categories:

- Internal (experience) context. This refers to the knowledge of the user and his/her understanding of the current information need.
- External (search) context. This category is related to the current research.
- Problem (content) context. This group of criteria includes comparison between the current research problem and research problems described.

---

1 Saracevic (1996) uses the term "texts" as cognitive structure, representations, metainformation about texts, computational resources and possibly other information for adaptation to different kinds of interactions and user intents. It includes images, data ... and their representations.

Schamber (1991) identified 10 categories of criteria. These categories are presentation quality, currency, reliability, verifiability, geographic proximity, specificity, dynamism, accessibility, accuracy, and clarity.

Cool et al. (1993) determined six categories of relevance criteria: topic, content/ information, format, presentation, values, and oneself.

In other research, Barry (1993, (1994) found 23 criteria used by users to make relevance judgement and then grouped them into seven categories: information content, user's previous experience and background, user's beliefs and preferences, other information and sources within the information environment, sources of document, document as a physical entity, and user's satisfaction.

Barry and Schamber (1995, 1998) identified 10 criteria: depth scope/specificity, accuracy/validity, clarity, currency, tangibility, quality of source, accessibility, availability of information, verification, affectiveness, topical appropriateness.

Schamber and Bateman (1996) started their research with 119 criteria and then reduced this to 83. They could group these criteria into five categories: currency, availability, clarity, credibility, and aboutness. Schamber and Bateman (1999) reported on five tests, in which respondents were asked to interpret, sort, select and rank 119 criteria selected from their earlier studies. The results were used to edit the set down to 40 items in nine categories:

- *Topicality.* About my topic

- *Availability.* Easy to obtain, Free or inexpensive

- *Novelty.* Unique or the only source, Original, New to me, Familiar

- *Currency.* Current

- *Quality of Information.* Well-written, Credible, Accurate, Understandable, Consistent, Focused

- *Presentation Characteristics.* Presentation of information, Suitable length, Comprehensive, Suitably general or specific, Detailed, Introductory, Overview

- *Source Characteristics.* Prominent, I know the author personally, I know the source, Reputable, Format of the source, Interactive

- *Information Characteristics.* Describes methods/techniques, Provides examples, Provides, graphics, Statistical approach, Research approach, Provides proof, Controversial, Provides bibliography or links, Provides background or history

- *Appeal of Information.* I like it, Validate my viewpoint, Interesting, Enjoyable.

Spink (1998) stated that cognitive correspondence, informativeness, novelty and information quality (accuracy, journal or author reputation) are the criteria by which cognitive relevance is inferred. It is clear that some of these factors of cognitive relevance will be hard to measure, and that they may interact with factors from other aspects of relevance, and thus may be difficult to measure separately.

2-7-1. Cognitive Factors of Relevance

This study deals with those factors, which affect cognitive relevance. Tang and Soloman (2001) in their laboratory experiment tabulated 15 criteria and grouped them in to three categories: topicality, information quality and cognitive state (Table 2-4). They stated that a criterion might be named differently and perhaps grouped under different categories. It depends on the insights of the scholars and users.

Table 2-4. Relevance Criteria (Tang & Soloman 2001)

Categories	*Topicality*	*Quality Of Information*	*Cognitive Requisite*
Criteria	- Covers the topic - Defines the topic - Provides background Information - Provides factual information and data	- Subject matter is important - Information is timely and up to date - Accuracy and trustworthiness - Clarity and well-written - In-depth presentation of information - Unique Approach	- Understandability - Newness - Similar to what I know - Adds to my knowledge - Information is interesting and enjoyable

2-8. Conclusion

Many researchers believe relevance is one of the most fundamental concepts in the theory of IR process. The concept has been much debated and developed since it was first adopted as the basis of measures of the effectiveness of IRS, and it continues to be used in evaluations today.

Saracevic (1975) pointed out relevance relating to concept within information science where the meaning and use of relevance is widely understood – at least within the context of IRS evaluation. Although use of relevance as the appropriate criterion of IRS effectiveness has been debated for more than four decades, a clear definition within the context of IRS evaluation has not emerged. Some researchers have also challenged the assumptions regarding relevance as a basis for IR evaluation.

In response to the inadequacies of system-oriented approach, efforts on user-oriented studies have been made. The user-oriented perspective seeks to characterise the

relationship between information and the user's problem situation and attempts to account for the various aspects of human cognitive processes used in making relevance judgements. The user-oriented approach fills the gap that exists between the ways that systems function and the ways that users express their needs.

In the user-oriented approach, relevance is an attribute or criterion reflecting the effectiveness of interactive exchange of information between users and IRSs in a dynamic communicative contact.. As a cognitive notion, relevance involves an interactive, dynamic establishment of a relation by inference, with intentions toward a context (Saracevic 1996).

A number of interactive IR models have been proposed by researchers and the literature on IR interaction is growing. In this chapter, three models that attempt to describe various aspects of interaction between system and users were reviewed. These models are: (i) Saracevic's (1996) stratified model of IR interactive; (ii) Belkin's (1996) episode model of IR interaction; (iii) Ingwersen's (1996) cognitive model.

According to Ingwersen (2001), Saracevic's model is "the most comprehensive model" of relevance. Ingwersen (2001) states that Saracevic's model is strongly associated with the cognitive model of Ingwersen and Belkin's episode model.

Saracevic's model represents IR interaction as the interplay between user levels and computer level through an interface level. In this model, relevance manifests itself in different level and each level involves different elements and and/or specific processes.

The three models represent some ideas of interactive IR processes and more research is needed in this growing body of knowledge. There is a limited knowledge about what needs to be modelled. More work needs to be done to determine all aspects of the user and his or her situation. A consideration of work on relevance criteria is helpful in this effort.

Several criteria and a great number of measures have been proposed and used for relevance assessment. However, there is a lack of agreement as to which are the best evaluation measures and to what extent they are affected by variability of relevance judgements. It is important both to the theoretical development of relevance and to the

design of IRSs to investigate the cognitive factors in relevance judgement. This study attempts to do this.

Chapter 3: Research Methods

3-1. Introduction

Recent studies on relevance view IR as a cognitive interaction between human and computer. As a cognitive notion, relevance involves an interactive, dynamic establishment of a relation by inference, with intentions toward a context (Saracevic 1996). This study has focused on cognitive aspects of the human processes involved in relevance judgements. Several criteria and a great number of measures have been proposed and used for relevance assessment. However, there is a lack of agreement as to which are the best evaluation measures and to what extent they are affected by variability of relevance judgements. Saracevic (1996)'s stratified model provides a necessary framework to incorporate relevance cognitive theory and a user approach in measuring relevancy. This Chapter presents some issues regarding the methodology of the research and gives details of the methods adopted to carry out the study. Methodology is a theory of methods as well as theory of observing and analysing, while method refers to the actual design and specific techniques used in the study (Wang and White 1999; Vakkari, Savolainen and Dervin 1996).

3-2. The Research Design

This study used a combination of qualitative and quantitative techniques to produce valid and reliable criteria and factors and their measures. Cassell and Symon (1994) pointed out that both quantitative and qualitative methods can be used together in the same study. The framework can also help to validate exploratory research, investigate systematic differences and test certain hypothesis about data (Creswell and Miller 1997). Two stages were used: exploratory stage and measurement stage.

The predominant approach used was surveys, supported by experimental studies. Wang and White (1999) noted that researchers in information science could use the survey method along with other methods in a single study. This study used surveys combined with experiment and short interviews. The experimental study was conducted among a number of PhD research students of all types in the field of science and engineering (first year, second year and third year and above, full-time between 24 and 42 years old). This research was conducted in two phases: a pilot study, and a series of experiments.

In the exploratory stage, the literature search and pilot study were used. This provided the basic framework and guided the selection of items from a theoretical background (Oppenheim 1992). The literature search was also used to develop ideas and understand the variety of psychometric measurement techniques (Churchill 1995). In summary, literature search and pilot study were used in formulating the problem for more precise investigation, increasing familiarity with the problem and collecting information about a set of criteria and factors that could be used to test hypotheses.

Oppenheim (1992) noted that a pilot study is normally concerned with the conceptualisation and refinement of the research problem. Tang, Shaw and Vevea (1999) used a pilot study for the exploratory stage of their study. In this project, the pilot study comprised a small-scale survey and interviews.

The measurement stage provides an environment for hypothesis testing and for measuring the different importance impact of factors that effect relevance (DeVellis 1991). The main goals of the measurement stage are a) to measure the attitudes of respondents with scale development; b) to investigate the amount of change over time in those attitudes; c) to be able to make inferences quantitatively about the significance of differences in measurement.

3-3. Research Questions

The primary purpose of this study was the identification of appropriate cognitive factors of relevance judgement. Secondary purpose was finding the importance of factors by

measuring the impact of each criterion and factor of cognitive relevance. The final purpose of this study was to investigate the stability of the cognitive factors over time. The following research questions were employed to guide this study:

1. What factors affect cognitive relevance?

2. What is the importance of each of those factors on relevance judgement?

3. Do cognitive factors of relevance judgement remain stable over time?

The first issue that will be addressed is how users make judgements for selecting relevant documents to their needs when using an interactive IRS. In order to explore this issue the first research question with the goal of identifying cognitive factors that influence relevance judgements formulated.

The second research question is deal with determining the importance of relevance criteria and factors - from most to least important - based on the score of each cognitive factor. These scores will be calculated by a statistical method subject to a statistical analysis to determine which factors have most influence on cognitive relevance.

Stability over time of cognitive factors has not been much debated in the studies of relevance criteria. This issue is a parameter that will be addressed. The third research questions designed to examine whether cognitive factors of relevance judgement remain stable over time.

3-4. Variables

Based on the results of pilot study and the study of experts attitude that will be discussed in Chapter 5, it has been proposed that the sixteen criteria (Table 3-1) are the independent variables of this study because they act as determinants of cognitive relevance judgement. Relevance judgement would be the dependent variable because it is influenced by these relevance criteria.

Table 3-1. Relevance Criteria as Independent variables

About my topic	The extent to which information is related to the user's topic
Unique or only sources	The extent to which the document is unique
Original	The extent to which the document is novel
New	The extent to which information is recent or up to date
Informative	The extent to which information is in-depth, provides a summary, interpretation or explanation, provides a sufficient variety of approaches
Easy to understand	The user's judgement is that he/she can understand or follow the information presented
Consistent	The extent to which information is consistent with or supported by other information within the field
Generally or specifically suitable	Information is specific to user's need and has sufficient depth
Know the author(s) personally	The extent to which the user has a personal or professional relationship with the author of a document
Know the journals or conferences	The extent to which a source of the document is well-known or reputable
The authors are eminent	The extent to which each author is well-known as a major researcher in the subject
Methods or techniques have described	The extent to which the document has sufficient detail about techniques or method
Controversial	User is encouraged by the text to engage in two-way interacting with statements made in the document
Good bibliography and links	The extent to which document has a sufficient bibliography and/or links to web resources
Background	The extent to which a document provides background
Validate viewpoint	The extent to which the user agrees with the information presented, or the information presented supports the user's point of view

3-5. Measurement

Measurement is the systematic assignment of numbers to a set of observations to reflect the status of each member of the set in terms of the variable property (Lewis-beck 1994). Blalock and Blalock (1982) pointed out problems of measurement are certainly crucial in the development of any science. They defined measurement as "the assignment of numbers to aspects of object or events according to rules"(P. 22).

Another definition of measurement that is more relevant to the social sciences is that measurement is a process involving both theoretical as well as empirical consideration:

> "Measurement focuses on the crucial relationship between the empirically grounded indicator(s) -that is, the observable response- and the underlying unobservable concept(s)" (Lewis-Beck 1994: 2).

Pors (2000) stated that measurement in information retrieval is more complicated because a lot of the measures such as the measure of relevance are composite measures. A composite measure is a relation of two or more basic measurements such as the measure of relevance, which can be measured in different ways but never directly.

Pedhazur and Pedhazur (1991) claimed that a great advantage in using measurement when it is compared with alternative approaches to the description of, or differentiation among, a set of objects is that one may apply some powerful tools of mathematics to the study of phenomena. He added that measurement constitutes a matching of numbers to aspects of objects. So, it is necessary to know what is matched to what, how the numbers can be interpreted meaningfully, and what mathematical manipulations may be applied.

A *scale* is a set of numerical values assigned to subjects, objects, or behaviours for the purpose of quantifying the measuring qualities. Scales can be used to measure attitudes, values, and interests. They can be used to measure the degree to which an individual possesses the characteristic of interest (Torgerson 1958).

Lewis-beck (1994) stated *Attitude* is an integral part of the personality (as to the way we think, feel, perceive, and behave toward a cognitive objective). Attitudinal scales measure the intensity of respondents' attitudes towards the various aspects of a situation

or issue and provide techniques to combine the attitudes towards different aspects into one overall indicator. There are three major types of attitudinal scale:

- the summated rating scale, also well known as the Likert scale;
- the equal-appearing-interval scale or differential scale, also known as the Thurstone scale;
- the cumulative scale, also known as Guttman scale.

The Guttman scale is one of the most difficult scales to construct and therefore is rarely used. Arranging statements in a perfect cumulative order is the main problem in using it (Kumar 1999).

In this study, two types of scales were used: the Likert scale and Thurstone scale.

The summated rating scale, more commonly known as the Likert scale, is based on the assumption that each item on the scale has equal 'attitudinal value', 'importance' or 'weight'. This type of scale was used in the second phase of the study.

In the first stage, the equal-appearing-interval scale or Thurstone scale was used. The central idea of a Thurstone scale is that, for each statement on the scale, if a person agreed with it, they were given a score equivalent to the strength of that statement or item. The Thurstone scale calculates a 'weight' or 'attitudinal value' for each item. The weight of each item is calculated on the basis of ratings assigned by a group of judges

3-6. Measurement Instrument Design

Spector (1992) noted that the development of a scale is a multi-step process. Figure 3-1 illustrates the five steps in this process:

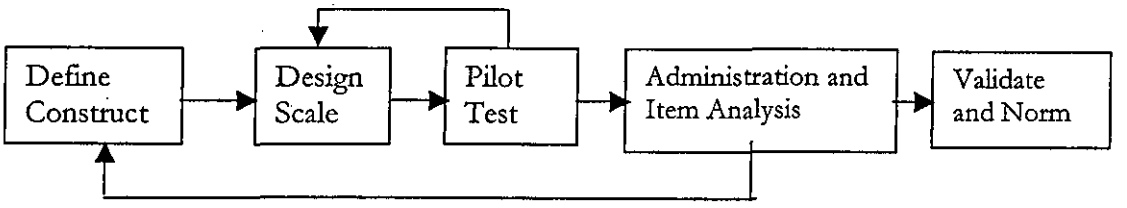


Figure 3-1. Step Process of Scale Development (from Spector 1992)

In this study, these five steps were undertaken. In the literature review, the construct was clearly and precisely defined. What the scale intended to measure was cognitive criteria that contributed to relevance judgement. Second, the exact format of the scale, a five-points Likert scale, was designed and an initial item pool written. Third, the initial version of the scale was pilot-tested with five respondents. They were asked to critique the scale. Fourth, a series of experiments were conducted. For each experiment item, analysis was performed and Cronbach's coefficient alpha was calculated. The scale was internally consistent. In the last step, the scale was validated.

Factor analysis was used for validating the scale. Factor analysis derives its factors by analyzing the pattern of covariation (or correlation) among items (Spector 1992). To compile norms, reliability and validity tests were performed on the norms of the instrument. In addition, descriptive statistics (mean and standard deviation) were calculated.

IR research is involved in cognitive structures, both in human minds and texts and particularly with regard to their interaction with IR (Ingwersen 1996). In this study, we conducted a series of experiments to evaluate the effectiveness of Interactive IRS by measuring relevancy. We adopted Saracevic's stratified model to provide a necessary framework to incorporate relevance cognitive theory and a user approach in measuring relevancy. The questionnaire was developed based upon issues identified from the literature and is improved by adopting a number of suggested enhancements. In order to identify and correct questionnaire problems, a pilot study was conducted. The pilot study involved a smaller sample of individuals. In this way, the questionnaire was pre-tested to ensure clarity and to remove ambiguities in language and meaning.

3-6-1. Email Survey

This research used electronic mail technology to send and receive the questionnaire. The emergence and continued growth of information technology has affected many areas as well as research. Questionnaire design with computers has affected data collection. Using electronic mail systems becomes more common in research studies and it may become an appropriate medium for collecting data. In particular, this kind of survey is also commonly used for the questioning of experts on a specific research area (Tse 1998). Ease of administration and the elimination of the tasks of folding and staffing effort and costs for both researcher and respondents are the biggest advantages of email survey. While physical transfer of messages from one place to another place is required in a mail survey, the transmission of email messages can occur almost instantaneously (Tse 1998). This advantage of an email survey increases the speed of data collection. Moreover, given the immediate return of undeliverable messages, the researcher does not need to wait for several days until undelivered mail is being returned. Compared to the traditional mail survey, response rates of email surveys can be higher, lower, or unchanged (Walczuch and Hofmaier 1999). Another good way was to have the questionnaire on the web. In a web survey, the questionnaire is presented in a user-friendly format. However, email survey was used to obtain responses from people who may not visit the web or who access the web using a text-based browser. Email survey work well because the user can answer the survey offline. However, the fact that they are easy to ignore and delete are disadvantages of email survey or web survey that may decrease the response rate of these surveys. To prevent increasing downloading time to view and open the attachment, the questionnaire in my study formed part of the body text of the email.

3-6-2. Pilot Study and Full-scale Experiments

The pilot study involved experiment, survey and interview. Five research students participated in the pilot study. They were asked to perform a search with a self-selected topic using ISI Web of Science and asked to make relevance judgements. In the stage following, they were asked to fill in a printed copy of the questionnaire (See Appendix 2). A short interview was also used to improve data quality and its purpose was to clarify

questions in the questionnaire and to identify any difficulties. The period of data collection for the pilot study was from April to May 2000.

For the full-scale study, a series of experiments were conducted. Those were multiple-phase and longitudinal, i.e., the study was conducted over a period of time and in multiple phases. Thirty research students were recruited from two universities in the UK for this part of the research.

3-7. Population and Sampling Method

One of the most critical influences on the quality of data that emerges from a survey is the choice of people to respond to questions. Because of the exploratory nature of the pilot study, a non-probability sampling technique (convenience sampling[1]) was used.

For the main study, real users with real information needs who were research students in science and engineering at Loughborough University or the University of Nottingham were selected. The number of participants in an experiment determines the power, or the probability of obtaining a significant result for a statistical test. In reality, however, it requires compromises between theoretical sampling and practical limitations such as costs and time (Oppenheim 1992). The numbers of participants used in some previous studies are shown in Table 3-2.

[1] A convenience sample is one where you get because people who are willing to complete the survey and are also available when you need them (Fink 2000).

Table 3-2. The Number of Subjects in Some Studies on Relevance

Study	Subjects	Environment	Relevance Judgement
Barry (1994)	18	Academic	Bibliographic citations
Cool et al. (1993)	300	Course assignment	Documents
Park (1995)	24	Academic	Bibliographic citations + Documents
Schamber, Eisenberg and Nilan (1990)	30	Academic	Information sources and citations
Schamber and Bateman (1996), Schamber and Bateman (1999)	500	Course assignment	Documents
Su (1992)	12	Academic	Bibliographic citations (Dialog)
Wang and Soergel (1998)	25	Academic	Bibliographic citations

The samples for my three experiments were 30 Ph.D. research students (19 in Loughborough University and 11 from the University of Nottingham); (7 female and 23 male). Twenty-one participants out of the thirty were studying science and nine engineering. Participants were motivated to take part in the study by the offer of a ten-pound book token. Participants were assured that no name would be identified in the final report and that the data would be kept confidential.

3-7-1. Participants in the Pilot Study and Full-Scale Experiments

Five research students in the Department of Information Science at Loughborough University were used in the pilot study. Each student was invited to participate in the pilot study by email or by an oral invitation. The respondents of the pilot study were

chosen for pre-testing the potential effectiveness of the questionnaire, because of their subject knowledge and also their willingness to co-operate.

30 research students participated in the full study. The subjects were all Ph.D. students of science and engineering at Loughborough University or the University of Nottingham. They were recruited by email messages posted on the General Noticeboard of Loughborough University, and advertisements posted to the George Green Library general noticeboard of the University of Nottingham.

Participants had a real information need relevant to their PhD thesis. All had done previous research on their topics. An invitation letter (See Appendix 5) was emailed and after their agreement to participate, a questionnaire was sent to them (See Appendix 6).

3-8. Data Collection

A survey questionnaire was designed for collecting data from participants. In the pilot study, data were collected by means of a questionnaire (printed copy) containing 40 items (Appendix 2). The questionnaire attempted to gather information about the search already performed on the topic by the participant, the attitude of participant about relevance criteria, and demographic data.

3-8-1. Pilot Study

Data were collected by means of a questionnaire (See Appendix 2) containing 40 items. The majority of these were Likert-type statements on a scale from "strongly agree" to "strongly disagree". Other questions asked for some information about user characteristics and his/her search strategies. Forty questions were used, of which 34 related to relevance judgement under seven broad headings:

- One measures topicality.

- Four measure novelty.

- Seven measure quality of information.

- Six measure presentation of information.

- Five are taken as indicators of source characteristics.

- Seven are taken as indicators of information characteristics.

- Four are taken as indicators of appeal of information.

The categorising of these items into the seven factors was then carried out through factor analysis.

3-8-2. Full-scale Experiments

A questionnaire (See Appendix 6) was designed. This questionnaire had three sections. The first section of the questionnaire had six questions and investigated characteristics of the search session, e.g., number of search strategies used, the number of search results, and the number of completely and partially relevant records. The second section of the questionnaire was designed to gain data that could measure the intensity of respondents' attitudes towards the cognitive relevance criteria. This section contains 23 statements; 16 statements were about cognitive relevance criteria; the 23rd statement invited respondents to add new criteria if they thought they were necessary. The final section of the questionnaire had five questions about personal information of respondents such as research topic, department, gender, year of the study and age.

3-8-3. Database

The Web of Science was used for searching tasks. The form of document surrogates, whether citations only or citations with index terms, abstracts, full texts, etc., should be appropriate to the hypothesis under test. Within the Web of Science, the Science Citation Index Expanded (SCI-EXPANDED) 1981-present, Social Science Citation Index, and Art and Humanities Citation Index bibliographic databases were used.

SCI was hosted by BIDS at the University of Bath between 1990 and 1999. The BIDS ISI was replaced by the Web of Science (WoS) service supported by MIMAS at the University of Manchester in late 1999.

The Science Citation Index Expanded is a major multidisciplinary database. Users can search by subject, author, abstract, journal, and/or cited reference lists. It covers much of the journal literature of the science. Furthermore, there are links from an article's bibliographic display to those articles subsequently published and indexed in the database that have cited that article. It indexes more than 5,700 journals across 164 scientific disciplines and contains over 17 million articles. Moreover, it contains searchable author abstracts for approximately 70% of the articles in database. Some of the disciplines covered include: Agriculture, Astronomy, Biology, Chemistry, Computer Science, Mathematics, Medicine, Pharmacology and Physics (Atkins 1999).

This IR system was chosen because it has been widely used in the UK universities, and its databases provide flexibility in supporting the search questions of science and engineering research students. Another reason for choosing this IR system was the accessibility of the database for all academics in the UK without any limitation. In the full-scale experiments, data were collected by means of an electronic questionnaire containing 28 items (Appendix 6). The majority of these were Likert-type statements on a scale from "strongly agree" to "strongly disagree". Other questions asked for some information about user characteristics and his/her search strategies.

Each experiment was replicated three times at 3 monthly intervals. The entire process was structured into two sessions: an online search session and a questionnaire answering session. First, each participant was asked to formulate his/her self-selected research topic and conduct searches on the ISI Web of Science. Second, each participant was asked to fill a questionnaire. The data collection process took about 45 minutes to one hour per subject.

3-9. Data Analysis

The data analysis was conducted in two phases: data processing and statistical analysis.

3-9-1. Data Processing

In this study, the main purpose of the questionnaire, and of each survey as a whole was measurement. The first objective of the data processing phase of the study was to get to the point where all the responses to questionnaire could be turned into numbers. A series of checking operations were performed on the complete data set. Frequency distributions on the main sampling variables and range checks for the variables (five-point attitude scale) were checking operations. These operations have been done to get a completely clean the data set.

3-9-2. Statistical Analysis

Three main functions of statistical techniques are: a) summarising the collected data and tabulating the data as clearly and effectively as possible, b) generalising the results from a sample to the entire population, and c) using a powerful method for data simplification, data exploration or data reduction (Howitt and Cramer 2000). In this study, a number of techniques are employed. These techniques are descriptive and inferential analysis, factor analysis, and Structural Equation Modelling (SEM).

3-9-2-1. Descriptive Analysis

Descriptive statistics techniques summarise the data collected from the sample of participants of this study. Descriptive statistical procedures include measures of central tendency and variability presented both numerically and visually (e.g. graphs) (Goodwin 2002). Powell (1997) pointed out that descriptive statistics form the predominant type of

data analysis. There are a number of basic functions that researchers in information science can perform including frequency distributions, measures of central tendency and variability. In part of this study, the median and the mean provided descriptions of the centre of distribution. The standard deviation provided a measure of variability.

3-9-2-2. Inferential Analysis

Much of the power of statistical analysis comes from the ability to draw general conclusions from a limited but representative set of data. In contrast to descriptive statistics, which simply summarise and describe the data, inferential statistics can perform certain more sophisticated functions. They are most commonly used to estimate population parameters or characteristics based on random sample statistics, and to test hypotheses using tests of statistical significance (Powell 1997). Statistics are used to test the null hypothesis, or the hypothesis of no relationship. To support a hypothesis that two or more variables are related, one must demonstrate that they are not unrelated. One must demonstrate that it is safe to conclude that the null hypothesis is wrong so as to conclude that the variables really are related (Powell 1997). In this study, the level of measurement was ordinal and the normality of population could not be assumed (because of small sample size). For these reasons, nonparametric statistics are appropriate for inferential analysis.

Nonparametric Statistics. To understand the idea of nonparametric statistics first requires a basic understanding of parametric statistics. A factor that often limits the applicability of parametric tests based on the assumption that the sampling distribution is normal is the size of the sample of data available for the analysis. Parametric statistics require the assumption of a normal population or distribution. If the sample drawn from the population has 100 or more cases, the normality assumption can almost always be relaxed (Powell 1997). Blalock and Blalock (1982) stated if the sample size is somewhat less than 100, the researcher should use statistical tests cautiously unless the approximation to normality is known to be good. In other word, the researcher can assume that the sampling distribution is normal even if he/she is not sure that the distribution of the variable in the population is normal, as long as our sample is large enough (e.g., 100 or more observations). However, if the sample is very small (30 in this

study), then parametric tests can be used only if the researcher is sure that the variable is normally distributed, and there is no way to test this assumption if the sample is small.

As mentioned above, a problem in measurement is that applications of tests that are based on the normality assumptions are further limited by a lack of precise measurement. Without going into too much detail, most common statistical techniques such as analysis of variance assume that the underlying measurements are at least of interval, meaning that equally spaced intervals on the scale can be compared in a meaningful manner. However, in this study, this assumption is not tenable, and the data represents a rank ordering (ordinal).

Nonparametric methods. Nonparametric methods were developed to be used in cases when the researcher knows nothing about the parameters of the variable of interest in the population (Howitt and Cramer 2000). In more technical terms, nonparametric methods do not rely on the estimation of parameters (such as the mean or the standard deviation) describing the distribution of the variable of interest in the population. Therefore, these methods are also sometimes (and more appropriately) called parameter-free methods or distribution-free methods (Ferguson and Takane 1989; Goodwin 2002).

Nonparametric tests were used to test relationships between variables. The Spearman R correlation coefficient was calculated (Siegel and Morgan 1996). Spearman R assumed that the variables under consideration were measured on at least an ordinal (rank order) scale. This is the appropriate nonparametric statistic for testing the relationship between categorised variables (Coakes and Steed 2001; Ferguson and Takane 1989, Vaughan 2001).

For testing the reliability and validity of the measurement scale, a correlational analysis was performed. In addition, a multivariate approach that includes a number of descriptive and inferential techniques were employed. Multivariate analysis was appropriate for this research because the variables involved in the study are sets of variables with a number of variables with each set. The multivariate procedures that were used were factor analysis and Structural Equation Modelling (SEM).

3-9-2-3. Factor Analysis

Another multivariate technique called factor analysis was used to generate statistically valid clusters of criteria based on the participants' importance ratings. The following reviews some of the important concepts related to the method of factor analysis.

Factor analysis refers to a statistical technique whose objective is to represent a set of variables in terms of a smaller number of hypothetical variables. The fundamental assumption of factor analysis is that there are underlying factors responsible for the covariations among the variables (Spector 1992). Typically, a factor analysis is performed when a person has a large number of variables and wants to obtain a sense of the general dimensions in these variables. The purpose of factor analysis is to find a new set of variables, fewer in number than the original variables, which express that which is common among the original variables (Howitt and Cramer 2000; Goodwin 2002).

In this study we had a collection of variables and needed an idea about what constructs might be used to explain the intercorrelations among these variables. In addition, we wanted to determine the validity of the measurement scale. Factor analysis operates on the correlations among variables, and starts with a matrix expressing the correlations of each variable with every other variable (Howitt and Cramer 2000). Factors are extracted based on the correlation coefficient values. The factor solution is further rotated and more factors are extracted until an optimal factor solution is reached.

3-9-2-4. Reliability and Validity

Reliability and validity are technical terms and to some extent are interconnected. These are concepts derived from measurement theory and from psychometrics (Oppenheim 1992). Baker (1988) noted that the most important criterion of the goodness of a measure is its validity. It is also important that a measure be consistent. This consistency in measurement is referred to as reliability. Zeller and Carmines (1980) pointed out reliability concerns the extent to which an experiment yields the same results on repeated trials.

Methods to Measure Reliability. To understand reliability coefficients, a brief discussion of the components of a score will be helpful. An observed or obtained score on an instrument can be divided into two parts.

Observed Score = True Score + Error

An instrument can be said to be reliable if it accurately reflects true scores. In other words, an instrument can said to be reliable to the extent that it minimises the error component. So, the reliability coefficient is the proportion of true variability to the total obtained variability (Zeller and Carmines 1980). For example, if the reliability coefficient is .85, this means that 85 percent of the variability in obtained scores could be said to represent true individual differences and 15 percent of the variability is due to random error. Correlation scores should be at least .70 (retest after a couple of weeks) and lower for long term retest (more than two months) (Goodwin 2002).

Repeated use of the instrument (stability[1]), similarity of items (homogeneity[2] or internal consistency) and equivalence[3] of instruments are three possible methods of testing reliability. In this study, the reliability of scale was considered in two ways:

Internal consistency reliability. This means that multiple items, designed to measure the same construct, should inter-correlate with one another. The more consistent the results given by repeated measurements, the higher the reliability of the measuring procedure (Zeller and Carmines 1980).

Test-retest reliability. This means that a scale yields consistent measurement over time and each respondent should get about the same score upon related testing. In this study, the experiment was carried out three times at 3-month intervals.

---

1 Stability means that the same results are produced with repeated testing.

2 Homogeneity or Internal Consistency means that items in the instrument measure the same concept.

3 Equivalence means that instrument produces the same results when an equivalent instrument is used or there is consistency among researchers using the same instrument.

Methods to measure validity. The validity of a scale is the extent to which it measures the construct that it is designed to measure. Reliability does not necessarily imply validity. Validity refers to whether a measure actually measures what it claims to measure. (Aron and Aron, 1999). According to Bryant (2000), "validity concerns whether a particular inference or conclusion that one wishes to make is accurate, reasonable, or correct."(P.101).

There are three types of validity: content (factorial) validity, construct (discriminant) validity, and criterion validity.

Content validity concerns the degree to which a measurement scale assesses all relevant aspects of the conceptual domain that it is intended to measure. This type of validity can be evaluated using multivariate statistical procedures, such as exploratory factor analysis (Bryant 2000).

Construct validity concerns whether a given measure actually assesses the underlying variable, or construct, that the measure is intended to represent. It also determines the extent to which the measure appears to comply with the theoretical implications of the topic being measured. A form of construct validity in the social sciences is discriminant validity. This refers to the degree to which multiple measures of different concepts are distinct. Researchers have also relied on factor analysis for evaluating discriminant validity.

Criterion validity concerns how accurately an instrument predicts a well-accepted indicator of a given concept, or a criterion. A multivariate approach to assessing criterion validity is structural equation modelling (SEM) (Byrne 1998).

The validity of a measurement scale concerns how thoroughly (content validity) and accurately (construct validity) it measures a theoretical concept of interest. In this study, cognitive relevance judgement is a specific theoretical concept of interest. It also concerns how useful it is in predicting important outcomes (criterion validity) (Bryant 2000).

In this study factor analysis and structural equation modelling have been used to evaluate the scale validity. Factor analysis was used for content and construct validity.

3-9-2-5. Longitudinal Analysis and Repeated Measurement

As mentioned before, the experiments were replicated three times during 2001. Because the response of each unit was observed on multiple occasions and data collected over a relatively short period, under uncontrolled circumstances (Hand and Crowder 1996), this is a repeated measurement or longitudinal study. Therefore, statistical techniques for analysing longitudinal or repeated measure data were used. Relevance judgement was regressed against several potential variables. This study intended to determine to what extent relevance judgement related to various factors such as Aboutness, Quality of information, and so on. For this, SEM analysis carried out.

3-9-3. Structural Equation Modelling (SEM)

Structural Equation Modelling (SEM), is a widely used procedure. SEM[1] may be considered a generalization of regression analysis, factor analysis or other multivariate analysis methods. Byrne (1998) pointed out SEM is a statistical methodology that takes a confirmatory, rather than exploratory, approach to multivariate data analysis. This technique is particularly useful when the associations in the data are complicated and it is assumed that underlying factors indirectly determine the measurement. In SEM, the assumed underlying factors are represented as latent variables and the aim of SEM is to estimate the strength of the associations among the latent variables and between the observed variables and the latent variables.

SEM merges a variety of statistical procedures: Factor analysis and Confirmatory Factor Analysis; Multiple regression Analysis; Path Analysis and Measurement modelling. SEM estimates two models simultaneously, the measurement model and the structural model.

The measurement model specifies how the latent variables are measured. It is also called Confirmatory Factor Analysis. Latent variables are unobserved/unobservable variables

---

1 The term structural equation modelling conveys two important aspects of the procedure: (a) that the causal processes under study are represented by a series of structural (i.e. regression) equations, and (b) that these structural relations can be modelled pictorially to enable a clearer conceptualisation of the theory under study (Byrne 1998).

such as Aboutness, Quality of Information, Characteristics of Information and Information Novelty. These are sometimes called factors. They are measured by indicators or observed variables, for example the indicators of Aboutness can be About my topic, Informativeness and so on.

The Structural model specifies how the latent variables are related. It is similar to regression or path analysis. When the Measurement Model and the Structural Model are estimated simultaneously, we are performing a full SEM analysis. SEM has major advantages over other techniques when analysing longitudinal models with multiple indicators. A key advantage is that measurement errors may be correlated over time (Li et al. 1998).

3-9-4. Software Tools for Statistical Analysis

Data analysis has been affected by the development of statistical software packages that run on PCs. Data analysis in this study was carried out using SPSS version 10. SPSS is the most commonly employed statistical package in the social sciences. The reasons for choosing SPSS as a software tool for data analysis were: first, it was available in the Department of Information Science of Loughborough University where the research was carried out; second, SPSS offers a user-friendly and accessible set of procedures which enables the researcher to execute all the procedures that are discussed here.

3-9-4-1. LISREL

LISREL is an acronym for the LInear Structural RELations model. It can be considered as an extension of path analysis, which can, in turn be viewed as an extension of multiple regression. In multiple regression, there is one dependent variable and multiple independent variables. Path analysis involves multiple independent variables and multiple dependent variables. Like path analysis, LISREL can have multiple independent and dependent variables. In addition, it uses both observed and latent variables.

The variables that we deal with in regression analysis are called observed variables because they can be directly observed or measured and we can collect data on them. LISREL involves a new type of variable called the latent variable. In contrast to observed variables, latent variables can not be measured or observed directly, nor can we collect data on them. Instead, they are formulated in terms of theoretical or hypothetical concepts.

For example, we can not directly observe or measure relevance judgements in information retrieval, nor can we collect data on it. Therefore, relevance judgement is a latent variable. In contrast, we can collect data on criteria of relevance judgement that were used by the participants when selecting records from an IRS. LISREL is the most widely used approach and most methodological publications rely on it. LISREL (version 8.52) was written by Karl G. Jöreskog and Dag Sörbom and distributed by Scientific Software International. A free student version can be downloaded from their WebPage (http://www.ssicentral.com/lisrel/mainlis.htm). It is limited in the number of observed variables that can be used. The major alternatives to LISREL are AMOS and EQS[1]. In this study, the data has been analysed by LISREL (the license for full usage rights and priority support had been provided by the Department of Information Science).

1 AMOS is provided by Small Waters Software, AMOS also has a contract with SPSS as an SPSS add-on. EQS was developed by Peter Bentler, EQS is now available as a standalone program. (http://www.mvsoft.com).

Chapter 4: The Pilot Study and The Study of Experts' Attitude

4-1. The Pilot Study

4-1-1. Introduction

One of the main objectives of the study was to identify cognitive factors that impact upon relevance judgement. As already mentioned, the pilot study was a model of the full research study, but on a smaller scale (briefer time frame and on fewer participants). It focused on those aspects of the full-scale study that were novel, untested and complex. The rationale of the pilot study was:

- To identify cognitive factors of relevance judgement;

- To obtain data to help the researcher plan the full-scale study.

There were other reasons to run the pilot study. The pilot study helped the researcher get familiar with the procedures of the full-scale study. The pilot study could also help the researcher decide between two approaches (collecting data in an interview versus using a self-administered survey).

The pilot study permitted discovery of response and coverage rates necessary for general sample size planning and assessment of the utility of individual items and sets of items to be used in the full-scale experiment.

4-1-2. Procedure

In the first stage of the study a structured questionnaire was designed. To improve the validity of questions, and in order to establish the participant's understanding of what was meant by each criterion that affects their relevance judgement, a short interview

with participants was used by the researcher. Although, the questionnaire removed the need for respondents to contact the researcher for clarification, when the questionnaire was distributed the interview was a way of detecting any misunderstanding.

The pilot study was carried out with five individuals (3 males and 2 females) in June 2000. The participants were research students from Department of Information Science of Loughborough University. Each student invited to participate in the pilot study received a letter through email or an oral invitation. The respondents to the pilot study were chosen for pre-testing the potential effectiveness of the questionnaire because of their subject knowledge and their willingness to co-operate.

A letter (Appendix 4) was sent to the research co-ordinator of each by email. The names and email addresses of 16 postgraduate students were thereby obtained. A sample of five students from the list of 16 students was selected. Each student invited to participate in the pilot study received a letter through email or by an oral invitation (Appendix 1).

The process of conducting the experiment, survey and interview was as follows: 1) the respondent was selected and asked to come to a office in which the experiment would be conducted; 2) the respondent was introduced to the experiment and instructed on searching in ISI Web of Science; 3) the respondent was asked to fill out the questionnaire after his/her search session; 4) the interview conducted. Finally the researcher thanked the respondent and terminated the process.

4-1-3. Questionnaire design

Data were collected by means of a questionnaire containing 50 items. The first six questions of the questionnaire were related to participants' search characteristics. Questions 7 through 40 solicited the attitude of participants about the criteria that may contribute to relevance judgement. These questions had five answer choices: strongly agree/agree/unsure/ disagree/strongly disagree. Categories of criteria that measured by mentioned questions were topicality, novelty, quality of information, presentation of information, source characteristics, information characteristics and appeal of information. Question 41 was an open-ended question that asked the idea of

participants about other criteria. The final questions asked for some information about user characteristics and his/her research topic.

A follow-up interview was carried out to gain further insight into the participants' response to the questionnaire. The interview yielded more detailed responses since people tend to speak more easily than they write. The researcher asked some open-ended questions orally. The interview included questions about whether the questionnaire was meaningful, whether questions needed rephrasing, and whether questions were missing.

4-1-4. Discussion

The results of the pilot study showed that the questionnaire should be improved. This improvement involved reducing the number of factors and reviewing the scale of measurement.

In the pilot study, there were many occasions in which the respondents could not judge some criteria of relevance of an article because of insufficient information about the article content in the bibliographic record and abstract. Consequently, a number of questions were deleted. Also it proved possible to merge some questions and reduce the number of criteria. A number of criteria were deleted or merged (Table 4-1). In the end, 16 criteria remained in the questionnaire.

Table 4-1. Deleted or Merged Criteria

No.	Statements which indicated a Criterion	No.	Statements which indicated a Criterion
1	They were **well known** to me.	10	The **overviews** were useful.
2	They were **well written.**	11	The **journals/conferences** that they appeared were **prominent.**
3	They were **credible.**	12	The journals/conferences **sources** are **reputable.**
4	They were **accurate.**	13	They had a **research approach.**
5	They were **focused.**	14	They had a **statistical approach.**
6	Information in them was **well presented.**	15	They provided **proof.**
7	The articles were **comprehensive.**	16	I **like** them.
8	They were very **detailed.**	17	They were **interesting.**
9	They were **introductory.**	18	They were **enjoyable** to read.

4-2. Measurement of Relevance Criteria: Experts' Attitude

4-2-1. Introduction

The equal-appearing-interval scale or Thurstone scale was used for measuring the attitude of experts. The central idea of a Thurstone scale is that, for each statement on the scale, if a person agrees with it, they were given a score equivalent to the strength of that statement or item. The Thurstone scale calculates a 'weight' or 'attitudinal value' for each item. The weight of each item is calculated on the basis of ratings assigned by a group of judges. The classic Thurstone approach to attitude scale construction involves two main stages. In the first stage, a number of attitude statements are written to span

the entire range of possible opinions, and these items are scaled with regard to their unfavourability or favourability towards a given attitude object. There are several Thurstonian techniques for scaling attitude items, including pairwise comparisons, equal-appearing intervals and successive intervals (Thurstone and Chave 1929). All of these methods require a group of participants to make judgements about each item, and all methods yield a set of item scale values that indicate how favourably or unfavourably the item's sentiment reflects the attitude object. Those items with scale values having large standard errors are discarded from the pool of items under consideration. In the second stage, participants are asked to indicate attitude statements with which they agree. Attitude estimates are developed for each individual by computing the mean (or median) scale value associated with endorsed items, and then these attitude estimates are used to develop empirical operating characteristic curves for each item.

4-2-2. Scale Development Procedure

The procedure for constructing the Thurstone scale was as follows:

A set of statements, which reflected the criteria of cognitive relevance, were collected from first questionnaire.

Between December 2000 to January 2001, twenty experts in information retrieval were identified from the IR Mailbase[1] list and an invitation email along with an email survey (a number of statements which reflect the criteria of cognitive relevance) (See Appendix 3) was sent to them. Ten out of 20 experts replied. The experts were asked to act as a member of panel of judges. The experts' organisations were: Dublin City University (School of Computer Applications), University of Sheffield (Department of Information Studies), Delft University of Technology (Library), Manchester Metropolitan University, Loughborough University (Department of Information Science), University of Pennsylvania (School of Nursing), University of Udine (Italy), Rutgers University (School of Communication, Information & Library Studies) (Royal School of Library

1 Mailbase was a service which runs electronic discussion lists for the UK higher education and research community. It has since been replaced by JISCmail.

and Information Science, Department of Information Studies & Centre for Informetric Studies, Denmark).

The panel of judges were asked to rate each statement on a scale of 1 (highly negative on the issue) to 11 (highly positive on the issue). Rate 1 means the weight for a statement is low and rate 11 means the weight for a statement is high. The weight (equivalent to the median value) for each statement is calculated on the basis of rating assigned by a group of judges.

The median value and standard deviation for each statement across all the judges' ratings were calculated.

4-2-3. Analysis of Responses

Table 4-2 shows the rating given by experts:

Table 4-2. Raw Data of Experts' Rating

	Criteria	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Expert 7	Expert 8	Expert 9	Expert 10	*Median*	*SD*
1	About my topic	9	11	11	11	11	9	9	8	10	11	10.5	1.2
2	Generally or specifically suitable	8	11	10	10	10	9	11	10	11	7	10.0	1.3
3	The authors are eminent	7	9	9	9	4	7	9	9	9	6	9.0	1.8
4	Informative	7	11	9	10	8	8	11	5	9	8	8.5	1.8
5	New	7	11	6	10	10	4	9	11	8	8	8.5	2.3
6	Original	4	10	8	11	8	6	8	10	10	8	8.0	2.1
7	Know the author(s) personally	3	6	9	9	9	7	8	8	7	9	8.0	1.9
8	Unique or only sources	4	8	7	11	1	9	8	5	9	9	8.0	3.0
9	Know the journals or conferences	5	7	5	8	9	7	10	8	6	5	7.0	1.8
10	Good bibliography and links	5	8	7	8	7	5	9	6	7	5	7.0	1.4
11	Validate viewpoint	3	6	8	9	4	7	7	6	9	4	6.5	2.1
12	Easy to understand	6	9	4	10	6	6	9	4	5	6	6.0	2.1
13	Background or history	6	7	6	6	7	5	8	5	9	5	6.0	1.3
14	Methods or techniques have described	6	5	7	9	5	5	6	5	8	5	5.5	1.4
15	Consistent	3	5	8	10	4	2	7	5	7	6	5.5	2.4
16	Controversial	3	6	8	9	2	2	8	5	4	4	4.5	2.6

After all of the items were judged, the scale values assigned by the experts for each item were tabulated and the median value used as the actual scale value of the item. Standard deviations were also calculated.

4-2-3-1. Scale Reliability and Validity

Internal consistency reliability was used to consider the reliability of the scale. In scale development procedure, the main concern was correlations among criteria. To investigate this, correlation coefficients were calculated. Table 4-3 shows correlation coefficients of each criterion.

Table 4-3. Correlation Coefficients of Each Criterion

	Criteria	Item Total Correlation Coefficient
1	About my topic	0.3012
2	Generally or specifically suitable	0.7572
3	The authors are eminent	0.6662
4	Informative	0.3404
5	New	0.7469
6	Original	0.5759
7	Know the author(s) personally	0.4954
8	Unique or only sources	0.2849
9	Know the journals or conferences	0.4738
10	Good bibliography and links	0.8438
11	Validate viewpoint	0.7030
12	Easy to understand	0.5366
13	Background or history	0.3984
14	Methods or techniques have described	0.5871
15	Consistent	0.8416
16	Controversial	0.7851

An examination of the reliability of the criteria revealed a Coefficient Alpha of 0.8977, suggesting that the 16 criteria are internally consistent.

Validity of the scale was examined by factor analysis. The matrix of correlation coefficients is tabulated in Table 4-4. Factor analysis uses the correlation matrix to try to determine which sets of variables cluster together.

Table 4-4. Correlation Matrix

	About my topic	Suitable	Informative	New	Original	The authors are eminent	Know the author(s)	Know the journals or	Unique or only sources	Good bibliography and	Validate viewpoint	Easy to understand	Background	Methods or techniques	Consistent	Controversial
About my topic	1															
Generally or specifically suitable	0.194	1														
Informative	0.524	0.584	1													
New	0.127	0.512	0.096	1												
Original	0.319	0.757	0.292	0.668	1											
The authors are eminent	-0.220	0.498	0.318	0.134	0.469	1										
Know the author(s)	0.257	0.549	0.187	0.214	0.601	0.115	1									
Know the journals or conferences	-0.119	0.428	0.217	0.520	0.338	-0.141	0.498	1								
Unique or only sources	0.162	0.230	0.498	-0.188	0.386	0.583	0.076	-0.074	1							
Good bibliography	0.271	0.843	0.759	0.524	0.590	0.465	0.527	0.513	0.193	1						
Validate viewpoint	0.197	0.542	0.432	-0.100	0.632	0.648	0.545	0.000	0.795	0.420	1					
Easy to understand	0.222	0.329	0.741	0.370	0.262	0.228	0.060	0.481	0.373	0.657	0.157	1				
Background or history	0.064	0.695	0.534	0.223	0.250	0.280	0.135	0.156	-0.033	0.676	0.247	0.301	1			
Methods or techniques	0.266	0.239	0.350	-0.014	0.389	0.490	0.278	-0.199	0.489	0.395	0.685	0.268	0.395	1		
Consistent	0.399	0.537	0.486	0.316	0.681	0.584	0.560	0.108	0.488	0.675	0.729	0.331	0.314	0.807	1	
Controversial	0.226	0.485	0.553	0.298	0.529	0.724	0.436	0.156	0.497	0.714	0.588	0.472	0.166	0.566	0.871	1

The factors and their associated Eigenvalues[1], percentage of variance explained and the cumulative percentages have been calculated. Table 4-5 displays the total variance explained.

Table 4-5. Total Variance Explained

	Initial Eigenvalues		
Component	Total	% of Variance	Cumulative %
1	7.085	44.281	44.281
2	2.551	15.945	60.226
3	1.580	9.873	70.099
4	1.342	8.385	78.485
5	1.144	7.153	85.637
6	0.932	5.826	91.463
7	0.720	4.498	95.961
8	0.513	3.207	99.168
9	0.133	0.832	100.000
10	0.000	0.000	100.000
11	0.000	0.000	100.000
12	0.000	0.000	100.000
13	0.000	0.000	100.000
14	0.000	0.000	100.000
15	0.000	0.000	100.000
16	0.000	0.000	100.000

1 This value shows how many factors are potentially statistically significant – the minimum value of a potentially significant eigenvalue is 1.0.

In reference to the eigenvalues, it would be expected that five factors would be extracted because they have eigenvalues greater than 1. If five factors are extracted then 85.6 % of the variance would be explained.

By producing a component matrix of components, five factors have been extracted. Table 4-6 shows the difference between high and low loadings in the component matrix.

Table 4-6. Component matrix

Component	1	2	3	4	5
Validate viewpoint	0.983				
Unique or only sources	0.939		0.358	-0.390	
The authors are eminent	0.906				-0.573
Consistent	0.767				
Methods or techniques	0.759	-0.331			
Controversial	0.754				
Know the journals or conferences	-0.320	0.937	0.325		
New		0.861			
Know the author(s) personally		0.650	-0.310		
Original	0.496	0.648			
Easy to understand			0.901		
Informative			0.687	0.306	0.337
Background or history				1.095	
Generally or specifically suitable		0.458		0.547	
Good bibliography and links		0.417	0.332	0.452	
About my topic					0.980

Table 4-7 shows the factor correlation matrix that indicates the relationship between factors.

Table 4-7. Factor Correlation Matrix

Component	1	2	3	4	5
1	1				
2	0.370	1			
3	0.237	0.171	1		
4	0.456	0.453	0.291	1	
5	0.307	0.262	0.026	0.252	1

4-2-4. Discussion

As Table 4-2 shows, the first criterion has the highest rank (10.5). It means that the experts believed that "About my topic" is an important criterion that effects cognitive relevance. "Generally or specifically suitable" with scale value of 10 out of 11 and "The authors are eminent" with scale value of 9.0 out of 11 are in the second and third rank. The experts also believed that "controversial" with scale value of 4.5 out of 11 is the criterion with the lowest rank. The findings of experts' attitudes toward rank order of cognitive relevance criteria are shown in Table 4-8.

Table 4-8. Rank order of cognitive criteria of relevance based on experts' attitude

Rank Order	Experts' Attitude
1	About my topic
2	Generally or specifically suitable
3	The authors are eminent
4	Informative
5	New
6	Original
7	Know the author(s) personally
8	Unique or only sources
9	Know the journals or conferences
10	Good bibliography and links
11	Validate viewpoint
12	Easy to understand
13	Background or history
14	Methods or techniques have described
15	Consistent
16	Controversial

Based on the Thurstone scale if the experts' ratings of an item scattered over the scale, this indicated that among these experts there is little agreement as to the degree to which that statement reflects their opinions. In this study the maximum standard deviation was 3.0 and there were no statements for discard. For the next stage of this research all sixteen statements were used.

Chapter 5: Results - Quantitative Analysis

5-1. Introduction

This chapter provides the results of the studies by presenting the findings of descriptive and inferential analysis of data. It describes the characteristics of participants of the study. Then, characteristics of respondents' search and their responses are presented. While this chapter focuses in particular on the describing the survey participants, it also traces the reliability and validity of the scales in the three studies.

5-2. Data Analysis of Study 1

5-2-1. Characteristics of Participants

The sample consisted of 30 Ph.D. students (19 in Loughborough University and 11 from the University of Nottingham); (7 female and 23 male). Twenty-one out of the thirty participants were studying science and nine engineering. The average age of the participants was 32 years (Standard Deviation 5.4). Nine students were in the first year of their study, 11 of them were in second year and 10 were in third year or above. Table 5-1 shows the characteristics of respondents.

Table 5-1. Characteristics of Respondents

Subjects	Gender	Age (years)	Department	Year Of the Study	Research Topic
Student 01	Male	42	Information Science	1	School library
Student 02	Male	24	Mathematical sciences	1	Fact Coulomb sums and their applications in ionic materials an oxides
Student 03	Male	32	Mathematical sciences	1	Non linear waves
Student 04	Male	32	Manufacturing engineering	1	Shearography
Student 05	Male	30	Mathematical sciences	1	Geographical Fluid Dynamics
Student 06	Male	24	Mechanical engineering	1	Mechatronics
Student 07	Female	35	Human Science	3	Grip Forces
Student 08	Male	24	Electronic and Electrical Engineering	3	Effects of Solar radiation spectral content on the performance of thin film photovoltaic modules
Student 09	Male	39	Information Science	2	Internet and Distance Learning
Student 10	Male	26	Mechanical engineering	1	Condition monitoring of bearings by ultrasound
Student 11	Male	35	Information Science	2	Impact assessment of Rural Information project: a needs analysis
Student 12	Male	25	Computer Science	3	Component-based software reuse
Student 13	Male	31	Electronic and Electrical Engineering	2	Speech Recognition
Student 14	Female	40	Human Science	3	Social Remembering
Student 15	Male	33	Food Science	3	Effect of Maillard reaction on virulence gene expression of listeria
Student 16	Male	31	Information Science	2	A Human Factors Experiment in Web-based User Interfaces
Student 17	Male	40	Manufacturing Engineering	3	Body Responses to combined effects of physical and psychosocial risk factors
Student 18	Male	34	Civil Engineering	3	Linear and non-linear viscoelastic behaviour of bitumminous binders and asphalt mixtures
Student 19	Male	24	Information Science	2	Information retrieval interaction,
Student 20	Male	31	Chemistry	1	Design and synthesis of UDP-Galp mutase inhibitors (anti-tuberculosis agents)

Student 21	Male	31	Genetics	2	Human DNA copy number analysis
Student 22	Female	32	Food Science	3	Virlence gene expression of L. monocytogenes in soft cheese
Student 23	Male	35	Civil Engineering	2	Riverbank Erosion Modelling
Student 24	Female	35	Genetics	3	Interplay between DNA replication and repair
Student 25	Male	29	Food Science	2	The rheological impacts of chemical agents on food proteins
Student 26	Female	25	Food Science	1	Wheat starch, molecular structure in relation to processing
Student 27	Male	35	Sport Science	3	The effects of functional knee bracing or taping on the tibiofemoral joint in athletes with an ACL-deficient knees
Student 28	Female	40	Management	2	Heuristic Rules of Strategic Learning Through Simulation
Student 29	Female	28	Physics	2	The Characteristic of Polymer under Different Strain Rates
Student 30	Male	36	Mining	3	Mining and Tunnel Excavation

Figures 5-1 to 5-5 illustrate the content of the Table 5-1 in graphic format.

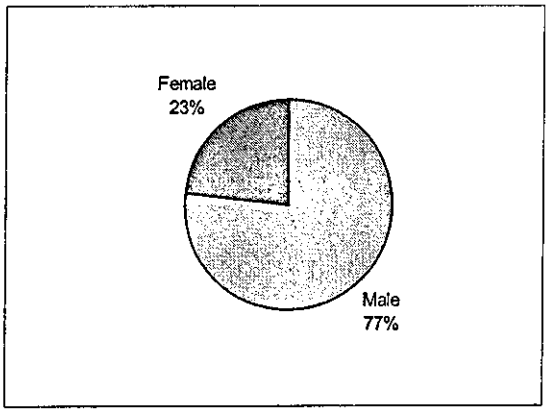


Figure 5-1. Percentage of Males and Females

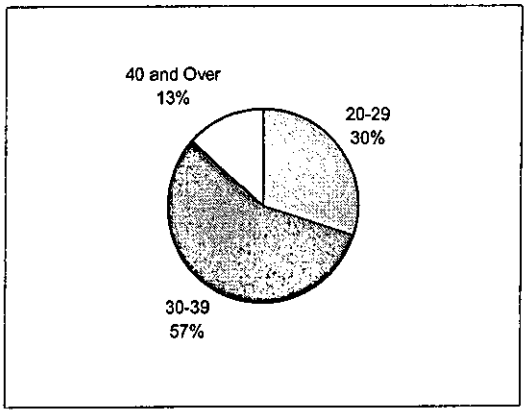


Figure 5-2. Percentage of Ages

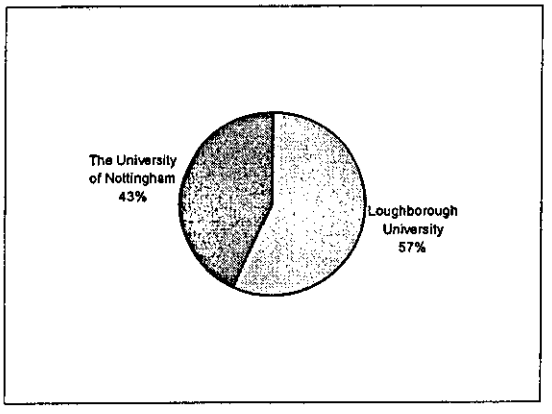


Figure 5-3. Percentage of Universities

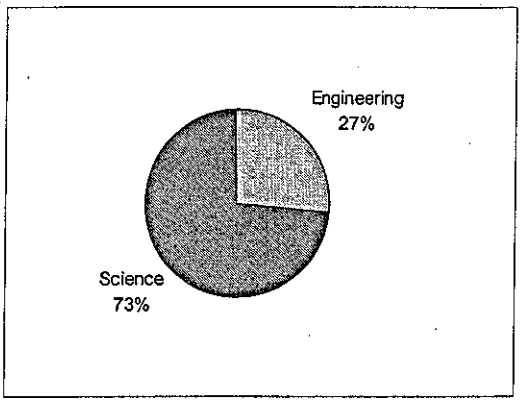


Figure 5-4. Percentage of Areas of Studies

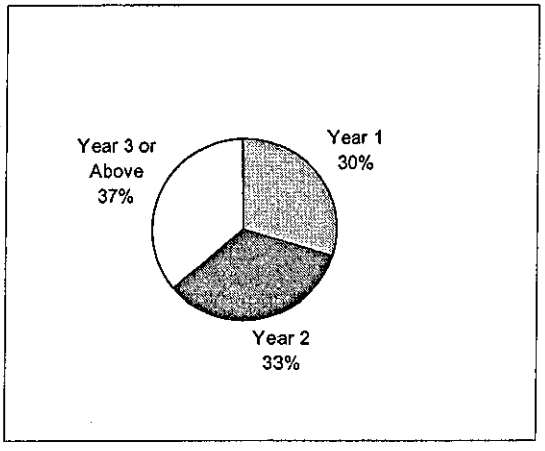


Figure 5-5. Percentage of Year of the Study

5-2-2. Characteristics of Respondents' Search

Table 5-2 shows the search characteristics and results from the study 1.

Table 5-2. Students' Search Characteristics (Study 1)

	No. of Search	No. of records	Completely Relevant	Partially Relevant	Not Relevant	Time Spent (Minutes)
Student 01	2	80	60	15	5	120
Student 02	1	44	11	4	29	20
Student 03	2	23	4	7	12	40
Student 04	3	5	3	2	0	15
Student 05	2	26	6	5	15	55
Student 06	2	34	0	1	33	20
Student 07	3	200	5	45	150	45
Student 08	6	24	6	14	4	30
Student 09	2	99	15	32	52	30
Student 10	10	56	22	27	29	130
Student 11	4	11	3	4	4	15
Student 12	2	22	11	3	8	20
Student 13	1	12	10	2	0	1
Student 14	5	20	12	8	0	40
Student 15	4	3	3	0	0	4
Student 16	4	17	7	9	1	10
Student 17	7	98	22	40	36	180
Student 18	3	108	22	32	54	10
Student 19	3	23	4	6	13	10
Student 20	3	18	12	6	0	5
Student 21	2	47	6	10	31	30
Student 22	5	10	4	3	3	10
Student 23	1	71	25	17	29	180
Student 24	1	116	8	59	49	30
Student 25	7	30	2	10	18	30
Student 26	5	3	0	1	2	3
Student 27	1	45	8	15	8	45
Student 28	1	37	8	22	7	15
Student 29	3	30	0	2	0	30
Student 30	11	260	3	9	250	120
Mean	**3.5**	**52.4**	**10.1**	**14.7**	**28.1**	**43.1**
SD	**2.57**	**58.44**	**11.68**	**14.78**	**51.32**	**50.09**

Table 5-2 shows that the average number of search strategies was 3.5. The number of records reviewed by respondents ranged from a low of 3 to a high of 260. They performed 106 search sessions in total and retrieved 1572 records. Participants reported that 19% of search results were completely relevant, 26% partially relevant and 55% non-relevant. Figure 6-6 reports this in graphic display.

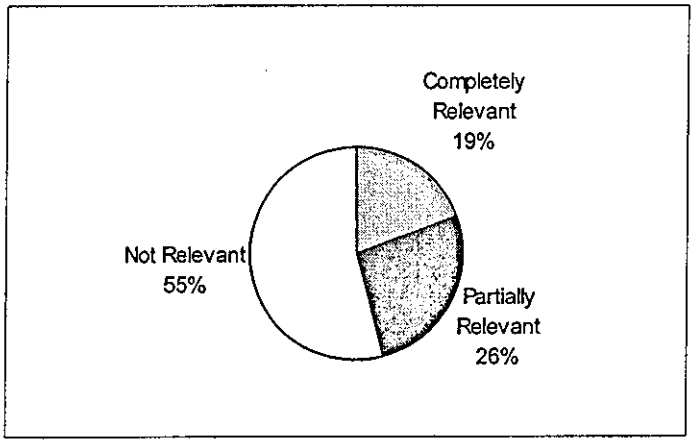


Figure 5-6. Percentage of Completely, Partial and Non-relevant Records (Study 1)

The average number of records retrieved was more than 52, of which on average 10 articles were deemed completely relevant, 13 articles were deemed partially relevant and 28 articles were deemed not relevant. The average time spent for each search was 43 minutes and 6 seconds.

5-2-3. Responses

Section 2 of the questionnaire was designed to collect data about respondents' attitudes towards the cognitive relevance criteria. A Likert-type attitude scale was constructed, in which the respondent was presented with statements and was asked to tick one of a row of boxes to indicate the degree to which he or she agreed or disagreed with each statement (Likert 1932). The advantages of this technique have been identified by Coolican (1994) as:

- Subjects prefer the Likert scaling technique because it is "more natural" to complete and because it maintains the subject's direct involvement.

- The Likert technique has been shown to have a high degree of validity and reliability.

- The Likert scale has been shown to be effective at measuring changes over time. The methodology of Likert scaling was as follows:

The unit values was assigned to each ordered category (From integer 1 for strongly disagree through 5 for strongly agree). After subjects responded by checking or marking one of the categories for each criterion items an N x K (subject by item) matrix of data was generated as shown in Table 5-3. Each subject's categorical value is provided in the body of the table.

Item analyses were then performed on the data. The median and standard deviation of each item are calculated and the Pearson *r* correlation of each item with the total score on all items is found. This correlation acts as a discrimination index for each criterion. If the criterion correlates highly with the total score, it is internally consistent and should be retained. In Table 5-3, criterion 6 (Easy to Understand) has a correlation coefficient 0.01 and criterion 13 (Controversial) has a correlation coefficient 0.06. There are low and can be eliminated from the set of criteria.

Table 5-3. Raw Data Matrix (Study 1)

	Relevance Criteria																
Subjects	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	*Total Score*
01	4	4	4	4	4	4	3	5	2	3	3	4	2	4	4	5	59
02	4	3	4	4	5	4	4	4	2	2	4	4	1	3	4	3	55
03	4	3	4	4	4	5	4	3	1	4	4	3	1	5	4	1	54
04	4	4	4	5	4	2	3	4	5	4	4	5	4	4	4	4	64
05	4	3	4	2	4	4	4	4	1	4	4	5	1	4	5	4	57
06	2	5	4	4	4	2	3	4	2	2	2	4	3	4	4	3	52
07	4	3	3	4	4	2	4	4	1	1	2	3	2	1	1	3	42
08	3	2	2	4	3	3	4	4	1	2	4	5	1	5	4	3	50
09	4	2	3	3	5	5	5	3	2	4	3	4	2	1	1	2	49
10	4	2	5	4	5	4	4	5	1	3	4	5	2	4	4	4	60
11	4	2	4	2	4	4	3	4	3	4	4	3	2	2	4	4	53
12	4	4	5	4	4	4	2	4	1	2	2	3	2	3	3	3	50
13	4	5	5	4	4	3	4	5	2	5	4	4	2	4	2	3	60
14	4	5	4	4	5	3	2	4	1	3	4	4	4	5	4	2	58
15	5	5	1	5	5	4	4	5	1	4	4	4	2	4	4	5	62
16	5	4	4	4	4	4	4	5	2	5	3	4	3	4	4	4	63
17	4	3	2	3	5	4	1	5	2	3	4	4	2	2	3	1	48
18	4	4	3	2	3	4	2	3	2	4	3	4	4	4	3	3	52
19	2	4	2	2	2	4	1	2	2	4	3	5	2	4	2	4	45
20	5	5	5	5	4	1	3	4	3	2	2	3	2	4	5	4	57
21	4	3	3	4	3	4	2	3	1	4	3	3	3	1	3	3	47
22	4	4	4	4	2	4	4	3	4	4	4	4	1	5	4	3	58
23	4	4	4	4	4	3	3	5	4	5	5	4	3	5	5	4	66
24	5	4	4	4	5	5	5	5	4	4	4	5	2	4	5	4	69
25	2	4	2	4	2	4	3	4	2	2	2	4	3	1	3	5	47
26	3	3	4	3	3	4	3	3	1	5	3	4	3	4	4	3	53
27	4	3	4	2	5	4	4	5	1	5	4	4	3	4	5	3	60
28	5	4	3	4	4	2	2	5	4	2	4	2	4	4	2	3	54
29	4	2	3	4	3	4	3	4	2	2	3	2	3	2	2	2	45
30	4	4	5	5	4	4	4	4	1	4	4	3	2	3	3	3	57
Correlation	*0.51*	*0.38*	*0.44*	*0.28*	*0.44*	*0.01*	*0.41*	*0.55*	*0.41*	*0.53*	*0.57*	*0.38*	*0.06*	*0.63*	*0.68*	*0.38*	
Median	**4.0**	**4.0**	**4.0**	**4.0**	**4.0**	**4.0**	**3.0**	**4.0**	**2.0**	**4.0**	**4.0**	**4.0**	**2.0**	**4.0**	**4.0**	**3.0**	
SD	**0.8**	**1.0**	**1.0**	**0.9**	**0.9**	**1.0**	**1.0**	**0.8**	**1.2**	**1.2**	**0.8**	**0.8**	**0.9**	**1.3**	**1.1**	**1.0**	

Where 1 = Strongly Disagree, 2 = Disagree, 3 = Unsure, 4 = Agree, 5 = Strongly Agree

C1 = About the topic, C2 = Unique or only sources, C3 = Original, C4 = New, C5 = Informative, C6 = Easy to Understand, C7 = Consistent, C8 = Generally or specifically suitable C9= Know the author(s) personally, C10 = Know the journals or conferences, C11 = The authors are eminent, C12 = Methods or techniques have described, C13 = Controversial, C14 = Good bibliography and links, C15 = Background or history, C16 = Validate viewpoint.

The results of the study 1 have been summarised in Table 5-4 and presented in percentages in Table 5-5.

Table 5-4. Frequencies of Responses (Study 1)

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16
Strongly Agree	5	5	5	4	8	3	2	10	1	5	1	6	0	5	5	3
Agree	20	12	14	18	14	18	12	13	4	12	16	15	4	15	13	9
Unsure	2	8	6	3	5	4	9	6	2	4	8	7	8	3	6	13
Disagree	3	5	4	5	3	4	5	1	11	8	5	2	13	3	4	3
Strongly Disagree	0	0	1	0	0	1	2	0	12	1	0	0	5	4	2	2

C1 = About the topic, C2 = Unique or only sources, C3 = Original, C4 = New, C5 = Informative, C6 = Easy to Understand, C7 = Consistent, C8 = Generally or specifically suitable C9= Know the author(s) personally, C10 = Know the journals or conferences, C11 = The authors are eminent, C12 = Methods or techniques have described, C13 = Controversial, C14 = Good bibliography and links, C15 = Background or history, C16 = Validate viewpoint.

Table 5-5. Percentages of Responses (Study 1)

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16
Strongly Agree	16.7	16.7	16.7	13.3	26.7	10.0	6.7	33.3	3.3	16.7	3.3	20.0	0.0	16.7	16.7	10.0
Agree	66.7	40.0	46.7	60.0	46.7	60.0	40.0	43.3	13.3	40.0	53.3	50.0	13.3	50.0	43.3	30.0
Unsure	6.7	26.7	20.0	10.0	16.7	13.3	30.0	20.0	6.7	13.3	26.7	23.3	26.7	10.0	20.0	43.3
Disagree	10.0	16.7	13.3	16.7	10.0	13.3	16.7	3.3	36.7	26.7	16.7	6.7	43.3	10.0	13.3	10.0
Strongly Disagree	0.0	0.0	3.3	0.0	0.0	3.3	6.7	0.0	40.0	3.3	0.0	0.0	16.7	13.3	6.7	6.7

C1 = About the topic, C2 = Unique or only sources, C3 = Original, C4 = New, C5 = Informative, C6 = Easy to Understand, C7 = Consistent, C8 = Generally or specifically suitable C9= Know the author(s) personally, C10 = Know the journals or conferences, C11 = The authors are eminent, C12 = Methods or techniques have described, C13 = Controversial, C14 = Good bibliography and links, C15 = Background or history, C16 = Validate viewpoint.

A summary of the attitudes of respondents toward cognitive relevance criteria based upon the results of study 1 are presented in Table 5-6.

Table 5-6. Cumulative Percentages of Responses (Study 1)

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16
Total Agree	83.3	56.7	63.3	73.3	73.3	70.0	46.7	76.7	16.7	56.7	56.7	70.0	13.3	66.7	60.0	40.0
Unsure	6.7	26.7	20.0	10.0	16.7	13.3	30.0	20.0	6.7	13.3	26.7	23.3	26.7	10.0	20.0	43.3
Total Disagree	10.0	16.7	16.7	16.7	10.0	16.7	23.3	3.3	76.7	30.0	16.7	6.7	60.0	23.3	20.0	16.7

Total Agree = strongly agree + agree, Total Disagree = strongly disagree + disagree

C1 = About the topic, C2 = Unique or only sources, C3 = Original, C4 = New, C5 = Informative, C6 = Easy to Understand, C7 = Consistent, C8 = Generally or specifically suitable C9= Know the author(s) personally, C10 = Know the journals or conferences, C11 = The authors are eminent, C12 = Methods or techniques have described, C13 = Controversial, C14 = Good bibliography and links, C15 = Background or history, C16 = Validate viewpoint.

5-2-4. Scale Reliability

The reliability of scale considered in two ways:

Internal consistency reliability. This means that multiple items, designed to measure the same construct, will inter-correlate with one another.

Test-retest reliability. This means that a scale yields consistent measurements over time. Each respondent should get about the same score upon repeated testing. In this study, the experiment was replicated by making the measurement three times in approximately 3 months intervals in March 2001, July 2001 and November 2001. The details of the other experiments are discussed in next sections.

The first stage of investigation of the criteria considered the correlation between the 16 criteria and examined the relationships between the individual criteria. Table 5-7 shows the matrix of the sixteen criterion items.

Table 5-7. Correlation Matrix of Cognitive Relevance Criteria (Study 1)

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16
C1	1															
C2	0.0751	1														
C3	0.2401	0.1300	1													
C4	0.2862	0.3917	0.1598	1												
C5	0.5444	-0.0115	0.245	0.1265	1											
C6	-0.0089	-0.4105	-0.1304	-0.3734	0.0309	1										
C7	0.2766	-0.2037	0.2493	0.2209	0.3126	0.2328	1									
C8	0.226	0.2276	0.0975	0.1722	-0.0935	-0.2949	-0.0353	1								
C9	0.4773	0.1658	0.1526	0.3004	0.5959	-0.1807	0.2217	0.1414	1							
C10	0.331	-0.1462	0.0488	-0.0507	0.3338	0.2266	0.2015	0.239	0.3127	1						
C11	0.1551	0.0672	0.1658	-0.3046	0.0386	0.4226	0.1768	0.0921	-0.0287	0.5009	1					
C12	0.0046	0.2971	-0.0573	0.0122	-0.0362	-0.3300	-0.4491	0.2447	0.1018	-0.1258	0.0831	1				
C13	-0.2832	0.0355	-0.0797	-0.2033	0.0672	0.1708	0.2054	0.0416	0.0167	0.2614	0.3201	-0.1857	1			
C14	0.1141	0.3903	0.2754	0.0942	0.0409	-0.1503	0.0708	0.1752	0.1324	0.4596	0.3106	-0.0329	0.3663	1		
C15	0.1699	0.1717	0.3216	0.0497	0.2136	0.0000	0.1895	0.1439	0.2931	0.3525	0.2869	-0.1144	0.3818	0.5926	1	
C16	-0.0085	0.2961	-0.0262	0.1262	-0.1546	-0.1333	0.135	0.2267	0.2655	-0.1026	0.0818	0.0391	0.3396	0.0868	0.3287	1

The matrix shows how each criterion item is correlated to the other items. The item total correlation and the reliability coefficient (Cronbach's Coefficient Alpha) was examined. An examination of the reliability of the criterion items as a unidimensional construct revealed a Coefficient Alpha of 0.68, suggesting that the 16 criteria are reasonably internally consistent. However criterion 'Easy to Understand' had item total correlation, -0.13 and criterion item 'Controversial' had item total correlation, -0.07. This suggests that these criteria may not be related to the construct (See Table 5-8).

Table 5-8. Reliability Analysis (Study 1)

Item-total Statistics	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Alpha if Item Deleted
About my topic	50.967	39.758	0.4079	0.6351	0.6593
Unique or only sources	51.300	40.493	0.2505	0.6487	0.6747
Original	51.267	39.582	0.297	0.3868	0.6691
New	51.167	41.937	0.1483	0.5216	0.6856
Informative	50.967	39.964	0.3186	0.7266	0.667
Easy to understand	51.267	45.375	-0.1364	0.5322	0.7172
Consistent	51.633	39.895	0.2707	0.5433	0.6724
Generally or specifically suitable	52.833	39.592	0.2466	0.4232	0.6765
Know the author(s) personally	50.800	39.131	0.4555	0.6059	0.6538
Know the journals or conferences	51.433	39.013	0.4754	0.7267	0.6521
The authors are eminent	51.467	37.706	0.3835	0.7023	0.657
Controversial	52.500	44.603	-0.0751	0.4823	0.7093
Methods or techniques have described	51.033	40.999	0.2659	0.6252	0.6731
Good bibliography and links	51.400	35.490	0.4859	0.7123	0.6397
Background or history	51.367	35.482	0.5727	0.5482	0.6292
Validate viewpoint	51.600	40.386	0.2417	0.562	0.6759
Reliability Coefficients Alpha = 0.6849		**Standardised item Alpha =0.6832**			

5-2-5. Scale Validity

As explained in chapter 4, factor analysis was applied to the experimental data. This statistical approach was used to analyse interrelationships among a number of variables and to explain these variables in terms of their common underlying dimensions (factors). This statistical approach involving finding a way of condensing the information contained in a number of variables into a smaller set of dimensions (factors) with a minimum loss of information (Hair et al. 1995).

According to Hair et al. (1995), there are four basic factor analysis steps:

- Data collection and generation of the correlation matrix
- Extraction of initial factor solutions
- Rotation and interpretation
- Construction of scales or factor scores to use in further analyses

Factor analysis derives its factors by analysing the pattern of correlation among criteria. With this method, two questions were addressed: (a) the number of factors that best present the criteria items; and (b) the interpretation of factors.

There are two basic factor analytic methods to extraction of initial factor solution. They are known as Principal Factor and Principal Component Analysis (Hair et. al 1995). The principal factor and principal component analysis models are both widely utilised. The selection of one method over the other is based on two criteria: (1) the objective of the researcher conducting the factor analysis and (2) the amount of prior knowledge about the variance in the variables. According to Hair et al. (1995), principal factor analysis has several problems. First, principal factor analysis suffers from "factor indeterminacy", which means that for any individual respondent, several different factor scores can be calculated for the factor model results. There is no single unique solution as found in component analysis, but in the most instances, the differences are not substantial. The second issue involves the calculation of the estimated communalities used to represent the shared variance. For larger-sized problems, the computations can take substantial computer time and resources. Also, the communalities are not always estimable or may be invalid (e.g., values greater than one or less than zero), requiring the deletion of the

variable from the analysis. The complications of principal factor analysis have contributed to the widespread use of principal component analysis.

There is much debates in the literature as to which technique of factor analysis is the most appropriate. In this study, Principal Component Analysis (PCA) was used. The first step in any PCA should be to determine the factorability of the correlation matrix as a whole. The correlation matrix was studied and showed a wide range of correlations. If all the items had very low correlations with each other, then the data would be heterogeneous and not appropriate for the analysis. The correlation matrix revealed a mixture of high, medium and low correlations, which is appropriate for Principal Component Analysis.

The Bartlett's test of sphericity[1] was significant (0.005). This test was used to determine for the factorability of the correlation matrix as a whole. The factors and their associated eigenvalues[2], percentage of variance explained and the cumulative percentages have been calculated. Table 5-9 displays the total variance explained.

1 Bartlett's test of sphericity calculates the determinant of the correlation matrix then the determinant is converted to a chi-square statistic and tested for significance. A p-value less than .05 indicates that the correlation matrix is not an identity matrix and it makes sense to factor.

2 This value shows that how many factors are potentially statistically significant – the minimum value of a potentially significant eigenvalue is 1.0.

Table 5-9. Total Variance Explained (Study 1)

	Initial Eigenvalues		
Component	Total	% of Variance	Cumulative %
1	3.403	21.269	21.269
2	2.601	16.256	37.525
3	2.110	13.190	50.714
4	1.477	9.229	59.943
5	1.169	7.307	67.251
6	0.995	6.219	73.469
7	0.907	5.667	79.137
8	0.766	4.787	83.923
9	0.605	3.782	87.705
10	0.454	2.836	90.542
11	0.385	2.403	92.945
12	0.372	2.326	95.271
13	0.270	1.687	96.958
14	0.240	1.500	98.458
15	0.163	1.017	99.475
16	0.084	0.525	100.000

Five factors were extracted because they have eigenvalues greater than 1. The five factors extracted account for 67.2 % of the variance.

Table 5-9 shows how output helps to determine the number of components/factors to be retained for further analysis. One good rule of thumb for determining the number of factors, is the "eigenvalue greater than 1" criteria. This criterion allowed the researcher to be fairly sure that extracted factors will account for at least the variance of one of the variables used in the analysis. There are other criteria for selecting the number of factors to keep, but this is the easiest to apply, since it is the default of most statistical computer programs (Hair et. al 1995).

Table 5-10 presents the result of examining the rotated component matrix. Five factors have been extracted by creating a component matrix. Table 5-10 also shows the difference between high and low loadings in the component matrix.

Table 5-10. Component Matrix (Study 1)

Component	1	2	3	4	5
Generally or specifically suitable	0.843				
About my topic	0.837				
Informative	0.825				
Know the journals or conferences		0.756	0.320		
New		-0.704			
The authors are eminent	0.443	0.640			
Easy to understand		0.620		-0.359	
Good bibliography and links			0.881		
Original			0.800		-0.398
Background or history			0.604		0.302
Unique or only sources		-0.351	0.426	0.397	
Controversial				0.874	
Consistent	0.346			-0.692	
Know the author(s) personally				0.521	
Validate viewpoint					0.908
Methods or techniques have described		0.329			0.676

The factor correlation matrix indicates the relationship between factors (See Table 5-11).

Table 5-11. Factor Correlation Matrix (Study 1)

Component	1	2	3	4	5
1	1.000				
2	-0.011	1.000			
3	0.380	0.042	1.000		
4	-0.060	-0.252	0.134	1.000	
5	0.130	0.085	0.375	0.087	1.000

The final step in factor analysis involves determining how many factors can be considered. This is discussed in the chapter 7 of the thesis. The number of factors to be interpreted largely depends on the underlying purpose of the analysis. In the present chapter, the purpose was to confirm the validity of the scale.

5-3. Data Analysis of Study 2

5-3-1. Characteristics of Respondents' Search

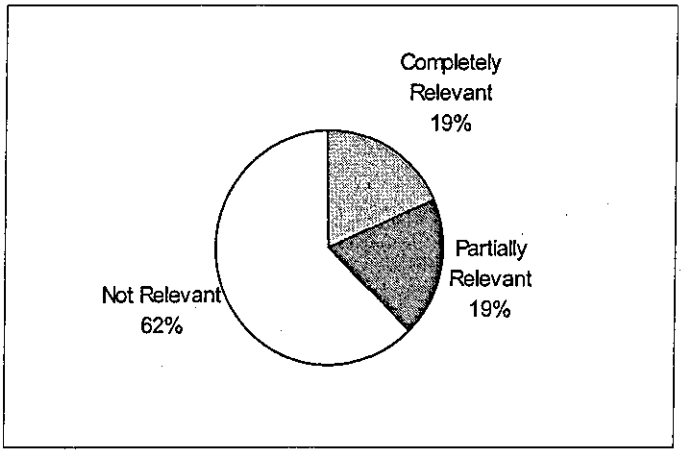
The second study was carried out in July 2001 with the same participants and same cognitive relevance criteria. Demographic data did not change from the 1st experiment to the 2nd experiment. The results of study 2 have tabulated in Table 5-12.

Table 5-12. Students' Search Characteristics (Study 2)

	No. of Search strategies	No. of records obtained	Completely Relevant	Partially Relevant	Not Relevant	Time Spent (Minutes)
Student 01	3	60	5	10	45	120
Student 02	1	17	4	6	7	30
Student 03	1	59	13	26	20	15
Student 04	6	1	1	0	0	15
Student 05	2	7	6	1	0	27
Student 06	3	10	4	2	4	15
Student 07	4	63	13	30	20	75
Student 08	3	32	17	9	6	40
Student 09	3	46	10	7	29	40
Student 10	10	56	47	8	1	180
Student 11	3	14	13	4	2	30
Student 12	4	55	20	35	0	15
Student 13	2	13	8	1	4	3
Student 14	5	22	14	5	8	40
Student 15	1	3	3	0	0	3
Student 16	3	29	17	6	6	11
Student 17	3	112	26	48	38	120
Student 18	5	187	35	54	99	20
Student 19	3	34	9	7	18	15
Student 20	2	8	3	1	4	16
Student 21	3	11	5	3	3	20
Student 22	1	6	2	2	2	10
Student 23	1	10	9	1	0	120
Student 24	1	10	10	0	0	10
Student 25	10	20	0	2	18	30
Student 26	6	3	0	1	1	10
Student 27	3	36	9	10	17	20
Student 28	1	37	8	13	16	30
Student 29	2	404	2	9	393	30
Student 30	15	820	1	20	799	80
Mean	**3.7**	**72.8**	**10.5**	**10.7**	**52.0**	**39.7**
SD	**3.15**	**92.03**	**10.54**	**14.16**	**87.64**	**42.94**

Table 5-12 shows that the average number of search strategies was 3.7. The number of records reviewed by respondents ranged from a low of 1 to a high of 404. They performed in total 110 search sessions and retrieved 1825 records. Participants reported that 19% of search results were completely relevant, 19% partially relevant and 62% non-relevant. Figure 5-7 reports this in graphic display.

Figure 5-7. Percentage of Completely, Partial and Non-relevant Records (Study 2)



The average number of retrieved records was more than 72, of which 11 articles were deemed completely relevant, 10 articles were deemed partially relevant and 52 articles were deemed not relevant. The average time spent for each search was 39 minutes and 42 seconds.

5-3-2. Responses

Responses of study 2 have been analysed in the same way as in study 1. Table 6-13 shows the N x K (subject by item) data matrix of the 2nd experiment. Each subject's categorical value is provided in the body of the table. Item analyses were performed on the data. The median and standard deviation of each item were calculated and the Pearson *r* correlation of each item with the total score on all items was calculated.

Table 5-13. Raw Data Matrix (Study 2)

	Relevance Criteria																
Subjects	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	*Total Score*
01	4	4	4	4	4	4	3	4	1	1	2	3	2	4	4	4	52
02	4	3	4	4	4	3	4	4	1	2	4	2	3	5	4	3	54
03	4	4	4	5	5	3	4	4	2	4	4	4	3	5	4	3	62
04	5	4	5	4	5	5	4	4	2	3	3	5	4	3	2	4	62
05	4	4	3	4	4	4	4	3	2	5	4	1	3	4	4	3	56
06	4	4	3	4	2	3	2	2	1	4	3	4	2	4	4	2	48
07	5	2	5	5	5	4	3	5	1	1	3	4	2	2	2	4	53
08	4	2	4	5	3	3	1	4	1	3	2	4	2	5	4	2	49
09	3	1	3	3	1	3	3	3	1	4	3	5	1	5	2	1	42
10	4	2	4	2	5	4	4	5	1	4	3	3	2	5	4	4	56
11	4	2	4	3	4	4	4	4	3	4	4	4	3	4	3	4	58
12	4	4	3	4	3	3	3	4	1	4	2	2	2	3	3	2	47
13	4	4	4	4	3	2	2	5	4	5	2	4	2	4	4	4	57
14	4	3	4	5	4	2	2	4	3	4	5	4	3	5	5	4	61
15	5	4	4	1	5	4	3	4	4	4	4	4	2	4	4	2	58
16	4	5	4	5	5	5	3	4	2	5	3	4	3	4	4	4	64
17	4	3	2	4	5	4	1	5	2	3	4	4	2	2	3	2	50
18	4	3	3	2	3	4	2	4	3	3	3	4	4	3	4	3	52
19	3	4	2	4	2	3	1	2	4	4	4	5	2	4	2	5	51
20	3	5	4	4	3	1	1	3	1	1	1	3	3	3	5	3	44
21	4	3	3	4	3	4	4	5	2	2	4	3	2	3	4	3	53
22	4	3	3	4	3	4	4	3	2	4	3	3	2	5	5	3	55
23	5	4	4	4	3	4	5	5	4	5	3	5	3	5	5	4	68
24	1	4	4	4	5	5	5	5	4	4	4	4	2	5	5	4	65
25	2	4	2	1	2	3	2	4	3	3	3	4	3	3	4	4	47
26	3	3	4	4	4	4	3	3	2	4	3	4	3	4	4	3	55
27	3	2	3	2	5	3	3	5	1	5	2	3	3	4	4	3	51
28	5	3	4	4	5	2	3	5	3	2	4	2	3	4	4	3	56
29	4	4	5	2	4	4	4	4	1	1	3	2	4	4	3	3	52
30	4	3	5	4	1	3	4	5	2	2	3	3	3	5	5	3	55
Correlation	0.24	0.27	0.44	0.25	0.51	0.45	0.59	0.38	0.54	0.37	0.44	0.18	0.32	0.35	0.31	0.52	
Median	**4.0**	**3.0**	**4.0**	**4.0**	**4.0**	**4.0**	**3.0**	**4.0**	**2.0**	**4.0**	**3.0**	**4.0**	**3.0**	**4.0**	**4.0**	**3.0**	
S.D.	**0.9**	**1.0**	**0.8**	**1.1**	**1.2**	**0.9**	**1.2**	**0.9**	**1.1**	**1.3**	**0.9**	**1.0**	**0.7**	**0.9**	**0.9**	**0.9**	

Where 1 = Strongly Disagree, 2 = Disagree, 3 = Unsure, 4 = Agree, 5 = Strongly Agree

C1 = About the topic, C2 = Unique or only sources, C3 = Original, C4 = New, C5 = Informative, C6 = Easy to Understand, C7 = Consistent, C8 = Generally or specifically suitable C9=. Know the author(s) personally, C10 = Know the journals or conferences, C11 = The authors are eminent, C12 = Methods or techniques have described, C13 = Controversial, C14 = Good bibliography and links, C15 = Background or history, C16 = Validate viewpoint.

Table 5-14 shows the results of the study 2 and the some results can be seen presented in percentages in Table 5-15.

Table 5-14. Frequencies of Responses (Study 2)

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16
Strongly Agree	5	2	4	5	10	3	2	10	0	5	1	4	0	10	6	1
Agree	18	13	15	17	7	13	10	13	5	12	10	14	3	12	16	11
Unsure	5	9	8	2	8	10	9	5	5	5	13	7	13	6	4	12
Disagree	1	5	3	4	3	3	5	2	9	4	5	4	13	2	4	5
Strongly Disagree	1	1	0	2	2	1	4	0	11	4	1	1	1	0	0	1

C1 = About the topic, C2 = Unique or only sources, C3 = Original, C4 = New, C5 = Informative, C6 = Easy to Understand, C7 = Consistent, C8 = Generally or specifically suitable C9= Know the author(s) personally, C10 = Know the journals or conferences, C11 = The authors are eminent, C12 = Methods or techniques have described, C13 = Controversial, C14 = Good bibliography and links, C15 = Background or history, C16 = Validate viewpoint.

Table 5-15. Percentages of Responses (Study 2)

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16
Strongly Agree	16.7	6.7	13.3	16.7	33.3	10.0	6.7	33.3	0.0	16.7	3.3	13.3	0.0	33.3	20.0	3.3
Agree	60.0	43.3	50.0	56.7	23.3	43.3	33.3	43.3	16.7	40.0	33.3	46.7	10.0	40.0	53.3	36.7
Unsure	16.7	30.0	26.7	6.7	26.7	33.3	30.0	16.7	16.7	16.7	43.3	23.3	43.3	20.0	13.3	40.0
Disagree	3.3	16.7	10.0	13.3	10.0	10.0	16.7	6.7	30.0	13.3	16.7	13.3	43.3	6.7	13.3	16.7
Strongly Disagree	3.3	3.3	0.0	6.7	6.7	3.3	13.3	0.0	36.7	13.3	3.3	3.3	3.3	0.0	0.0	3.3

C1 = About the topic, C2 = Unique or only sources, C3 = Original, C4 = New, C5 = Informative, C6 = Easy to Understand, C7 = Consistent, C8 = Generally or specifically suitable C9= Know the author(s) personally, C10 = Know the journals or conferences, C11 = The authors are eminent, C12 = Methods or techniques have described, C13 = Controversial, C14 = Good bibliography and links, C15 = Background or history, C16 = Validate viewpoint.

Attitudes of respondents toward cognitive relevance criteria based upon the study 2 are presented in Table 5-16.

Table 5-16. Cumulative Percentages of Responses (Study 2)

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16
Total Agree	76.7	50.0	63.3	73.3	56.7	53.3	40.0	76.7	16.7	56.7	36.7	60.0	10.0	73.3	73.3	40.0
Unsure	16.7	30.0	26.7	6.7	26.7	33.3	30.0	16.7	16.7	16.7	43.3	23.3	43.3	20.0	13.3	40.0
Total Disagree	6.7	20.0	10.0	20.0	16.7	13.3	30.0	6.7	66.7	26.7	20.0	16.7	46.7	6.7	13.3	20.0

Total Agree = strongly agree + agree, Total Disagree = strongly disagree + disagree

C1 = About the topic, C2 = Unique or only sources, C3 = Original, C4 = New, C5 = Informative, C6 = Easy to Understand, C7 = Consistent, C8 = Generally or specifically suitable C9= Know the author(s) personally, C10 = Know the journals or conferences, C11 = The authors are eminent, C12 = Methods or techniques have described, C13 = Controversial, C14 = Good bibliography and links, C15 = Background or history, C16 = Validate viewpoint.

5-3-3. Scale Reliability

Reliability of the scale of study 2 was assessed by internal consistency. Table 5-17 shows the correlation matrix of the items.

Table 5-17. Correlation Matrix of Cognitive Relevance Criteria (Study 2)

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16
C1	1															
C2	-0.055	1														
C3	0.389	0.014	1													
C4	0.18	0.117	0.229	1												
C5	0.233	0.068	0.318	0.057	1											
C6	0.056	0.013	0.116	-0.093	0.376	1										
C7	0.108	-0.041	0.435	-0.043	0.248	0.525	1									
C8	0.185	-0.216	0.337	-0.056	0.416	0.146	0.333	1								
C9	-0.083	0.282	-0.135	-0.125	0.008	0.071	0.05	0.135	1							
C10	-0.132	0.019	-0.336	-0.031	0.029	0.151	0.107	-0.13	0.401	1						
C11	0.083	-0.11	-0.109	0.064	0.244	0.239	0.267	0.037	0.44	0.162	1					
C12	-0.091	-0.083	-0.068	0.057	-0.156	0.165	-0.222	-0.097	0.398	0.282	0.052	1				
C13	0.163	0.298	0.339	-0.143	0.23	0.081	0.181	0.075	0.069	-0.147	0.054	-0.17	1			
C14	-0.13	-0.158	0.225	0.101	-0.153	-0.081	0.392	-0.085	0.103	0.351	0.173	0.00	-0.105	1		
C15	-0.171	0.272	0.088	0.059	-0.03	-0.207	0.167	0.176	0.196	0.115	-0.043	-0.252	0.134	0.451	1	
C16	-0.089	0.284	0.184	0.11	0.219	0.174	0.161	0.166	0.429	0.03	0.133	0.185	0.29	0.00	0.05	1

The item total correlation and the reliability coefficient (Cronbach's Coefficient Alpha) was examined. The Coefficient Alpha was 0.618, suggesting that the 16 criteria are reasonably internally consistent (See Table 5-18).

Table 5-18. Reliability Analysis (Study 2)

Item-total Statistics	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Alpha if Item Deleted
About my topic	50.6000	35.9724	.0986	.6206
Unique or only sources	51.1000	35.5414	.1146	.6198
Original	50.7667	33.9092	.3203	.5916
New	50.8000	35.5448	.0707	.6306
Informative	50.7667	31.4954	.3400	.5829
Easy to understand	50.9667	33.4126	.3213	.5897
Consistent	51.4000	30.7310	.4432	.5639
Generally or specifically suitable	50.4000	34.4552	.2416	.6016
Know the author(s) personally	52.3000	31.6655	.3923	.5748
Know the journals or conferences	51.1000	33.5414	.1701	.6171
The authors are eminent	51.2667	33.7885	.3166	.5915
Controversial	50.9000	36.5069	.0204	.6346
Methods or techniques have described	51.8333	35.4540	.2080	.6068
Good bibliography and links	50.4333	34.7368	.2058	.6065
Background or history	50.6333	35.0678	.1688	.6118
Validate viewpoint	51.2333	32.8747	.4042	.5789
Reliability Coefficients Alpha = 0.618				

5-3-4. Scale Validity

Scale validity in study 2 was examined by performing factor analysis. Bartlett's test of sphericity was significant (0.04). Total variance explained consists of the factors and their associated eigenvalues, percentage of variance explained and the cumulative percentages have been calculated (See Table 5-19).

Table 5-19. Total Variance Explained (Study 2)

	Initial Eigenvalues		
Component	Total	% of Variance	Cumulative %
1	2.827	17.669	17.669
2	2.330	14.561	32.230
3	1.802	11.265	43.495
4	1.705	10.656	54.151
5	1.315	8.221	62.371
6	1.071	6.693	69.064
7	1.022	6.389	75.453
8	0.909	5.680	81.133
9	0.809	5.058	86.191
10	0.614	3.838	90.029
11	0.487	3.046	93.076
12	0.418	2.610	95.686
13	0.309	1.934	97.620
14	0.153	0.959	98.579
15	0.137	0.858	99.437
16	0.090	0.563	100.000

Seven factors can be extracted because they have eigenvalues greater than 1. More than 75.4 % of the variance would be explained, by those seven factors.

A Component matrix (Table 5-20) and Factor Correlation Matrix (Table 5-21) have been produced.

Table 5-20. Component Matrix (Study 2)

Component	1	2	3	4	5	6	7
Good bibliography and links	0.904	0.179	-0.230	-0.116			0.120
Background or history	0.746	-0.310	0.242	0.206		-0.282	
Easy to understand	-0.211	0.954	0.104	-0.167	0.121	0.157	-0.186
Consistent	0.421	0.782				-0.149	
Unique or only sources			0.970	-0.487			
Methods or techniques have described		0.212	0.717		-0.125	-0.186	
Know the author(s) personally		-0.111	-0.314	1.094			-0.138
Informative	-0.229	0.271	0.154	0.364	0.333	-0.172	0.160
Know the journals or conferences		0.151			0.883	-0.130	0.242
Generally or specifically suitable	0.101	-0.157	0.209	0.286	0.532	0.433	
The authors are eminent	0.307	0.214		-0.258	0.522	0.170	
Original	0.223	0.247	0.115	0.301	-0.419	0.138	0.408
Controversial	-0.170		-0.196			0.942	
Validate viewpoint		0.106	0.511	0.190		0.533	
New	0.166	-0.248		-0.279	0.117	0.160	0.971
About my topic	-0.222			0.178	0.198	-0.144	0.657

Table 5-21. Factor Correlation Matrix (Study 2)

Component	1	2	3	4	5	6	7
1	1						
2	-0.003	1					
3	0.123	0.054	1				
4	0.026	0.417	0.391	1			
5	0.054	-0.034	0.101	-0.055	1		
6	0.139	-0.121	0.085	-0.162	0.310	1	
7	-0.054	0.351	0.067	0.331	-0.281	-0.207	1

5-4. Data Analysis of Study 3

5-4-1. Characteristics of Respondents' Search

The study 3 was carried out in December 2001 and January 2002 with the same subjects and the same cognitive relevance criteria. Demographic data slightly changed from the study 1 and 2. The majority of changes were about the ages of participants and year of their study. The results of the search characteristics of the study 3 have been tabulated in Table 5-22.

Table 5-22. Students' Search Characteristics (Study 3)

	No. of Search Strategies	No. of Search Results	Completely Relevant	Partially Relevant	Not Relevant	Time Spent (Minutes)
Student 01	2	42	16	8	20	95
Student 02	1	11	5	0	6	15
Student 03	1	26	7	10	9	5
Student 04	5	13	1	3	9	20
Student 05	1	19	11	2	6	35
Student 06	2	26	3	2	21	20
Student 07	3	12	1	5	6	90
Student 08	10	3	3	0	0	40
Student 09	4	41	23	7	11	35
Student 10	112	101	62	32	7	200
Student 11	4	23	16	4	2	40
Student 12	3	153	30	50	73	30
Student 13	2	16	10	2	4	5
Student 14	7	30	8	5	17	45
Student 15	6	1	1	0	0	10
Student 16	5	25	2	3	20	12
Student 17	2	125	35	67	23	90
Student 18	6	297	48	121	128	15
Student 19	3	27	5	7	15	15
Student 20	4	14	3	2	5	15
Student 21	4	16	7	10	3	20
Student 22	5	3	2	1	0	15
Student 23	2	52	6	15	31	60
Student 24	1	100	75	23	2	15
Student 25	8	10	0	2	6	50
Student 26	5	12	1	8	3	10
Student 27	2	42	19	8	13	50
Student 28	1	39	6	8	25	25
Student 29	1	0	0	0	0	15
Student 30	10	200	2	23	177	100
Mean	**7.4**	**49.3**	**13.6**	**14.3**	**21.4**	**39.7**
SD	**19.9**	**66.9**	**18.8**	**25.3**	**38.9**	**41.0**
Mean without student no 10	**3.8**	**47.5**	**11.9**	**13.7**	**21.9**	**34.2**
SD without student no 10	**2.6**	**67.4**	**16.8**	**25.5**	**39.5**	**28.2**

Table 5-22 shows that there was a student (student No. 10) who spent more than three hours on this experiment. He performed 112 search strategies and then made a judgement for degree of relevancy of obtained records. This of course increased the means and standard deviations of these two variables. Because of this, I decided to report the results with two scores, one including that participant's response and one without that response. The average number of search strategies was 3.7 (7.4 with no. 10). Participants recorded that 28% of search results were completely relevant, 29% partially relevant and 43% non-irrelevant. Figure 6-8 reports this in graphic display.

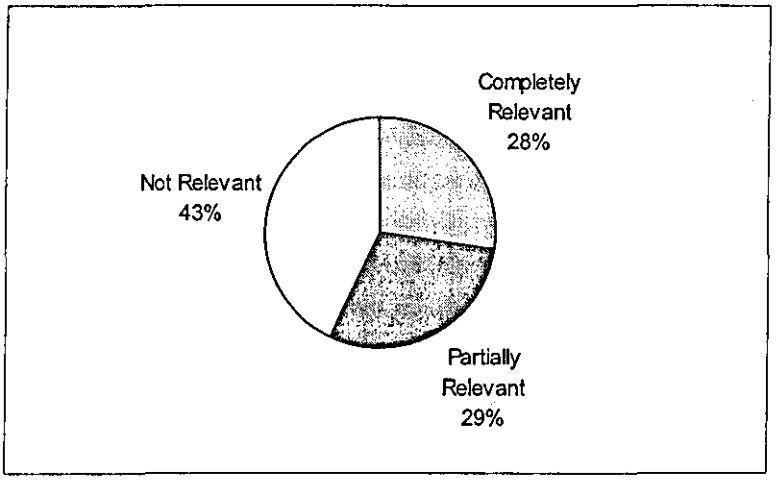


Figure 5-8. Percentage of Completely, Partial and Non-relevant Records (Study 3)

The average number of records retrieved was more than 39, of which an average articles were deemed completely relevant, 13 articles were deemed partially relevant and 14 articles were deemed not relevant. The average time spent for each search was 28 minutes and 12 seconds (39 minutes and 42 seconds with no. 10).

5-4-2. Responses

A data matrix of study 3 (Table 5-23) was produced and item analyses were performed on the data. The median and standard deviation of each item were calculated and the Pearson *r* correlation of each item with the total score on all items was found.

Table 5-23. Raw Data Matrix (Study 3)

	Relevance Criteria																
Subjects	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	*Total Score*
01	4	4	4	4	4	4	3	4	2	3	3	3	2	4	4	4	56
02	4	4	4	4	4	3	3	4	2	3	4	2	3	4	2	4	54
03	4	3	3	4	4	3	4	2	2	3	3	4	3	4	4	4	54
04	4	5	4	4	4	3	3	4	2	2	2	4	3	4	3	3	54
05	4	3	3	4	4	3	3	3	2	4	3	2	3	4	4	3	52
06	4	4	4	4	3	3	3	3	2	3	3	4	3	4	4	3	54
07	4	3	4	5	5	5	4	4	2	2	3	4	3	3	2	2	55
08	5	4	4	4	3	3	5	5	1	1	1	3	3	2	2	3	49
09	4	2	3	3	2	3	3	3	1	4	3	3	2	4	3	3	46
10	2	4	4	4	4	3	4	4	1	3	4	4	2	4	4	4	55
11	4	4	4	4	5	4	3	4	3	4	4	4	3	4	4	4	62
12	4	4	4	5	3	2	2	4	1	3	3	4	2	3	3	2	49
13	4	4	4	2	2	2	2	4	4	5	4	4	2	4	2	4	53
14	4	4	5	5	4	3	2	5	4	5	4	3	2	5	4	3	62
15	5	5	5	2	5	5	3	5	1	4	5	5	3	4	4	3	64
16	2	5	5	4	5	4	4	5	2	4	3	4	3	5	4	4	63
17	5	4	4	5	5	4	2	5	4	4	4	4	2	3	2	2	59
18	4	3	3	2	3	4	2	3	2	4	3	4	3	4	4	2	50
19	2	4	4	4	4	2	2	2	1	4	1	2	2	1	2	4	41
20	4	5	2	5	3	1	2	4	2	2	2	3	1	2	3	1	42
21	4	4	4	4	3	4	2	4	2	4	3	4	3	4	3	4	56
22	4	4	4	3	4	4	4	4	3	4	4	4	4	4	4	3	61
23	3	4	3	4	3	4	4	4	1	4	3	4	3	4	4	3	55
24	5	4	4	4	4	4	4	5	4	4	4	4	2	5	5	5	67
25	3	4	3	2	4	2	2	4	4	2	2	4	4	4	4	1	49
26	4	3	4	4	3	4	3	4	2	5	3	4	3	4	4	4	58
27	4	2	4	5	4	3	2	4	1	3	2	5	2	3	4	2	50
28	5	3	4	4	4	2	3	4	2	2	2	3	4	3	4	4	53
29	2	4	4	1	1	4	4	2	1	1	2	4	3	3	2	4	42
30	4	3	5	4	3	4	4	4	2	3	4	3	3	4	4	3	57
Correlation	0.34	0.21	0.56	0.16	0.57	0.60	0.25	0.64	0.46	0.49	0.74	0.35	0.21	0.72	0.52	0.35	
Median	**4.0**	**4.0**	**4.0**	**4.0**	**4.0**	**3.0**	**3.0**	**4.0**	**2.0**	**3.5**	**3.0**	**4.0**	**3.0**	**4.0**	**4.0**	**3.0**	
SD	**0.9**	**0.8**	**0.7**	**1.0**	**1.0**	**1.0**	**0.9**	**0.9**	**1.0**	**1.1**	**1.0**	**0.8**	**0.7**	**0.9**	**0.9**	**1.0**	

Where 1 = Strongly Disagree, 2 = Disagree, 3 = Unsure, 4 = Agree, 5 = Strongly Agree

C1 = About the topic, C2 = Unique or only sources, C3 = Original, C4 = New, C5 = Informative, C6 = Easy to Understand, C7 = Consistent, C8 = Generally or specifically suitable C9= Know the author(s) personally, C10 = Know the journals or conferences, C11 = The authors are eminent, C12 = Methods or techniques have described, C13 = Controversial, C14 = Good bibliography and links, C15 = Background or history, C16 = Validate viewpoint.

Frequencies of responses of the study 3 and their percentages are tabulated in Table 5-24 and Table 5-25.

Table 5-24. Frequencies of Responses (Study 3)

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16
Strongly Agree	5	4	4	6	5	2	1	6	0	3	1	2	0	3	1	1
Agree	19	17	19	17	13	12	9	17	5	12	9	18	3	18	17	12
Unsure	2	7	6	2	9	10	10	4	2	8	12	7	16	6	5	10
Disagree	4	2	1	4	2	5	10	3	14	5	6	3	10	2	7	5
Strongly Disagree	0	0	0	1	1	1	0	0	9	2	2	0	1	1	0	2

C1 = About the topic, C2 = Unique or only sources, C3 = Original, C4 = New, C5 = Informative, C6 = Easy to Understand, C7 = Consistent, C8 = Generally or specifically suitable C9= Know the author(s) personally, C10 = Know the journals or conferences, C11 = The authors are eminent, C12 = Methods or techniques have described, C13 = Controversial, C14 = Good bibliography and links, C15 = Background or history, C16 = Validate viewpoint.

Table 5-25. Percentages of Responses (Study 3)

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16
Strongly Agree	16.7	13.3	13.3	20.0	16.7	6.7	3.3	20.0	0.0	10.0	3.3	6.7	0.0	10.0	3.3	3.3
Agree	63.3	56.7	63.3	56.7	43.3	40.0	30.0	56.7	16.7	40.0	30.0	60.0	10.0	60.0	56.7	40.0
Unsure	6.7	23.3	20.0	6.7	30.0	33.3	33.3	13.3	6.7	26.7	40.0	23.3	53.3	20.0	16.7	33.3
Disagree	13.3	6.7	3.3	13.3	6.7	16.7	33.3	10.0	46.7	16.7	20.0	10.0	33.3	6.7	23.3	16.7
Strongly Disagree	0.0	0.0	0.0	3.3	3.3	3.3	0.0	0.0	30.0	6.7	6.7	0.0	3.3	3.3	0.0	6.7

C1 = About the topic, C2 = Unique or only sources, C3 = Original, C4 = New, C5 = Informative, C6 = Easy to Understand, C7 = Consistent, C8 = Generally or specifically suitable C9= Know the author(s) personally, C10 = Know the journals or conferences, C11 = The authors are eminent, C12 = Methods or techniques have described, C13 = Controversial, C14 = Good bibliography and links, C15 = Background or history, C16 = Validate viewpoint.

Overall attitudes of respondents are summarised in Table 5-26.

Table 5-26. Cumulative Percentages of Responses (Study 3)

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16
Total Agree[1]	75.0	75.0	75.0	75.0	66.7	45.8	33.3	79.2	20.8	58.3	37.5	70.8	12.5	70.8	62.5	50.0
Unsure	9.1	18.2	18.2	4.5	27.3	27.3	27.3	4.5	9.1	18.2	27.3	13.6	54.5	18.2	13.6	22.7
Total Disagree	16.7	4.2	4.2	20.8	8.3	20.8	37.5	12.5	70.8	25.0	33.3	12.5	33.3	12.5	25.0	20.8

Total Agree = strongly agree + agree, Total Disagree = strongly disagree + disagree

C1 = About the topic, C2 = Unique or only sources, C3 = Original, C4 = New, C5 = Informative, C6 = Easy to Understand, C7 = Consistent, C8 = Generally or specifically suitable C9= Know the author(s) personally, C10 = Know the journals or conferences, C11 = The authors are eminent, C12 = Methods or techniques have described, C13 = Controversial, C14 = Good bibliography and links, C15 = Background or history, C16 = Validate viewpoint.

5-4-3. Scale Reliability

To test the reliability of the scale in this study, a correlation matrix was produced (See Table 5-27).

1 'Strongly Agree' responses and 'Agree' responses categorised under the term 'Total Agree' and all 'Strongly Disagree' responses and 'Disagree' responses can categorised under the term 'Total Disagree'.

Table 5-27. Correlation Matrix of Cognitive Relevance Criteria (Study 3)

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16
C1	1															
C2	-0.161	1														
C3	0.019	0.201	1													
C4	0.221	-0.070	0.052	1												
C5	0.170	0.205	0.290	0.428	1											
C6	0.104	-0.042	0.436	0.136	0.274	1										
C7	-0.081	0.012	0.178	0.066	0.015	0.435	1									
C8	0.428	0.366	0.439	0.311	0.438	0.219	0.051	1								
C9	0.288	0.160	0.069	0.010	0.247	-0.032	-0.267	0.366	1							
C10	0.054	-0.078	0.195	0.034	0.141	0.210	-0.331	0.155	0.342	1						
C11	0.252	0.103	0.374	0.061	0.273	0.477	0.039	0.380	0.379	0.582	1					
C12	0.060	0.084	0.168	0.198	0.092	0.393	0.019	0.238	0.048	0.054	0.251	1				
C13	0.028	-0.070	0.130	0.382	0.137	0.294	0.348	-0.068	0.043	-0.194	-0.036	0.109	1			
C14	0.059	0.034	0.267	0.163	0.135	0.410	0.146	0.302	0.417	0.466	0.661	0.272	0.222	1		
C15	0.088	-0.110	0.034	0.030	0.256	0.178	0.113	0.161	0.142	0.298	0.304	0.272	0.198	0.611	1	
C16	-0.127	0.053	0.394	0.129	-0.042	0.239	0.387	-0.095	-0.017	0.273	0.212	-0.145	0.075	0.303	0.117	1

The matrix shows how each criterion item is correlated to the other items. The item total correlation and alpha measure for all the items were calculated (See Table 5-28).

Table 5-28. Reliability Analysis (Study 3)

Item-total Statistics	Scale Mean If Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Alpha if Item Deleted
About my topic	50.233	39.013	0.216	0.519	0.733
Unique or only sources	50.300	40.631	0.092	0.429	0.742
Original	50.200	37.614	0.485	0.570	0.712
New	50.300	41.114	-0.005	0.583	0.759
Informative	50.433	35.909	0.458	0.594	0.709
Easy to understand	50.767	35.564	0.499	0.646	0.705
Consistent	51.033	40.033	0.116	0.676	0.743
Generally or specifically suitable	50.200	35.683	0.556	0.710	0.701
Know the author(s) personally	51.967	36.999	0.326	0.485	0.723
Know the journals or conferences	50.767	36.392	0.347	0.674	0.721
The authors are eminent	51.033	33.826	0.658	0.721	0.687
Controversial	50.433	39.289	0.236	0.373	0.731
Methods or techniques have described	51.367	40.723	0.102	0.544	0.740
Good bibliography and links	50.400	34.662	0.643	0.772	0.691
Background or history	50.667	36.920	0.406	0.576	0.715
Validate viewpoint	50.900	38.576	0.211	0.603	0.735
Reliability Coefficients Alpha = 0.7355		**Standardised item Alpha =0.7375**			

5-4-4. Scale Validity

Validity of the scale was examined by factor analysis. Bartlett's test of sphericity was significant (0.03). Eigenvalues, percentage of variance explained and the cumulative percentages were computed (See Table 5-29).

Table 5-29. Total Variance Explained (Study 3)

	Initial Eigenvalues		
Component	Total	% of Variance	Cumulative %
1	3.870	24.187	24.187
2	2.294	14.337	38.524
3	1.739	10.866	49.389
4	1.458	9.111	58.500
5	1.315	8.218	66.717
6	1.055	6.596	73.314
7	0.999	6.242	79.556
8	0.786	4.913	84.469
9	0.584	3.648	88.116
10	0.456	2.847	90.964
11	0.433	2.703	93.667
12	0.300	1.873	95.540
13	0.235	1.469	97.010
14	0.223	1.395	98.405
15	0.172	1.073	99.478
16	0.083	0.522	100.000

Component matrix (Table 5-30) and Factor Correlation Matrix (Table 5-31) revealed six factors with eigenvalues greater than 1. More than 73.3 % of the variance would be explained, if those six factors were extracted.

Table 5-30. Component Matrix (Study 3)

Component	1	2	3	4	5	6
Know the journals or conferences	0.880			-0.440		
Good bibliography and links	0.812					
The authors are eminent	0.728					
Know the authors personally	0.581		-0.367			0.315
Background or history	0.563					-0.369
New		0.795		-0.370		
Informative		0.721				
About my topic		0.643				
Generally or specifically suitable		0.637				0.376
Validate viewpoint	0.365		0.744		-0.445	
Consistent			0.723	0.469		
Original			0.629			0.332
Easy to understand			0.535		0.466	
Controversial				0.879		
Methods or techniques have described					0.929	
Unique or only sources						0.923

Table 5-31. Factor Correlation Matrix (Study 3)

Component	1	2	3	4	5	6
1	1					
2	0.226	1				
3	0.108	0.027	1			
4	0.101	-0.044	0.037	1		
5	0.198	0.211	0.205	0.233	1	
6	0.233	0.189	-0.019	-0.044	0.078	1

To summarise, this chapter has comprehensively examined the reliability and validity of three studies. Based on detailed statistical analyses sixteen items, that were selected through the pilot study and the study of experts' attitude, were examined for reliability and validity. Internal consistency reliability were tested for study 1 (α=0.68), study 2 (α=0.61), and study 3 (α=0.73). Overall, the results of correlation analyses revealed that the criteria were reasonably internally consistent and three measurement scales were reliable.

Content validity and construct validity were assessed through factor analysis which is a powerful multivariate analysis. In order to identify the number of components, factor analysis technique was applied to study 1 (five factors emerged which explained 67.25% of the variance), study 2 (seven factors emerged which explained 75.45% of the variance), and study 3 (six factors emerged which explained 73.3% of the variance). The results of factor analysis of three studies confirmed the validity of the scale in each study.

Chapter 6: Results – Measurement Models

6-1. Introduction

The purpose of this chapter is to present the result of the longitudinal study that was undertaken to identify the links between the cognitive factors of relevance judgement and their criteria. In this study, a 16-item scale that measured the criteria was used (Appendix 7). For each item, respondents were presented with five possible responses from "strongly agree" to "strongly disagree". There were coded from 5 to 1. These item scores represented ordinal scale measurement. Joreskog and Moustaki (2001) argued that when the measured variables are ordinal data, the categorical nature of these variables should be taken into account. In this study, the participants were 30 PhD students from two universities in the UK and in the science and engineering area. The attitudes of the same individuals were measured on three occasions during 2001. The purpose of this longitudinal study was to assess the level of stability of attitudes over time.

6-2. Hypothesised Factor-Analytic Model

As described in chapter 4, an exploratory factor analysis was carried out to examine the covariation among a set of observed variables in order to gather information on their underlying latent factors. Byrne (1998) stated that exploratory factor analysis is designed for the situation where links between the observed and latent variables are unknown or uncertain. The analysis proceeds in an exploratory mode to determine how, and to what extent, the observed variables are linked to their underlying factors. The criteria of cognitive relevance as the observed variables are listed in Table 6-1.

Table 6-1. Criteria and Factors of Cognitive Relevance Judgement

Factors	Criteria	Definition
Aboutness	About my topic	The extent to which information is related to the participant's topic
	Informative	The extent to which information is in-depth, provides a summary, interpretation or explanation, provides a sufficient variety of approaches
	Generally or specifically suitable	Information is specific to participant's need and has sufficient depth
	Background	The extent to which a document provides background or context information
	Good bibliography and links	The extent to which document has a sufficient bibliography and/or links to other information
	Controversial	The extent to which the participant is encouraged by the text to engage in two-way interacting with statements made in the document
Quality of Information	Validate viewpoint	The extent to which the participant agrees with the information presented, or the information presented supports the participant's point of view
	Consistent	The extent to which information is consistent with or supported by other information within the field
	Easy to understand	The participant's judgement is that he/she can understand or follow the information presented
	Methods or techniques have described	The extent to which the document has sufficient detail about techniques or methods
Characteristics of Information	Know the author(s)	The extent to which the participant has familiarity or professional relationship with the author of a document
	Know the journals or conferences	The extent to which the participant has familiarity with the journal or the extent to which a source of the document is well-known or reputable
	The authors are eminent	The extent to which each author is well-known as a major researcher in the subject
Information Novelty	New	The extent to which information is recent or up to date
	Original	The extent to which the document is novel
	Unique or only sources	The extent to which the document is unique

Figure 6-1 presents the factor-analytic model based on exploratory factor analysis (See Chapter 3).

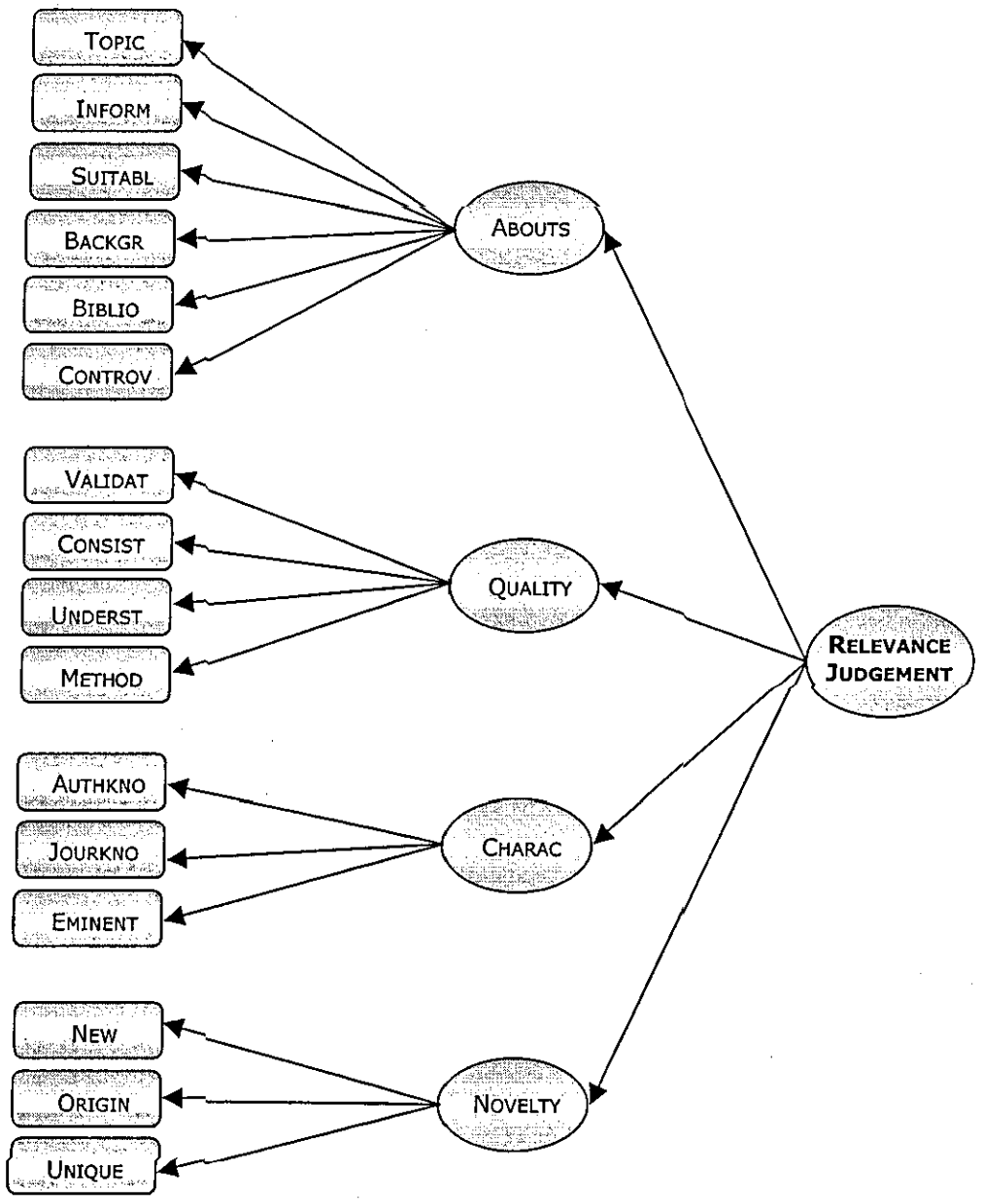


Figure 6-1. Hypothesised Factor-Analytic Model

In Figure 6-1, variables on the left side are the independent observed variables, i.e., TOPIC (About my topic), INFORM (Informative), SUITAB (Generally or specifically suitable), BACKGR (Background or history), BIBLIO (Good bibliography and links), CONTROV (Controversial), VALIDAT (Validate my viewpoint, CONSIST (Consistent), Underst (Easy to understand), METHOD (Methods or techniques have described), AUTHKNO (Know the author(s) personally), JOURKNO (Know the journals or conferences), EMINENT (The authors are eminent), NEW (New), ORIGIN (Original), and UNIQUE (Unique or only sources).

Variables on the right side of Figure 6-1 (curves) are independent latent variables. These are: ABOUTS (Aboutness); QUALITY (Quality of Information); CHARAC (Characteristics of Information); and NOVELTY (Information Novelty). Each latent variable has a number of indicator variables. The relationships among the latent variables (factors) determine how the independent variables influence or affect the dependent variables.

As mentioned in Chapter 3, the software used in this part of the study was LISREL. These are two basic types of variables in LISREL: latent variables and observed variables. Latent variables are those that formulated in terms of theoretical or hypothetical concepts, i.e., constructs which are not directly measurable or observable and that can be used as indicators of latent variables (Diamantopoulos and Siguaw 2000). In other words, latent variables are represented or measured by one or more observed variables. This model can be translated into LISREL notation. LISREL describes the relations among all the variables by a set of equations (Hayduk 1987; Schumacker and Lomax 1996). According to Vaughan (1999) the LISREL methodology involves the following process:

- To identify variables to be used and specify an initial model which indicates the relationships among variables;
- To collect data on the observed variables;
- To test the model against the data collected;
- To revise the model if necessary and re-test it.

6-3. The Measurement Models

The research aims require the building and testing of models depicting factors that influence relevance judgement. For Study 1, construction of the measurement model was carried out with the relevance judgement criteria 'regressed' on the latent factors. This construction reflects the theoretical model for the exploratory study of relevance judgement presented as Figure 6-1.

6-3-1. The Sub-models for the Studies

The graphical presentation of models provided in the following sections display the parameter estimates and the fit statistics generated by LISREL. The parameters displayed in the paths are the standardised coefficients. The paths leading back from the factors in the ovals point to the related composite variables formed from the data of Study 1, Study 2 and Study 3. The coefficients indicate the magnitude of the contribution of each independent latent variable to the dependent latent variable (here relevance judgement)

6-3-1-1. Sub-models of Study 1

Figure 6-2 to Figure 6-6 presents the sub-models of Study 1.

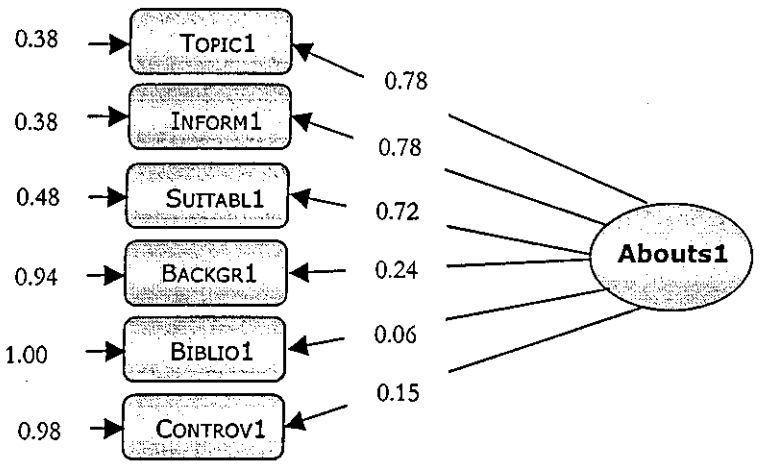


Figure 6-2. Sub-model of Aboutness (Study 1)

In this sub-model, Aboutness (Abouts1) is influenced by six criteria (TOPIC1, INFORM1, SUITABL1, BACKGR1, BIBLIO1, and CONTROV1). The standard path coefficients can be interpreted by using a general rule of thumb. Effects are meaningful, but weak when path coefficient are >0.05 and <0.1; moderate when they are >0.11 and <0.25, powerful or strong to very strong when they are >0.26 (Pedhazur and Kerlinger 1982). Following this general rule on interpretation of the path coefficients, the paths of Aboutness to TOPIC1 (0.78), INFORM1 (0.78), SUITABL1 (0.72) are very strong, while the paths to BACKGR1 (0.24) and CONTROV1 (0.15) are moderate, and of that BIBLIO1 (0.06) is weak.

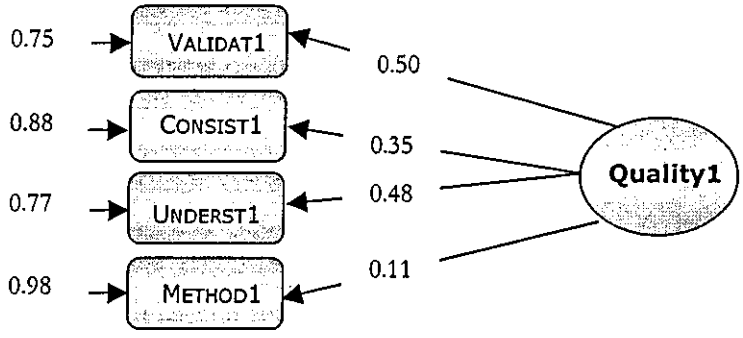


Figure 6-3. Sub-model of Quality of Information (Study 1)

This sub-model contains Quality of information as dependent variable and four criteria, which are VALIDAT1, CONSIST1, UNDERST1, and METHOD1. As this measurement model indicates the effects of VALIDAT1 (0.50), CONSIST1 (0.35), UNDERST1 (0.48) are strong. The path coefficient of METHOD1 (0.11) can be interpreted as insignificant.

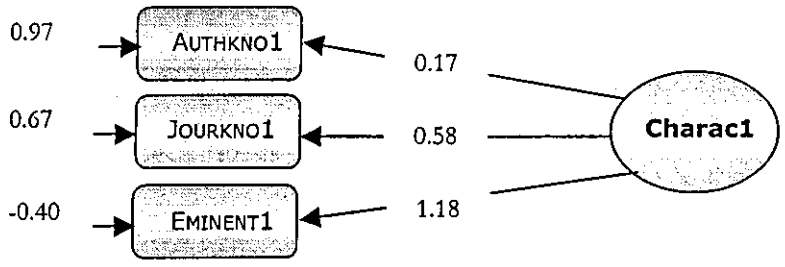


Figure 6-4. Sub-model of Characteristics of Information (Study 1)

As shown on this sub-model, the effect of AUTHKNO1 on Charac1 is (0.17) is moderate. The effect of JOURKNO1 on Charac1 is 0.58, and EMINENT1 on Charac1 is 1.18. These coefficients are powerful.

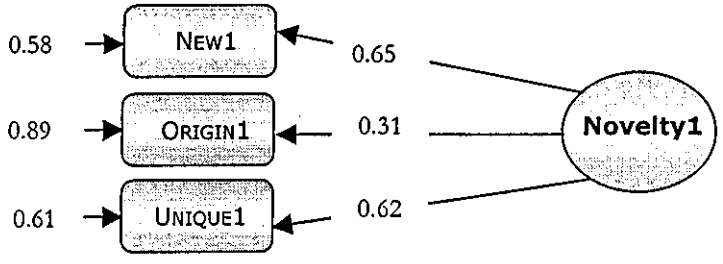


Figure 6-5. Sub-model of Information Novelty (Study 1)

As this sub-model shows, the coefficient paths from Novelty1 to NEW1, ORIGIN1, and UNIQUE1 are 0.65, 0.31 and 0.62 receptively. It means NEW1 and UNIQUE1 are two important criteria of Novelty1 and ORIGIN1 is less important.

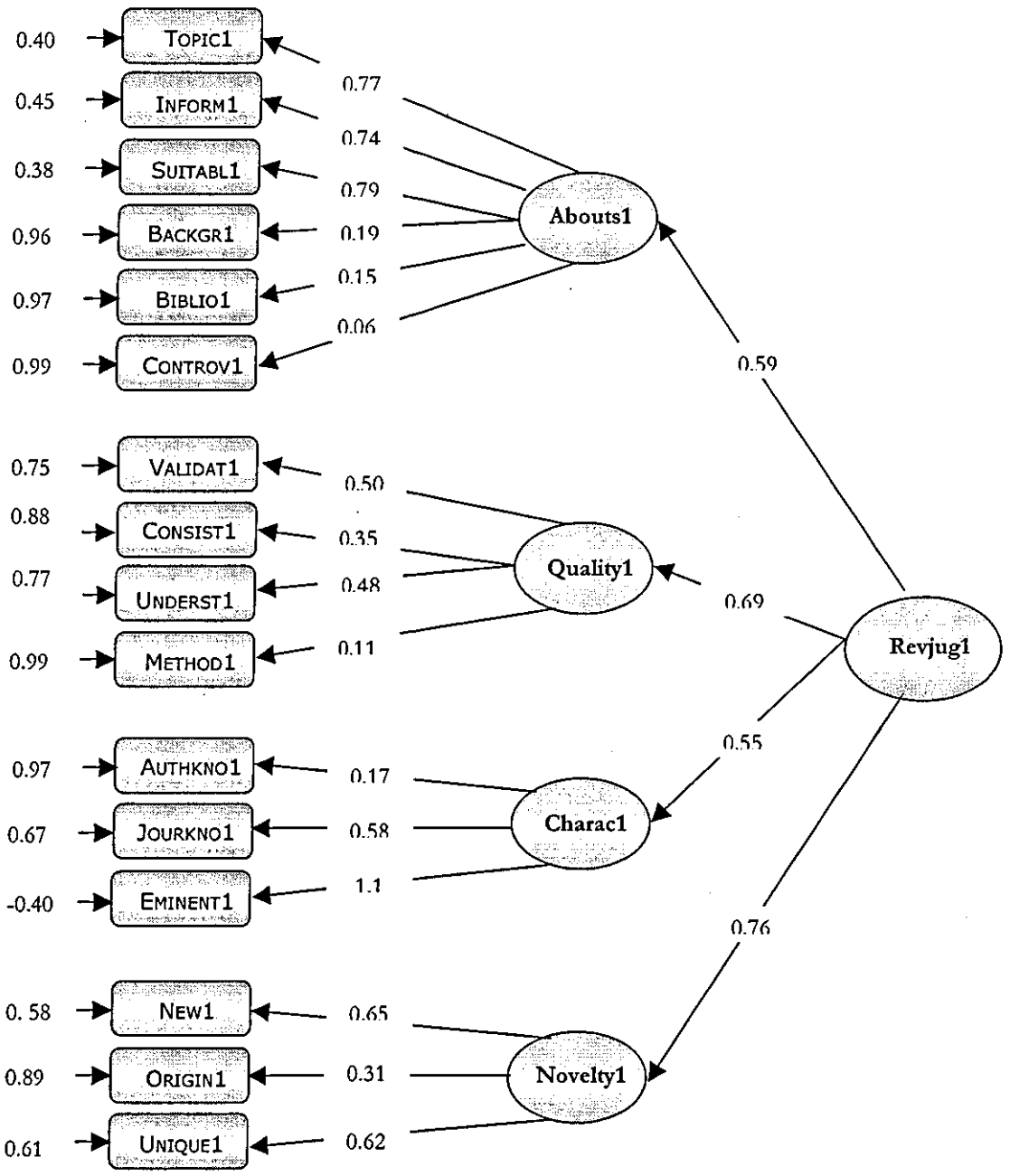


Figure 6-6. Measurement Model of Relevance Judgement (Study 1)

Figure 6-6 illustrates the overall measurement model for Study 1. The coefficient paths from Revjug1 to Abouts1 (0.59) (described as TOPIC1, INFORM1, SUITABL1, BACKGR1, BIBLIO1, and CONTROV1) and Quality1 (0.69) (described as VALIDAT1, CONSIST1, UNDERST1, and METHOD1) Charac1 (0.55) (described as AUTHKNO1, JOURKNO1, and EMINENT1) and Novelty1 (0.76) (described as NEW1, ORIGIN1, and UNIQUE1) are high. This model suggests that Aboutness, Quality of Information, Characteristics of Information and Information Novelty are significant factors in affecting relevance judgement.

6-3-1-2. Sub-models of Study 2

The measurement sub-models of Study 2 are presented in this section.

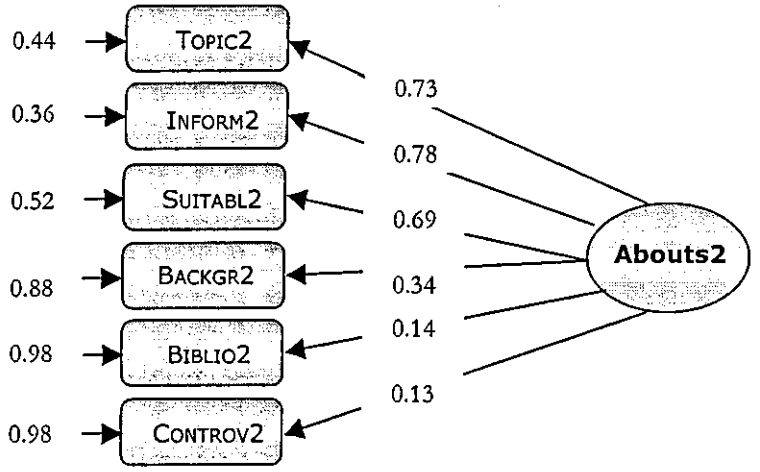


Figure 6-7. Sub-model of Aboutness (Study 2)

This sub-model shows that the path coefficient of Abouts2 to TOPIC2 (0.74), INFORM2 (0.78), and SUITABL2 (0.69) are stronger than BACKGR2 (0.34), BIBLIO2 (0.14), and CONTROV2 (0.13). The coefficient of BIBLIO2 and CONTROV2 are not statistically significant.

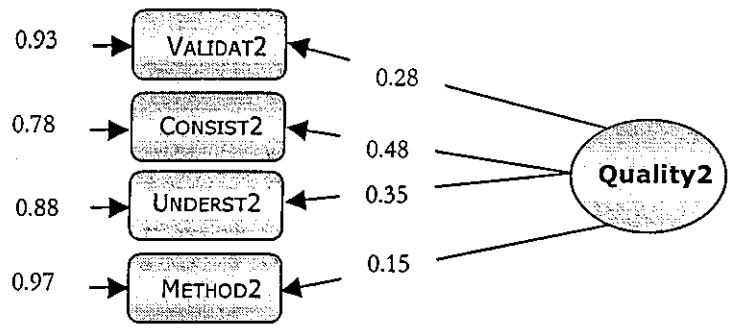


Figure 6-8. Sub-model of Quality of Information (Study 2)

This figure demonstrates that the coefficient paths of VALIDAT2 (0.28) and METHOD2 (0.15) are moderate, while CONSIST2 (0.48) UNDERST2 (0.35) are strong.

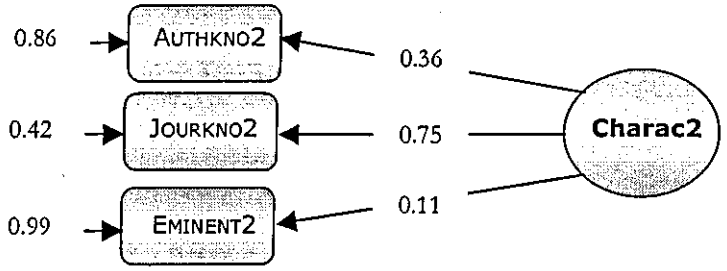


Figure 6-9. Sub-model of Characteristics of Information (Study 2)

The three coefficient paths from Charac2 to AUTHKNO2 (0.36), JOURKNO2 (0.75), and EMINENT2 (0.11) show the criteria AUTHKNO2 and JOURKNO2 have strong influence on Characteristics of information.

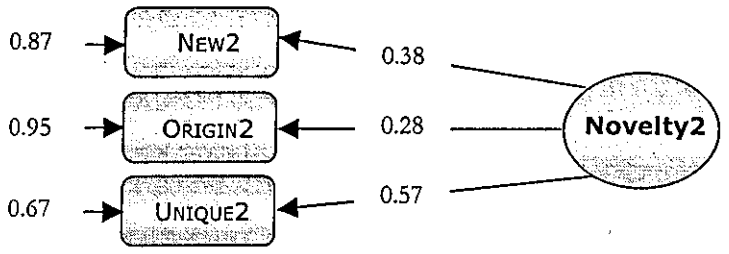


Figure 6-10. Sub-model of Information Novelty (Study 2)

As displayed in Figure 6-10, the coefficient paths of Novelty2 to UNIQUE2 is 0.57, and of that NEW2 and ORIGIN 2 are 0.38 and 0.28 respectively.

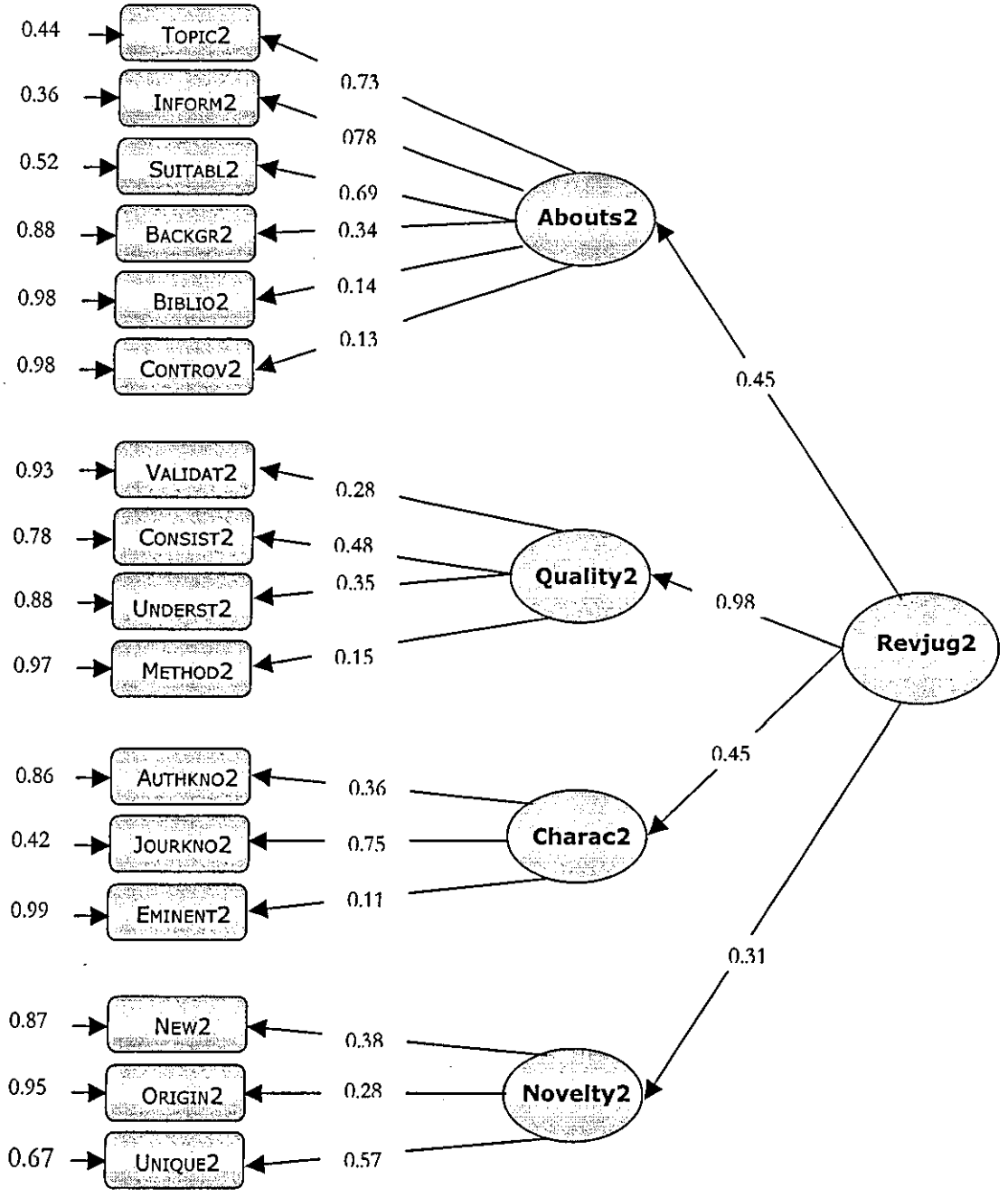


Figure 6-11. The Measurement Model of Relevance Judgement (Study 2)

A comparison of the four path coefficients shows that Quality of Information (Quality2: 0.98) (described as VALIDAT2, CONSIST2, UNDERST2, and METHOD2) was the most important factor affecting relevance judgement, followed by Aboutness (Abouts2: 0.45) (described as TOPIC2, INFORM2, SUITABL2, BACKGR2, BIBLIO2, CONTROV2), Characteristics of Information (Charac2: 0.45) (described as AUTHKNO2, JOURKNO2, and EMINENT2) and Information Novelty (Novelty2: 0.31) (described as NEW2, ORIGIN2, and UNIQUE2).

6-3-1-3. Sub-models of Study 3

In this section, four sub-models of Study 3 and a measurement model of whole study are discussed.

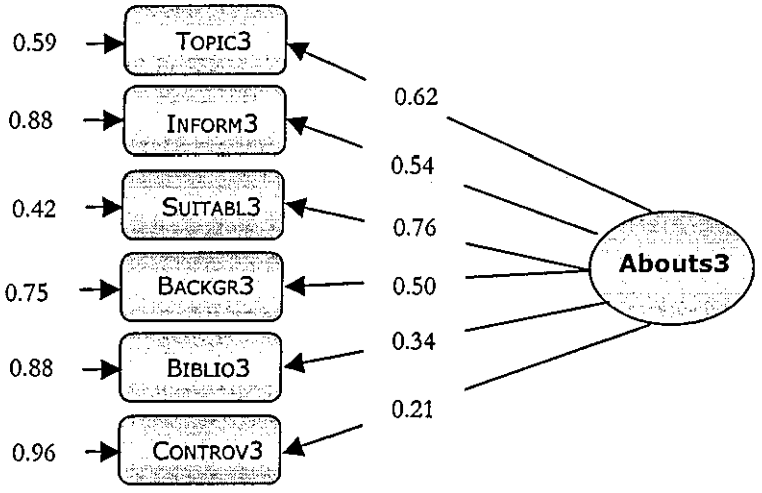


Figure 6-12. Sub-model of Aboutness (Study 3)

This sub-model measures the influence of six criteria, i.e., TOPIC3, INFORM3, SUITABL3, BACKGR3, BIBLIO3, and CONTROV3 on Abouts3. As Figure 6-12 shows, the strongest influence is shown by SUITABL3, with 0.76 coefficient. The contribution of TOPIC3 (0.62), INFORM3 (0.54), and BACKGR3 (0.50) are strong. BIBLIO3 (0.34) and CONTROV3 (0.21) have less influence.

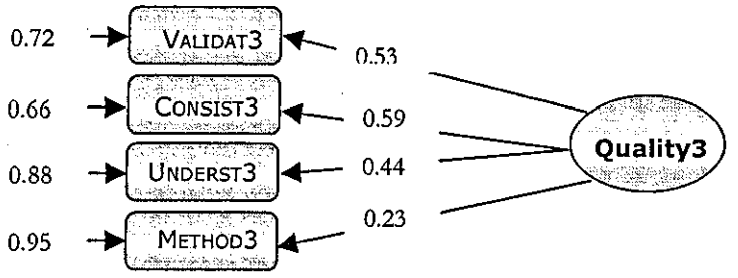


Figure 6-13. Sub-model of Quality of Information (Study 3)

The paths from Quality3 to four hypothesised criteria, VALIDAT3 (0.53), CONSIST3 (0.59), UNDERST3 (0.44), and METHOD3 (0.23) in this measurement model indicate that VALIDAT3, CONSIST3, and UNDERST3 have the strongest effect on quality of information (Quality3).

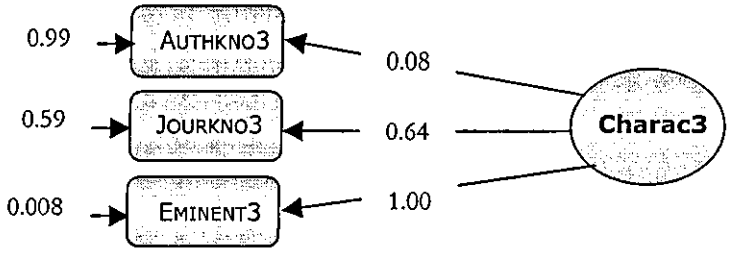


Figure 6-14. Sub-model of Characteristics of Information (Study 3)

This sub-model represents three criteria, i.e., AUTHKNO3, JOURKNO3, and EMINENT3 that, effect characteristics of information (Charac3). The criteria JOURKNO3 and EMINENT3 with 0.64 and 1.00 path coefficients respectively have the highest influence on characteristics of information (Charac3).

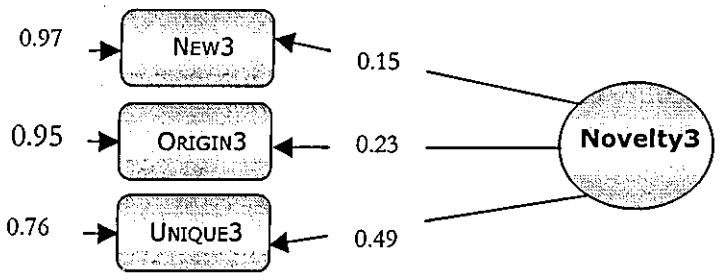


Figure 6-15. Sub-model of Information Novelty (Study 3)

The coefficient paths presented in Figure 6-15 show that UNIQUE3 (0.49) affects information novelty (Novelty3). NEW3 (0.15) and ORIGIN3 (0.23) are less important in this sub-model.

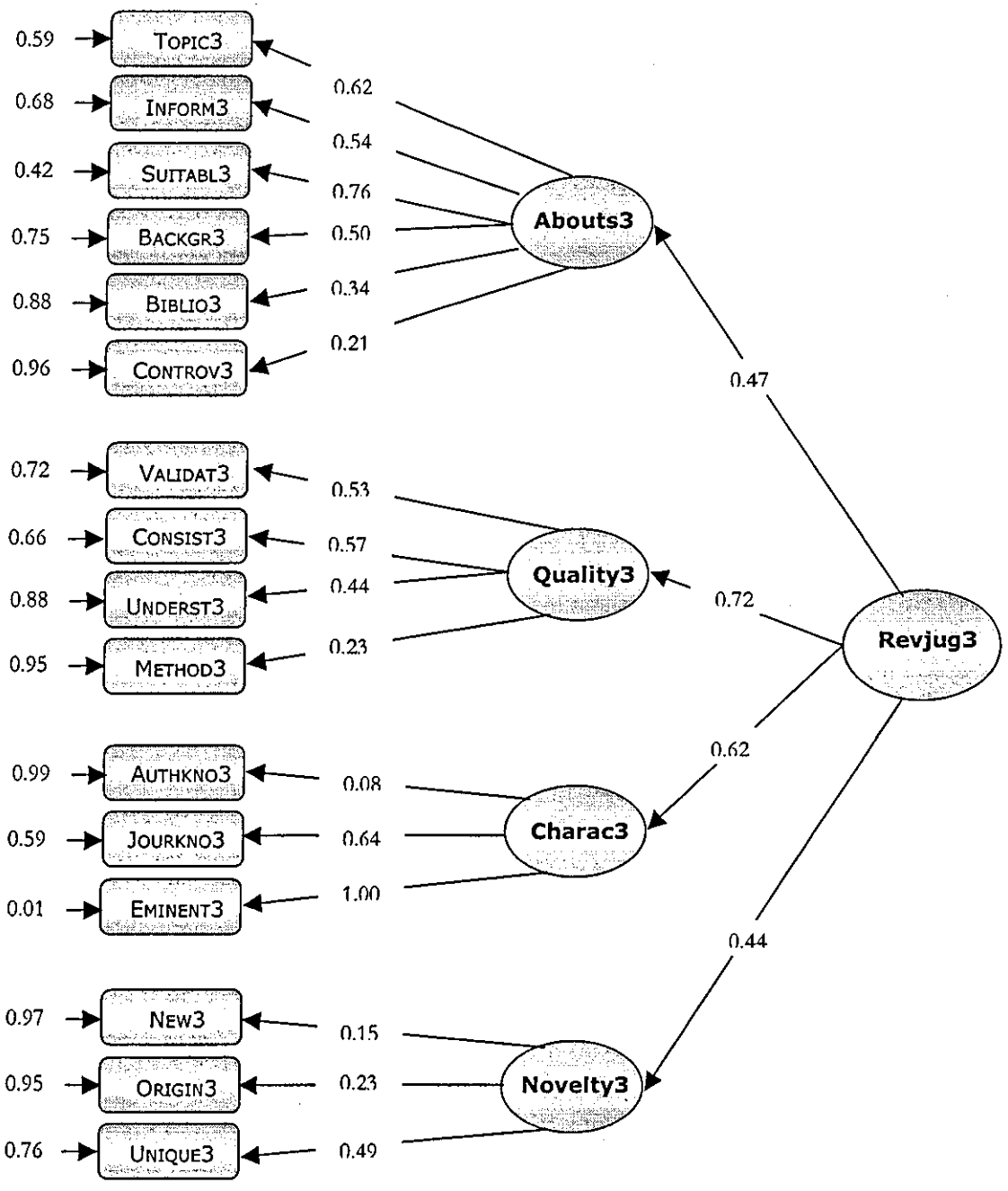


Figure 6-16. The Measurement Model of Relevance Judgement (Study 3)

Overall, this model shows that Aboutness (described as TOPIC3, INFORM3, SUITABL3, BACKGR3, BIBLIO3, and CONTROV3), Quality of Information (described as VALIDAT3, CONSIST3, UNDERST3, and METHOD3), and Characteristics of Information (described as AUTHKNO3, JOURKNO3, and EMINENT3) as well as Information Novelty (described as NEW3, ORIGIN3, and UNIQUE3) are all important in establishing relevance judgement.

In this Chapter, we have used tables and figures to illustrate the relationship that we found among relevance criteria in three studies. Table 6-2 shows the results of the three studies.

Table 6-2. Path Coefficient and Errors of Study 1, 2 and 3

Factors		Criteria	Study 1		Study 2		Study 3	
			Coefficient	*Error*	*Coefficient*	*Error*	*Coefficient*	*Error*
Aboutness	<--	About my topic	0.77	0.40	0.72	0.44	0.62	0.59
Aboutness	<--	Informative	0.74	0.45	0.78	0.36	0.54	0.68
Aboutness	<--	Generally or specifically suitable	0.79	0.38	0.69	0.52	0.76	0.42
Aboutness	<--	Background	0.19	0.96	0.34	0.88	0.50	0.75
Aboutness	<--	Good bibliography and links	0.16	0.97	0.13	0.98	0.34	0.88
Aboutness	<--	Controversial	0.07	0.99	0.14	0.98	0.21	0.96
Quality of Information	<--	Validate viewpoint	0.50	0.75	0.28	0.93	0.53	0.72
Quality of Information	<--	Consistent	0.35	0.88	0.48	0.78	0.59	0.66
Quality of Information	<--	Easy to understand	0.48	0.77	0.35	0.88	0.44	0.88
Quality of Information	<--	Methods or techniques have described	0.11	0.99	0.15	0.97	0.23	0.95
Characteristics of Information	<--	Know the author(s)	0.17	0.97	0.36	0.86	0.08	0.99
Characteristics of Information	<--	Know the journals or conferences	0.58	0.67	0.75	0.42	0.64	0.59
Characteristics of Information	<--	The authors are eminent	1.18	0.004	0.11	0.99	1.00	0.008
Information Novelty	<--	New	0.65	0.58	0.38	0.87	0.15	0.97
Information Novelty	<--	Original	0.31	0.89	0.28	0.95	0.23	0.95
Information Novelty	<--	Unique or only sources	0.61	0.61	0.57	0.67	0.49	0.76

Chapter 7: Discussion and Conclusions

7-1. Introduction

Researches have examined many criteria that contribute to relevance judgements (Barry 1994; Bateman 1998; Maglaughlin and Sonnenwald 2002; Schamber and Bateman 1999; Tang and Solomon 2001; Wang and Soergel 1998). However, previous research has not examined the importance of each criterion and the relationship between the criteria. Moreover, the relationship among the latent factors of relevance judgement has not been studied yet. This chapter begins with an overview of the solutions based on the measurement and structural models, which depicted factors influencing relevance judgement, presented in the previous chapter. The chapter will then address the conclusions and implications.

7-2. Research Aims and Objectives

The primary purpose of this study was to identify cognitive criteria or rationale that participants employ when they make relevance judgements. The second purpose of research was to measure the importance of each criterion of cognitive relevance. The present study was designed:

- To get an idea of the total variance accounted for by the relevance criteria and factors (that is, how much of the variation in cognitive relevance judgement can be explain with the aid of the criteria that we have chosen and how much variation remains unexplained).

- To examine which are the most important criteria and which are less important or insignificant.

- To find out how powerful each criterion is after its link with other criteria that have been held constant.

- To examine whether cognitive factors of relevance judgement (aboutness, information quality, information characteristics and novelty) remain stable over time and thus represent constant factors of cognitive relevance judgement.

In this study, a longitudinal design adopted to overcome a weakness in previous relevance criteria studies, which relied on taking a snapshot of a sample of participants at a particular point. The aims of the study were achieved by developing theoretical model by a series of measurement sub-models using structural equation modelling. This technique derived information about the data to fix the parameters in the measurement model (by using one-factor modelling of each factor). Then, the structural model enabled parameters relating to the latent factors to be identified. The model of cognitive relevance judgement comprised sixteen criteria and four factors (Aboutness, Quality of Information, Characteristics of Information, and Information Novelty).

7-3. Findings

The present study addresses two main issues related to cognitive factors of relevance judgement. First, what is the importance of the various cognitive factors of relevance judgement and the relationship among them? And second, do these factors remain stable over time?

To answer the first question, numerous criteria were identified in the literature that influence relevance judgement. An initial pool of forty items was developed. The items were constructed into Likert scale statements. These were then pre-tested and the items were reduced to set of sixteen criteria. These were administrated to a sample of experts in information retrieval research area (See Chapter 5). A principal component analysis with varimax rotation suggested a four-factor solution with Eigenvalues exceeding 1.0, which accounted for 78% of the total variance of the items. Therefore, the 16 criteria were divided into four latent variables (factors). The first factor was Aboutness, the second was Quality of Information, the third was Characteristics of Information and the fourth was Novelty. A conceptual model was then developed to explore the relationship between the criteria and factors of cognitive relevance.

Another purpose of this study was to clarify relationships among responses to 16 criteria of relevance judgement (observed variables). Our approach to testing the relationships of the measurement models was through a statistical technique called Confirmatory Factor Analysis (CFA).

A conceptual model of the three studies was that the dependent endogeneous latent variable, cognitive relevance judgement (Revjug), is influenced by four exogeneous independent latent variables, Aboutness (Abouts), Quality of Information (Quality), Characteristics of Information (Charac) and Information Novelty (Novelty).

7-3-1. Findings of Study 1

Measurement model of study 1 was developed to examine the relationship between the criteria and their latent variables. Table 7-1 shows the path coefficients, R-squared and Error of the measurement model of Study 1.

Table 7-1. Path Coefficients, R-Squared and Error of the Measurement Model of Study 1

	Coefficient	R-squared	Error
Factor / **Criteria**	**Aboutness**		
About my topic (TOPIC1)	0.77	0.60	0.40
Informative (INFORM1)	0.74	0.55	0.45
Generally or specifically suitable (SUITABL1)	0.79	0.62	0.38
Background (BACKGR1)	0.19	0.036	0.96
Good bibliography and links (BIBLIO1)	0.16	0.027	0.97
Controversial (CONTROV1)	0.07	0.005	0.99
Factor / **Criteria**	**Quality of Information**		
Validate viewpoint (VALIDATE1)	0.50	0.25	0.75
Consistent (CONSIST1)	0.35	0.12	0.88
Easy to understand (UNDERST1)	0.48	0.23	0.77
Methods or techniques have described (METHOD1)	0.11	0.013	0.99
Factor / **Criteria**	**Characteristics of Information**		
Know the author(s) (AUTHKNO1)	0.17	0.029	0.97
Know the journals or conferences (JOURKNO1)	0.58	0.33	0.67
The authors are eminent (EMINENT1)	1.18	0.99	0.004
Factor / **Criteria**	**Information Novelty**		
New (NEW1)	0.65	0.42	0.58
Original (ORIGIN1)	0.31	0.11	0.89
Unique or only sources (UNIQUE1)	0.61	0.39	0.61

The relationship between factors and a higher level variable or dependent latent variable (i.e. relevance judgement) is illustrated in Figure 7-1.

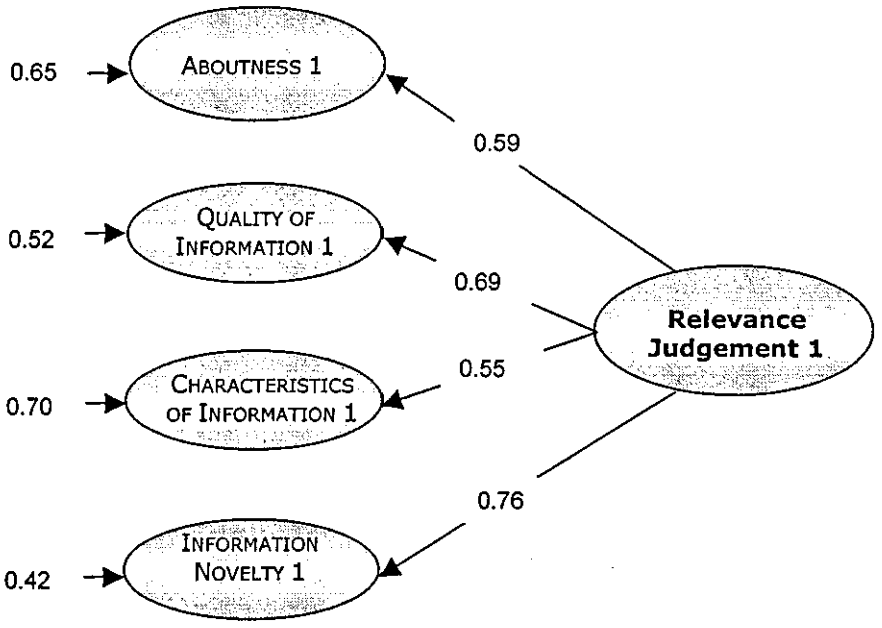


Figure 7-1. The Measurement Model of Relevance Judgement (Study 1)

The measurement model of relevance judgement in study 1 presented in Figure 7-1 indicates that Information Novelty (0.76), Quality of Information (0.69), Aboutness (0.59) and Characteristics of Information (0.55) have strong effects on relevance judgement.

The measurement equations for the measurement model of study 1 are:

Abouts1 = 0.59*Revjug1, Errorvar.= 0.65 , R^2 = 0.35

Quality1 = 0.69*Revjug1, Errorvar.= 0.52 , R^2 = 0.48

Charac1 = 0.55*Revjug1, Errorvar.= 0.70 , R^2 = 0.30

Novelty1 = 0.76*Revjug1, Errorvar.= 0.42 , R^2 = 0.58

The squared multiple correlation, R^2, is displayed for each equation. The R-squared of 0.35 in the Abouts1 (Aboutness of study 1) equation shows that 35% of the variance in relevance judgement is explained by Aboutness. The R-squared of Quality of Information (Quality1), Characteristics of Information (Charac1), and Information Novelty (Novelty1) are 48%, 30 % and 58%, respectively.

7-3-2. Findings of Study 2

Table 7-2 shows the path coefficients, R-squared and Error of the measurement model of Study 2.

Table 7-2. Path Coefficients, R-squared and Error of the Measurement Model of Study 2

	Coefficient	R-squared	Error
Factor / **Criteria**	**Abouts**		
About my topic (TOPIC2)	0.72	0.56	0.44
Informative (INFORM2)	0.78	0.63	0.36
Generally or specifically suitable (SUITABL2)	0.69	0.48	0.52
Background (BACKGR2)	0.34	0.11	0.88
Good bibliography and links (BIBLIO2)	0.13	0.017	0.98
Controversial (CONTROV2)	0.14	0.02	0.98
Factor / **Criteria**	**Quality of Information**		
Validate viewpoint (VALIDAT2)	0.28	0.07	0.93
Consistent (CONSIST2)	0.48	0.23	0.78
Easy to understand (UNDERST2)	0.35	0.12	0.88
Methods or techniques have described (METHOD2)	0.15	0.02	0.97
Factor / **Criteria**	**Characteristics of Information**		
Know the author(s) (AUTHKNO2)	0.36	0.13	0.86
Know the journals or conferences (JOURKNO2)	0.75	0.57	0.42
The authors are eminent (EMINENT2)	0.11	0.012	0.99
Factor / **Criteria**	**Information Novelty**		
New (NEW2)	0.38	0.12	0.87
Original (ORIGIN2)	0.28	0.06	0.95
Unique or only sources (UNIQUE2)	0.57	0.33	0.67

Figure 7-2 presents the relationship between four factors and relevance judgement in study 2.

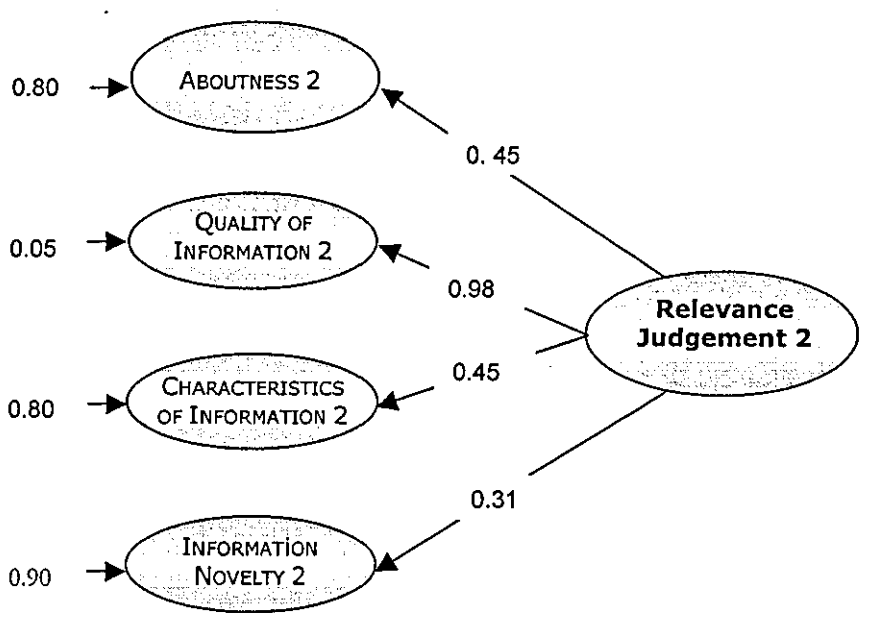


Figure 7-2. The Measurement Model of Relevance Judgement (Study 2)

As can be seen in Figure 7-2, Quality of Information (0.98) is a significant factor of relevance judgement. This is followed by Aboutness (0.45), Characteristics of Information (0.45) that can be interpreted as acceptable, but lower than Quality of Information. Information Novelty (0.31) seems to be less significant in study 2.

The measurement equations for the measurement model of study 2 are:

Abouts2 = 0.45*Revjug2, Errorvar.= 0.80 , R^2 = 0.20

Quality2 = 0.98*Revjug2, Errorvar.= 0.50, R^2 = 0.95

Charac2 = 0.45*Revjug2, Errorvar.= 0.80 , R^2 = 0.20

Novelty2 = 0.31*Revjug2, Errorvar.= 0.90 , R^2 = 0.15

The R-squared of Quality of Information (Quality2) is 0.95 which is significant. the percent of variance explained by Aboutness (Abouts2) (20%), Characteristics of Information (Charac2) (20%) and Information Novelty (Novelty1) (15%) are extremely low in Study 2.

7-3-3. Findings of Study 3

Path coefficients (direct effects), squared multiple correlation, R^2, and errors of the measurement model of Study 3 are reported in Table 7-3.

Table 7-3. Path Coefficients, R-Squared and Error of the Measurement Model of Study 3

	Coefficient	R-squared	Error
Factor / Criteria	Abouts		
About my topic (TOPIC3)	0.62	0.41	0.59
Informative (INFORM3)	0.54	0.32	0.68
Generally or specifically suitable (SUITABL3)	0.76	0.58	0.42
Background (BACKGR3)	0.50	0.25	0.75
Good bibliography and links (BIBLIO3)	0.34	0.12	0.88
Controversial (CONTROV3)	0.21	0.043	0.96
Factor / Criteria	Quality of Information		
Validate viewpoint (VALIDAT3)	0.53	0.28	0.72
Consistent (CONSIST3)	0.59	0.34	0.66
Easy to understand (UNDERST3)	0.44	0.12	0.88
Methods or techniques have described (METHOD3)	0.23	0.054	0.95
Factor / Criteria	Characteristics of Information		
Know the author(s) (AUTHKNO3)	0.08	0.007	0.99
Know the journals or conferences (JOURKNO3)	0.64	0.41	0.59
The authors are eminent (EMINENT3)	1.00	0.99	0.008
Factor / Criteria	Information Novelty		
New (NEW3)	0.15	0.024	0.97
Original (ORIGIN3)	0.23	0.053	0.95
Unique or only sources (UNIQUE3)	0.49	0.24	0.76

Figure 7-3 presents the relationship between four factors and relevance judgement in study 3.

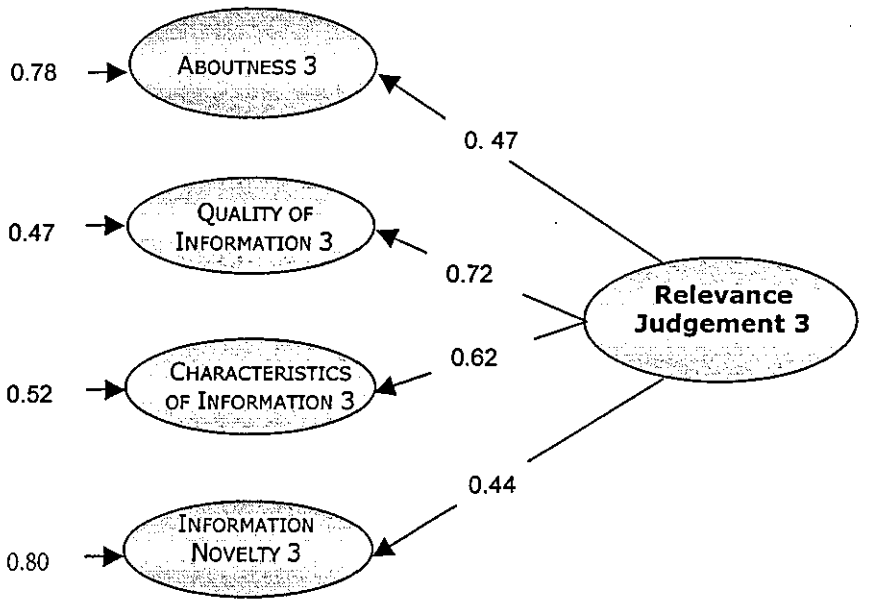


Figure 7-3. The Measurement Model of Relevance Judgement (Study 3)

Figure 7-3 presents the path coefficient of Quality of Information (0.72) that is a highly recognised factor of relevance judgement. This is followed by Characteristics of Information (0.62), Aboutness (0.47) and Information Novelty (0.44) that can be interpreted as good, but lower than Quality of Information.

The measurement equations for the measurement model of study 3 are:

Abouts3 = 0.47*Revjug3, Errorvar.= 0.78 , R^2 = 0.22

Quality3 = 0.72*Revjug3, Errorvar.= 0.47 , R^2 = 0.53

Charac3 = 0.62*Revjug3, Errorvar.= 0.62 , R^2 = 0.38

Novelty3 = 0.44*Revjug3, Errorvar.= 0.80 , R^2 = 0.20

The R-squared of Quality of Information (Quality3) is 0.53 which is significant. The percent of variance explained by Aboutness (Abouts3) (22%), Characteristics of Information (Charac3) (38%) and Information Novelty (Novelty1) (20%) are low in Study 3.

7-3-4. Findings of the Longitudinal Study

The second purpose of the study was to assess the stability of the cognitive factors of relevance judgement over time.

The model in Study 1, Study 2 and Study 3 comprise four factors of relevance judgement called Aboutness, Quality of Information, Characteristics of Information and Information Novelty. The measurement model is applied at each time point. The objective of the panel model was to answer two questions: Has the level of relevance judgement increased or decreased over time? Has the variance of relevance judgement increased or decreased over time?

A conceptual model for relevance judgement is shown in Figure 7-4.

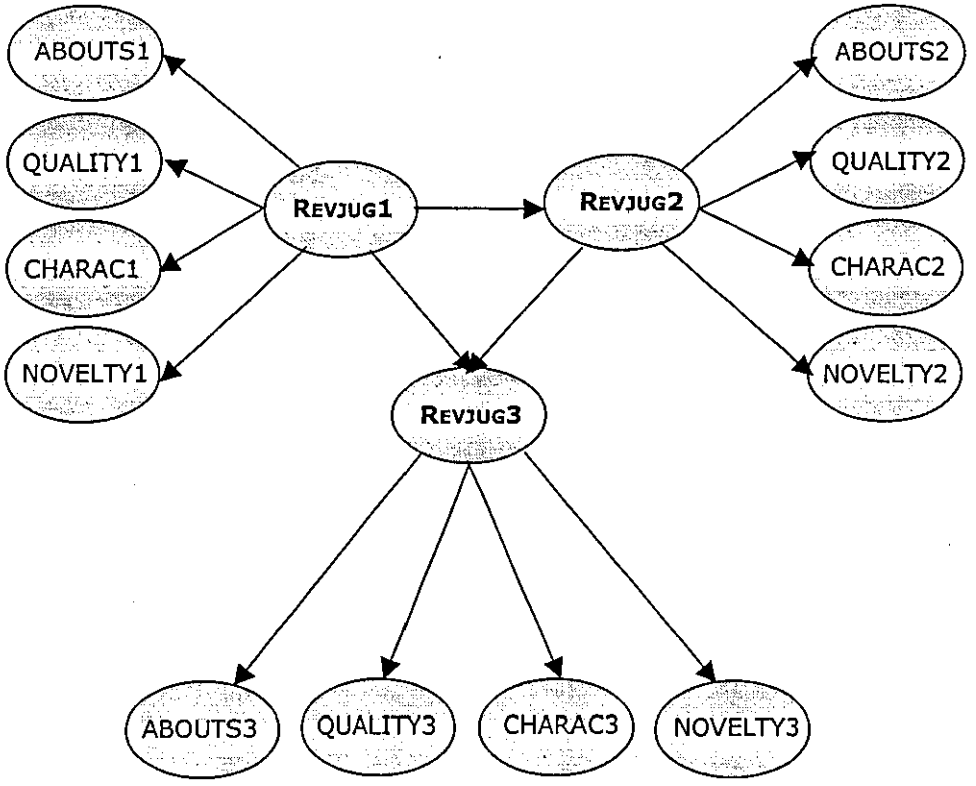


Figure 7-4. Conceptual Panel Model for Relevance Judgement

The model involves a structural model in the middle of the diagram in which Relevance Judgement at time 2 is predicted by Relevance Judgement at time 1 and Relevance Judgement at time 3 is predicted by Relevance Judgement at time 1 and time 2.

In evaluating the structural part of the model, we have focused on the relationships between the three dependent latent variables (i.e. Relevance Judgement 1, 2 and 3). Figure 7-5 presents the results of the analysis.

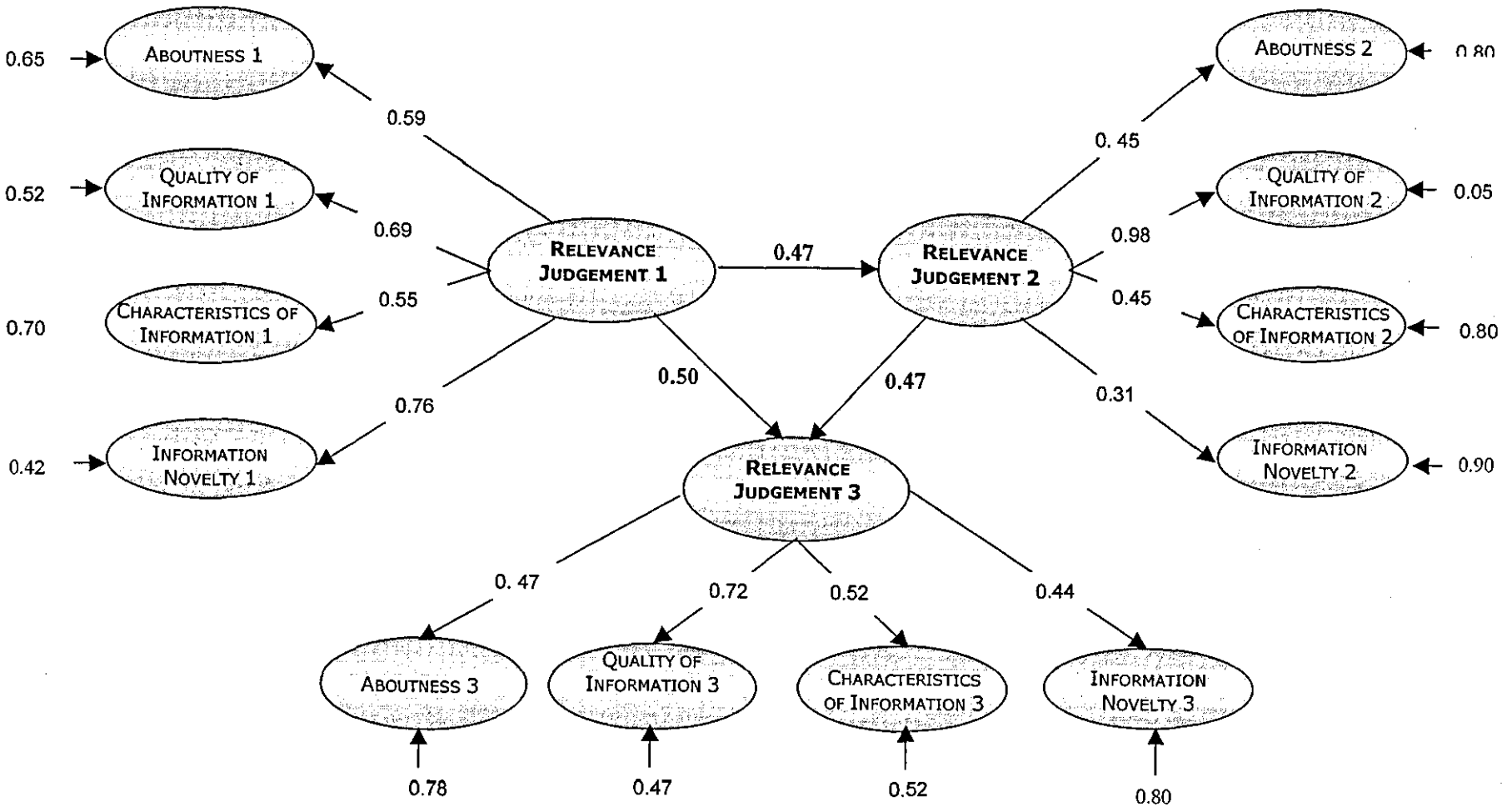


Figure 7-5. Structural Model for Relevance Judgement

The structural equations of the model show that the stability coefficients are statistically significant, suggesting that relevance judgement at time 2 can be predicted from relevance judgement at time 1, to some extent. But the prediction is not very accurate. This does not necessarily mean that the factors are not stable over time. However, the sample size was not large enough to make them significant. Other variables outside of the model may be needed to make this prediction more accurate.

To answer the second question, whether the factors remain stable across time, we looked at the stability loadings in the longitudinal structural model. The results support our hypothesis regarding the stability of the factors of relevance judgement. All these stability loadings, regardless of their strength, are statistically significant.

The results of this study enable reconstruction of previous theories on criteria for relevance judgement. Previous investigations of criteria have concentrated on eliciting relevance criteria from users of IRSs. Previous researches have compiled a number of criteria. However, it is difficult to reach a consensus in classifying these criteria because many of them have multiple meanings, in particular criteria of cognitive relevance. In addition, these criteria are situationally dependent. It is generally assumed in the literature that these criteria comprise all factors of relevance and hence are applicable to relevance judgements in general. This research concentrated on cognitive relevance and its associated factors.

The results of this study are not in conflict with, but complementary to, the results that were previously obtained by other researchers (e.g. Cool et al. 1993; Barry 1994; Schamber and Bateman 1999; Bateman 1998; Wang and Soergel 1998; Tang and Solomon 2001; Maglaughlin and Sonnenwald 2002). In this study, a statistical model of cognitive relevance has been developed with four factors: (1) Aboutness; (2) Quality of Information; (3) Characteristics of Information; and (4) Information Novelty.

The measurement results of this study support some arguments of previous research such as finding of Wang and Soergel (1998). Based on that model, topicality, orientation, and quality are the three most frequently used criteria in evaluating bibliographic surrogates. Wang and Soergel (1998) stated that novelty, discipline, and recency are also critical in relevance assessments.

The results of SEM, in this study to some extent support the findings of Schamber and Bateman (1999). They conducted a series of tests and generated five major criteria groups with a total of 23 criteria. The first group is Aboutness, within this group the criteria include "about my topic," "appropriate," "pertinent," "relevant," and "usable." The second group is Currency, which contains elements of "current," "recent," and "up-to-date." The third group is Availability, which is made up by "available," "accessible," "convenient," and "easy to get." The fourth group Clarity consists of factors such as "clear," "readable," and "understandable." The final group Credibility is represented by "credible," "expert," "I know the publication," "I know the source," "prominent," "reliable," "reputable," and "well-written." Bateman (1998) presented a model that consists of six dimensions: Information Quality, Information Credibility, Information Completeness, Information Topicality, Information Currency, and Information Availability. The results of this study are not completely consistent with Bateman (1998).

7-4. Conclusion

The purpose of this study was to investigate the factors that users employ when they make relevance judgement. This study attempted to quantitatively measure the cognitive factors affect relevance judgement. This section starts with a statement of conclusions drawn from the longitudinal study, which was summarised in Chapter 6 and previous sections of this chapter. The conceptual framework assumed that Aboutness, Quality of Information, Characteristics of Information and Information Novelty are four independent latent factors that would represent relevance judgement. The results of structural equation modelling indicate that a four-factor structure is necessary to measure cognitive relevance judgement. The results also imply that the hypothesised four-factor model of the relevance judgement is reasonable. Thus, the hypothesis put forth in the conceptual framework was supported in this study. The study also examined stability of the cognitive factors of relevance judgement longitudinally in three occasions. The results showed that they are stable over time.

A comparison of the path coefficients of Aboutness, Quality of Information, Characteristics of Information and Information Novelty showed that the contribution of all factors in three studies are important and, in fact, are statistically significant. The path coefficients obtained through SEM ranged from 0.45 to 0.78 for Aboutness, from 0.47 to 0.98 for Quality of Information, from 0.45 to 0.55 for Characteristics of Information and from 0.31 to 0.76 for Information Novelty. This variability and range of values can be interpreted as the impact of situational relevance. The importance of each factor varied depending on the participant's situation. This situation comprised their topic and discipline, their information problem, their knowledge of the topic.

The research reported here supports the development of Interactive Information Retrieval Systems (IRSs) and an associated theory of relevance. This research would enhance the concept of relevance judgement to users in most situations. Moreover, this can help researchers to understand more about the situational variation in relevance judgement. Like previous studies this research highlights the complexity of the cognitive relevance. In fact, the identification of a mathematical model of relevance judgement can be useful for future studies that seek to better understand the effects of relevance criteria in various information retrieval situations.

7-5. Limitations of the study

The limitations of the study are considered from two perspectives, those specifically concerning the methodological technique of SEM, and those of a more practical nature. The methodological limitations are identified as relating to the modelling approach and, the impact of the sample size on the results. The practical limitations include capturing the same students three times during a year. These limitations and the ways in which they were addressed in this study are discussed in detail below.

7-5-1. Methodological Limitations

Modelling approach

The modelling analysis conducted was based on a singular model at a time with data from one sample. No attempt was made to divide the sample into groups, such as science and engineering PhD students, by year of their study (first year, second year and final year student) or by gender or other attributes. The relatively small sample sizes prevented these treatments. An improvement in the approach for this study, would have been the use of two or more independent samples, for example, science PhD students and engineering PhD students to a single model.

This study used one particular approach to SEM, namely a confirmatory approach in which the posited theoretical models for Study 1, Study 2 and Study 3 were systematically tested through a series of measurement models which included sixteen criteria and four factors of relevance judgement. However, it is recognised that other approaches could have been used. For example, another approach is the fitting of alternative models, to the same data, by placing relevance criteria on the different factors and developing an alternative model.

The role of sample size on the results obtained

The relationship between sample size and the probability of obtaining significant results has been discussed in Chapter 3. The samples size in the three studies in this investigation just give an idea for using structural equation modelling (SEM). This sample size provided parameters in the measurement models and consequently the full models were fixed. This small sample size decreased the statistical power (i.e., the probability of obtaining significant results) and the numbers of free parameters to be estimated. In this research, the limitation of sample size reduced the number of free

parameters to be estimated and the researcher could not develop and test a relatively complex model.

7-5-2. Practical Limitations

Any investigation involving longitudinal design will carry with it practical limitations. The longitudinal nature of Study 1, 2 and 3 necessitated the "capture" of the PhD students three times during a year. Fortunately, this requirement did not pose major problems.

In this study, in the case that respondents dealt with more than one document, measurement was based on overall impression of their relevance judgements. However, the survey statements are more akin to statements about the relevance of a single document rather than a set of documents. If time and resources had permitted, the survey questions could have been applied after each single document was retrieved, i.e. the user does not judge the relevant documents overall, but instead he or she assesses each single document. However, because of lack of time and because of the risk of participant fatigue an overall assessment of all documents retrieved was requested instead. In addition, dynamic notion of relevance and temporal nature of relevance assessments were not taken into account. Since users' perceptions and objectives change during interaction, their information needs change during a search session and therefor relevance assessments will change during the progress of a search. Both these issues may have affected the criteria used by participants to measure relevance, and might have influenced the results of this study. Other research being carried out in this department[1] is attempting to assess changes in searchers' relevance assessment over time.

7-6. Implications

1 J. Back, unpublished results.

This section focuses on the implications of the results from several points of view. Implications for theoretical development of the cognitive relevance judgement and implications for the design of interactive information retrieval systems are two major implications.

7-6-1. Implications for Theoretical Development

This research has found that users apply Aboutness and Quality of Information more than other factors in their relevance judgements.

Many studies (Schamber, Eisenberg and Nilan 1990; Boyce 1982; Barry 1994; Wang 1998; Bateman 1998) have shown that topicality is the most frequently used factor for relevance judgement. Moreover, they have established that relevance is a multidimensional construct and users employ multiple criteria for their relevance judgements. This is the first time that a numeric value has been given to the cognitive relevance criteria with a multivariate technique (SEM). Like Barry and Schamber (1999) and Bateman (1998) this research confirmed that users not only select the documents that are related to their topics but also they consider the quality of information in their judgements.

7-6-2. Implications for Interactive IRS

The findings of this research have implications for developing an information retrieval system that better satisfies users' needs. The findings suggest the following features are needed in such a system:

- An interactive IRS should provide an interface that allows users to specify the criteria that they view as important for their selection purpose. This consideration increases consideration may increase users' satisfaction.

- Since the system is interactive, it should allow users to express their satisfaction with retrieved documents. Using the users' feedback, the system would retrieve documents similar to those ranked high satisfaction.

- With an interactive IRS, users should have control of what pieces of information they want in the document summary. Upon receiving this input, the system would create a document summary based on users' needs.

7-6-3. Directions for Further Research

Directions for further research can be considered as follows:

On the basis of the general approach in this investigation, it is recommended that further research be conducted in order to:

- re-test and re-validate the models developed in Studies 1, 2 and 3 on a larger sample size, and different populations and across different contexts;

- re-test the models in Studies 1, 2 and 3 is a way that addresses the limitations described under Methodological limitations in this Chapter;

- test with sub-groups, such as science and engineering PhD students or first year, second year and final year students, in order to confirm generalisability, replicability and validity of the measurement models and the structural model of relevance judgement.

Further research is recommended to confirm the stability over time of the cognitive factors of relevance judgement identified with a longer interval and the lack of influence of background and individual attributes.

References

Aron, E. N. & A. Aron, 1999. *Statistics for Psychology.* N.J. ; London: Prentice Hall.

Atkins, H., 1999. The ISI Web of Science - Links and Electronic Journals. *D-Lib Magazine,* 5(9), (URL: http://www.dlib.org/dlib/september99/atkins/09atkins.html).

Baker, T. L., 1988. *Doing Social Research.* New York; London: McGraw-Hill.

Barry, C. L., 1993. A Preliminary Examination of Clues to Relevance Criteria within Document Representations. *Proceedings of American Society for Information Science,* 81-86.

Barry, C. L., 1994. User Defined Relevance Criteria: An Exploratory Study. *Journal of the American Society for Information Science,* **45**(3), 149-159.

Barry, C. L. & L. Schamber, 1995. User-Defined Relevance Criteria: A Comparison of Two Studies. In: *Proceedings of the American Society for Information Science.* Information Today, Medford, pp. 103-111.

Barry, C. L. & L. Schamber, 1998. Users' Criteria for Relevance Evaluation: A Cross-Situational Comparison. *Information Processing and Management,* **34**(2/3), 219-236.

Bateman, J., 1998. Changes in Relevance Criteria: A Longitudinal Study. In: *Proceedings of the 61st Annual Meeting of the American Society for Information Science,* Vol. 35, pp. 23-32.

Belkin, N. & A. Vickery, 1985. *Interaction in Information Systems: A Review of Research from Document Retrieval to Knowledge-Based Systems.* London: The British Library.

Belkin, N. J. *et al.*, 1996. Cases, Scripts, and Information-Seeking Strategies: On the Design of Interactive Information Retrieval Systems. *Expert Systems with Applications,* **9**(3), 379-396.

Blalock, H. M. & A. B. Blalock, 1982. *Methodology in Social Research.* New York: McGraw-Hill.

Borgman, C. L., 1986. The User's Mental Model of an Information Retrieval System: An Experiment on a Prototype Online Catalog. *International Journal of Man-Machine Studies,* **24**(1), 47-64.

Borlund, P., 2000. Experimental Components for the Evaluation of Interactive Information Retrieval Systems. *Journal of Documentation,* **56**(1), 71-90.

Boyce, B., 1982. Beyond Topicality: A Two Stage View of Relevance and the Retrieval Process. *Information Processing and Management,* **18**(3), 105-109.

Bruce, H. W., 1994. A Cognitive View of the Situational Dynamism of User-Centered Relevance Estimation. *Journal of the American Society for Information Science,* **45**(5), 142-148.

Bryant, F. B., 2000. Assessing the Validity of Measurement. *In:* L. G. Grimm & P. R. Yarnold (Eds.), *Reading and Understanding More Multivariate Statistics.* Washington, DC ; London: American Psychological Association, pp. 99-146.

Burch, J. G. & F. R. Strater, 1974. *Information Systems: Theory and Practice.* Santa Barbara: Hamilton.

Byrne, B. M., 1998. *Structural Equation Modeling with LISREL, PRELIS, and SIMPLIS : Basic Concepts, Applications, and Programming.* N.J. ; London: Lawrence Erlbaum.

Cassell, C. & G. Symon, 1994. *Qualitative Methods in Organizational Research: A Practical Guide.* London: Sage.

Chowdhury, G. G., 1999. *Introduction to Modern Information Retrieval.* London: Library Association.

Churchill, G. A., 1995. *Marketing Research: Methodological Foundations.* Fort Worth, TX ; London: Dryden Press.

Cleverdon, C. W., J. Mills & E. M. Keen, 1966. *Factors Determining the Performance of Indexing Systems (Volume 1: Design).* Cranfield: College of aeronautics.

Coakes, S. J. & L. G. Steed, 2001. *SPSS : Analysis without Anguish : Version 10.0 for Windows.* Brisbane ; Chichester: John Wiley & Sons.

Cool, C. *et al.*, 1993. Characteristics of Text Affecting Relevance Judgements. In: *National Online Meeting,* Vol. 14, Learned Information (Europe), pp. 77.

Coolican, H., 1994. *Research Methods and Statistics in Psychology.* London: Hodder & Stoughton.

Cooper, W. S., 1971. A Definition of Relevance for Information Retrieval. *Information Storage and Retrieval,* 7(1), 19-37.

Cosijn, E. & P. Ingwersen, 2000. Dimensions of Relevance. *Information Processing and Management,* 36(7), 533-550.

Creswell, J. W. & G. A. Miller, 1997. Research Methodologies and the Doctoral Process. *In:* L. F. Goodchild, *et al.* (Eds.), *Rethinking the Dissertation Process: Tackling Personal and Institutional Obstacles.* Jossey-Bass, pp. 33-46.

Daniels, P. J., 1986. Progress in Documentation: Cognitive Models in Information Retrieval - an Evaluative Review. *Journal of Documentation,* 42(4), 272-304.

DeVellis, R. F., 1991. *Scale Development: Theory and Applications.* Newbury Park; London: Sage Publications.

Diamantopoulos, A. & J. Siguaw, 2000. *Introducing LISREL: A Guide for the Uninitiated.* London: Sage Publications.

Draper, S., 1998. Mizzaro's Framework for Relevance. Http://www.psy.gla.ac.uk/~steve/stefano.html (24/05/2000).

Ellis, D., 1992. The Physical and Cognitive Paradigms in Information Retrieval Research. *Journal of Documentation,* 48(1), 45-64.

Ellis, D., 1996. The Dilemma of Measurement in Information Retrieval Research. *Journal of the American Society for Information Science,* 47(1), 23-36.

Ferguson, G. A. & Y. Takane, 1989. *Statistical Analysis in Psychology and Education.* New York; Tokyo: McGraw-Hill.

Fink, A., 2000. *How to Sample in Surveys.* Thousand Oaks, Calif.; London: Sage Publications.

Froelich, T. J., 1994. Relevance Reconsidered: Toward an Agenda for the 21st Century. *Journal of the American Society for Information Science,* **45**(3), 124-134.

Goodwin, C. J., 2002. *Research in Psychology: Methods and Design.* New York: John Wiley & Sons.

Greenberg, M. A. C. & B. L. Crissey, 1976. *Models in the Policy Process.* New York: Russell Sage Foundation. In: Widman, Lawrence E.; Loparo, Kenneth A. and Nielsen, Norman R. (1989). Artificial Intelligence, Simulation and Modeling. New York: John Wiley & Sons.

Hair, J. F. *et al.*, 1995. *Multivariate Data Analysis with Readings.* New Jersey: Prentice Hall.

Hand, D. & M. Crowder, 1996. *Practical Longitudinal Analysis.* New York: Chapman and Hall.

Harman, D., 1995. The Second Text Retrieval Conference (TREC-2). *Information Processing and Management,* **31**(3), 271-289.

Harter, S. P., 1996. Variations in Relevance Assessments and the Measurment of Retrieval Effectiveness. *Journal of the American Society for Information Science,* **47**(1), 37-49.

Hayduk, L. A., 1987. *Structural Equation Modeling with LISREL: Essentials and Advances.* Baltimore: Johns Hopkins University Press.

Hert, C. A., 1997. *Understanding Information Retrieval Interactions: Theoretical and Practical Implications.* Greenwich: Ablex.

Howard, D. L., 1994. Pertinence as Reflected in Personal Constructs. *Journal of the American Society for Information Science,* **45**(3), 172-185.

Howitt, D. & D. Cramer, 2000. *An Introduction to Statistics in Psychology: A Complete Guide for Students.* Harlow: Pearson Education.

Ingwersen, P., 1996. Cognitive Perspectives of Information Retrieval Interactions: Elements of Cognitive IR Theory. *Journal of Documentation,* **52**(1), 3-50.

Ingwersen, P., 2001. Cognitive Information Retrieval. *Annual Review of Information Science and Technology,* **34**(1999), 3-52.

Joreskog, K. G. & I. Moustaki, 2001. Factor Analysis of Ordinal Variables: A Comparison of Three Approaches. *Multivariate Behavioral Research,* **36**(3), 347-388.

Kowalski, G., 1997. *Information Retrieval Systems: Theory and Implementation.* Boston: Kluwer Academic Publishing.

Kumar, R., 1999. *Research Methodology: A Step by Step Guide for Beginners.* London; Thousand Oaks, Calif.: Sage.

Lancaster, F. W., 1968. *Information Retrieval Systems: Characteristics, Testing and Evaluation.* New York: John Wiley & Sons.

Lewis-Beck, M. S., 1994. *Basic Measurement.* London: Toppan.

Li, F. *et al.*, 1998. Analyzing Measurement Models of Latent Variables through Multilevel

Confirmatory Factor Analysis and Hierarchical Linear Modeling Approaches. *Structural Equation Modeling,* 5(3), 294-306.

Likert, R., 1932. A Technique for the Measurement of Attitudes. *In:* G. Summers (Ed.) *Attitude Measurement.* Chicago, IL: Rand McNally & Company, pp. 149-158.

Maglaughlin, K. L. & D. H. Sonnenwald, 2002. User Perspectives on Relevance Criteria: A Comparison among Relevant, Partially Relevant, and Not-Relevant Judgments. *Journal of the American Society for Information Science and Technology,* 53(5), 327-342.

Mizzaro, S., 1997. Relevance: The Whole History. *Journal of the American Society for Information Science,* 48(8), 810-832.

Mizzaro, S., 1998. How Many Relevances in Information Retrieval? *Interacting with Computers,* 10(3), 305-322.

Perspectives of Cognitive Science, 1981. Norman, D. (Ed.) Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Oppenheim, A. N., 1992. *Questionnaire Design, Interviewing and Attitude Measurement.* London: Pinter.

Park, T. K., 1993. The Nature of Relevance in Information Retrieval: An Empirical Study. *Library Quarterly,* 63(3), 318-351.

Park, H., 1995. *Inferential Relevance and Its Implications to Inferential Information Retrieval.* PhD Dissertation. Rutgers University.

Pedhazur, E. J. & F. N. Kerlinger, 1982. *Multiple Regression in Behavioral Research: Explanation and Prediction.* New York: Holt Rinehart and Winston.

Pedhazur, E. J. & S. L. Pedhazur, 1991. *Measurement, Design, and Analysis: An Integrated Approach.* Hillsdale, N.J.: Lawrence Erlbaum Associates.

Pors, N. O., 2000. Information Retrieval, Experimental Model and Statistical Analysis. *Journal of Documentation,* 56(1), 55-70.

Powell, R. R., 1997. *Basic Research Methods for Librarians.* London: Ablex.

Quade, E. S., 1985. Modelling Techniques. *In:* L. E. Widman, K. A. Loparo & N. R. Nielsen (Eds.), *Artificial Intelligence, Simulation and Modeling.* New York: John Wiley & Sons, pp.

Regazzi, J. J., 1988. Performance Measures for Information Retrieval Systems -- an Experimental Approach. *Journal of the American Society for Information Science,* 39(4), 235-251.

Rothenberg, J., 1986. Object-Oriented Simulation: Where Do We Go from Here? *In:* L. E. Widman, K. A. Loparo & N. R. Nielsen (Eds.), *Artificial Intelligence, Simulation and Modelling.* New York: John Wiley & Sons, pp. 464-496.

Salton, G. & M. J. McGill, 1983. *Introduction to Modern Information Retrieval Systems.* Auckland: McGraw-Hill.

Salton, G., 1992. The State of Retrieval System Evaluation. *Information Processing and Management,*

28(4), 441-449.

Saracevic, T., 1975. Relevance: A Review of and a Framework for the Thinking on the Notion in Information Science. *Journal of the American Society for Information Science,* **26**(3), 321-343.

Saracevic, T. *et al.*, 1988. A Study of Information Seeking and Retrieving. I. Background and Methodology. *Journal of the American Society for Information Science,* **39**(3), 162-176.

Saracevic, T., 1996. Relevance Reconsidered. *In:* P. Ingwersen & N. Pors (Eds.), *Information Science: Integration in Perspectives.* Copenhagen: The Royal School of Librarianship, pp. 201-218.

Schamber, L., M. B. Eisenberg & M. S. Nilan, 1990. A Re-examination of Relevance: Toward a Dynamic Situational Definition. *Information Processing and Management,* **26**(6), 785-776.

Schamber, L., 1991. Users' Criteria for Evaluation in a Multimedia Environment. In: *Proceedings of the American Society for Information Science (ASIS) 54th Annual Meeting,* Vol. 28, pp. 126-133.

Schamber, L., 1994. Relevance and Information Behavior. *Annual Review of Information Science and Technology,* **29**(7), 3-48.

Schamber, L. & J. Bateman, 1996. User Criteria in Relevance Evaluation: Toward Development of a Measurement Scale. In: *Proceeding of the 59th Annual Meeting of the American Society for Information Science,* Vol. 33, pp. 218-225.

Schamber, L. & J. Bateman, 1999. Relevance Criteria Uses and Importance: Progress in Development of a Measurement Scale. In: *Proceedings of the 62nd Annual Meeting of the American Society for Information Science,* Vol. 36, pp. 381-389.

Schuldt, B. A. & J. W. Totten, 1994. Electronic Mail Vs. Mail Survey Response Rates. *Marketing Research,* **6**(1), 36-39.

Schumacker, R. E. & R. G. Lomax, 1996. *A Beginner's Guide to Structural Equation Modeling.* Mahwah, N.J.: Lawrence Erlbaum Associates.

Siegel, A. F. & C. J. Morgan, 1996. *Statistics and Data Analysis : An Introduction.* Chichester: John Wiley & Sons.

Smeaton, A., 1992. Progress in the Application of Natural Language Processing to Information Retrieval Tasks. *The Computer Journal,* **35**(3), 268-278.

Spector, P. E., 1992. *Summated Rating Scales.* Thousand Oaks, CA: Sage Publications.

Sperber, D. & D. Wilson, 1995. *Relevance: Communication and Cognition.* Oxford: Blackwell.

Spink, A., H. Greisdorf & J. Bateman, 1998. From Highly Relevant to Not Relevant: Examining Different Regions of Relevance. *Information Processing and Management,* **34**(5), 599-621.

Spink, A., 1999. Towards a Theoretical Framework for Information Retrieval in an Information Seeking Context. *In:* T. D. Wilson & D. K. Allen (Eds.), *Exploring the Contexts of Information Behaviour.* Sheffield: Taylor Graham, pp. 21-34.

Spink, A. & H. Greisdorf, 2001. Regions and Levels: Measuring and Mapping Users' Relevance Judgments. *Journal of the American Society for Information Science and Technology,* **52**(2), 161-173.

Su, L. T., 1992. Evaluation Measures for Interactive Information Retrieval. *Information Processing and Management,* **28**(4), 503-516.

Tague-Sutcliffe, J. M., 1996. Some Perspectives on the Evaluation of Information Retrieval Systems. *Journal of the American Society for Information Science,* **47**(1), 1-3.

Tang, R. & P. Solomon, 1998. Toward an Understanding of the Dynamics of Relevance Judgment: An Analysis of One Person's Search Behavior. *Information Processing and Management,* **34**(2/3), 237-256.

Tang, R., W. M. Shaw & J. L. Vevea, 1999. Towards the Identification of the Optimal Number of Relevance Categories. *Journal of the American Society for Information Science,* **50**(3), 254-264.

Tang, R. & P. Solomon, 2001. Use of Relevance Criteria across Stages of Document Evaluation: On the Complementarity of Experimental and Naturalistic Studies. *Journal of the American Society for Information Science and Technology,* **52**(8), 676-685.

Thurstone, L. L. & E. J. Chave, 1929. *The Measurement of Attitude: A Psychophysical Method and Some Experiments with a Scale for Measuring Attitude toward the Church.* Chicago, Ill.: The University of Chicago Press.

Torgerson, W. S., 1958. *Theory and Methods of Scaling.* New York: John Wiley & Sons.

Tse, A. C. B., 1998. Comparing the Response Rate, Response Speed and Response Quality of Two Methods of Sending Questionnaires: E-mail Vs. Mail. *Journal of the Market Research Society,* **40**(4), 353-361.

Vakkari, P., R. Savolainen & B. Dervin, 1996. Information Seeking in Context. In: *Proceeding of the international conference on research in information needs, seeking and use in different contexts, Tampere, Finland*Taylor Graham, London

Van Rijsbergen, C. J., 1979. *Information Retrieval.* London: Butterworths.

Vaughan, L. Q. & J. M. Tague-Sutcliffe, 1997. Measuring the Impact of Information on Development: A Lisrel-Based Study of Small Businesses in Shanghai. *Journal of the American Society for Information Science,* **48**(10), 917-931.

Vaughan, L. Q., 1999. The Contribution of Information to Business Success: A LISREL Model Analysis of Manufacturers in Shanghai. *Information Processing and Management,* **35**(2), 193-208.

Vaughan, L. Q., 2001. *Statistical Methods for the Information Professional : A Practical, Painless Approach to Understanding, Using, and Interpreting Statistics.* Medford, N.J.: Information Today.

Voorhees, E. M., 2000. Variation in Relevance Judgements and the Measurement of Retrieval Effectiveness. *Information Processing and Management,* **36**(6), 697-716.

Walczuch, R. M. & K. Hofmaier, 1999. Measuring Customer Satisfaction on the Internet. http://www-edocs.unimaas.nl/files/rm00051.pdf (12/10/2002).

Walker, G. & J. Janes, 1999. *Online Retrieval: A Dialogue of Theory and Practice.* Englewood: Libraries Unlimited.

Wang, P. & D. Soergel, 1998. A Cognitive Model of Document Use During a Research Project. Study I. Document Selection. *Journal of the American Society for Information Science,* **49**(2), 115-133.

Wang, P. & M. D. White, 1999. A Cognitive Model of Document Use During a Research Project. Study Ll. Decisions at the Reading and Citing Stages. *Journal of the American Society for Information Science,* **50**(2), 98-114.

Wang, P., 2001. Methodologies and Methods for Used Behavioral Research. *Annual Review of Information Science and Technology,* **34**(1999), 53-99.

Widman, L. E., K. A. Loparo & N. R. Nielsen, 1989. *Artificial Intelligence, Simulation and Modelling.* New York: John Wiley & Sons.

Wilson, T. D., 1999. Models in Information Behavior Research. *Journal of Documentation,* **55**(3), 249-270.

Zeller, R. A. & E. G. Carmines, 1980. *Measurement in the Social Sciences: The Link between Theory and Data.* Cambridge ; New York: Cambridge University.

Appendices

Appendix 1: Cover Letter to Invite A Postgraduate Student for Participating in the Pilot Study

Dear .....

I am a research student in the Department of Information Science here. I am conducting a research study under the supervision of Professor Charles Oppenheim into the Evaluation of Information Retrieval Systems Effectiveness. In order to carry out this research, I would like to invite you to help me in my research by undertaking for me to a search of the ISI Web of Science and then filling out a questionnaire. This has number of statements about what influenced your decisions to select items as relevant during the search session. You will be asked to tick boxes to confirm whether you agree or not with those statements. This exercise will not take much of your time (about 45 minutes) and will be fun to do. Your contribution will help this research, and the results can help you and others to improve their searches in the future. If you are interested in helping me and have the time, please contact me by email and we will agree an appointment.

I would like to thank you in advance for your co-operation.

Amir Ghaebi

PhD Student
Department of Information Science
Loughborough University
A.Ghaebi@lboro.ac.uk

Appendix 2: Questionnaire for Evaluation of Information Retrieval Systems Effectiveness

Dear Colleague,

The purpose of this experiment is to evaluate Interactive Information Retrieval Systems, particularly Science Citation Index on Web of Science. The results will be useful to both research academics and information systems providers. As a researcher and user of those systems, your views are very important to this investigation. It would be appreciated if answer the following questions.

Yours sincerely,

Amir Ghaebi

PhD Student
Department of Information Science
Loughborough University
A.Ghaebi@lboro.ac.uk

1. How many different search strategies did you use in the session?
2. How many articles did you get as search results in your final search?
3. How many articles were completely related to your search topic?
4. How many articles were partially related to your search topic?
5. How many articles were not relevant to your search topic?
6. How much time did you spend for searching?

The following statements refer to your thought or feeling about the search results. If you dealt with more than one article, provide your overall impression of how you had relevance assessment.

Please rate how strongly you agree or disagree with the following statements and tick one checkbox for each statement.

	Strongly Agree	Agree	Unsure	Disagree	Strongly Disagree
7. The articles returned were **about my topic?**	☐	☐	☐	☐	☐

I selected a number of the articles as relevant because:

8. They were **unique or only sources.**	☐	☐	☐	☐	☐
9. They contained **original** ideas or results.	☐	☐	☐	☐	☐
10. They were **new** to me.	☐	☐	☐	☐	☐
11. They were **well known** to me.	☐	☐	☐	☐	☐
12. They were very **Informative.**	☐	☐	☐	☐	☐
13. They were **well written.**	☐	☐	☐	☐	☐
14. They were **credible.**	☐	☐	☐	☐	☐
15. They were **accurate.**	☐	☐	☐	☐	☐
16. They were **understandable.**	☐	☐	☐	☐	☐
17. They were **consistent.**	☐	☐	☐	☐	☐
18. They were **focused.**	☐	☐	☐	☐	☐
19. Information in them was **well presented.**	☐	☐	☐	☐	☐
20. The articles were **comprehensive.**	☐	☐	☐	☐	☐
21. They were **generally/specifically suitable** for my needs.	☐	☐	☐	☐	☐
22. They were very **detailed.**	☐	☐	☐	☐	☐
23. They were **introductory.**	☐	☐	☐	☐	☐
24. The **overviews** were useful.	☐	☐	☐	☐	☐
25. The **journals/conferences** that they appeared were **prominent.**	☐	☐	☐	☐	☐
26. I know their **authors** personally.	☐	☐	☐	☐	☐
27. I know the **journals/conferences.**	☐	☐	☐	☐	☐
28. The journals/conferences **sources** are **reputable.**	☐	☐	☐	☐	☐
29. The **authors** are **eminent.**	☐	☐	☐	☐	☐
30. They described **methods/techniques.**	☐	☐	☐	☐	☐
31. They had a **research approach.**	☐	☐	☐	☐	☐
32. They had a **statistical approach.**	☐	☐	☐	☐	☐
33. They were **controversial.**	☐	☐	☐	☐	☐
34. They provided **proof.**	☐	☐	☐	☐	☐

35. They provided **a good bibliography and links.** ❒ ❒ ❒ ❒ ❒

36. They provided **background or history.** ❒ ❒ ❒ ❒ ❒

37. **I like** them. ❒ ❒ ❒ ❒ ❒

38. They **validate my viewpoint.** ❒ ❒ ❒ ❒ ❒

39. They were **interesting.** ❒ ❒ ❒ ❒ ❒

40. They were **enjoyable** to read. ❒ ❒ ❒ ❒ ❒

41. If there were other criteria that caused you selected the articles as relevant, please write it: .....

Finally, please could you provide some information about yourself and your research.

Completed by: Gender: Male❑ Female❑ Age:

Department: Year of study: 1st ❑ 2nd ❑ 3rd or above❑

Research Topic:

Tel: Email: Date:

Thank you very much for helping me.

Appendix 3: Invitation Letter to the Experts in Information Retrieval

Dear .....

I am a PhD student in the Department of Information Science at Loughborough University. I am conducting a research study under the supervision of Professor Charles Oppenheim into the Evaluation of Information Retrieval Systems Effectiveness. The emphasis of this study is on measuring cognitive factors that contribute to relevance assessment. In order to carry out this research, I have developed a measurement scale. Professor Oppenheim recommends that I contact you as one of the elite researcher in this field. This email is being sent to just 10 such elite researcher in the UK. I would like to invite you to help me in my research by rating a number of statements about cognitive elements of relevance assessment. If you are interested in helping me and have the time, please rate the following statements and send your responses back to me by replying this email.

I would like to thank you in advance for your co-operation.

Please rate each statement on scale of 1 to 11. A rating of 1 means that you rate the criterion as extremely unimportant to a relevance judgement, and 11 means you rate the criterion as the most important criterion in a relevance judgement. Please put your response directly after the indicated arrow (RATE>) for each statement.

User selects bibliographic records as relevant during a search session if:

1. They are about the user's topic.

RATE>

2. They are unique or the only sources.

RATE>

3. They contain original ideas or results.

RATE>

4. They are new to the user.

RATE>

5. They are very informative.

RATE>

6. They are easy to understand.

RATE>

7. They are consistent with user's previous knowledge.

RATE>

8. They are generally/specifically suitable for user's needs.

RATE>

9. User knows their authors personally.

RATE>

10. User knows the journals/conferences.

RATE>

11. The authors are eminent.

RATE>

12. They describe methods/techniques.

RATE>

13. They are controversial.

RATE>

14. They provide a good bibliography and links.

RATE>

15. They provide background or history.

RATE>

16. They validate user's viewpoint.

RATE>

Yours truly,

Amir Ghaebi

PhD Student
Department of Information Science
Loughborough University

Appendix 4: The Letter to Research Co-ordinator of the Departments

Dear .....

I am a research student in the Department of Information Science here. I am conducting a research study under the supervision of Professor Charles Oppenheim into the Evaluation of Information Retrieval Systems Effectiveness. In order to carry out this research, I would like to invite 30 postgraduate students in the Science Faculty to assist in my study. I would be very grateful if you will let me have the email addresses of all your full-time MPhil/Ph.D. students, so that I can informally invite them to help me in my research. This will not take much of their time. If you have any question about this request, please do not hesitate to let me or Professor Oppenheim know.

I would like to thank you in advance for your co-operation.

Amir Ghaebi

PhD Student
Department of Information Science
Loughborough University
A.Ghaebi@lboro.ac.uk

Appendix 5: The Draft of the Letter to Invite A Postgraduate Student for Participating in the Study

Dear .....

I am a research student in the Department of Information Science here. I am conducting a research study under the supervision of Professor Charles Oppenheim into the Evaluation of Information Retrieval Systems Effectiveness. In order to carry out this research, I would like to invite you to help me in my research by undertaking for me to a search of the ISI Web of Science and then filling out a questionnaire. This will not take much of your time and will be fun to do (about 30 minutes). Your contribution will help this research, and the results can help you and others to improve their searches in the future. If you are interested in such research and have the time, please contact me by email and we will agree an appointment.

I would like to thank you in advance for your co-operation.

Amir Ghaebi

PhD Student
Department of Information Science
Loughborough University
A.Ghaebi@lboro.ac.uk

Appendix 6: Cover Letter for Participating in an Experiment

Dear Colleague,

As I mentioned in my last email, let me invite you to participate in my experiment. For this purpose, I would like to ask you perform an online search of ISI WEB of SCIENCE (It is needed your ATHENS user ID and password). After getting the results of search, you will be asked to select your relevant records. In this stage you will be asked to fill out a questionnaire (that is followed by this email). If you have any difficulties to answer a question or to understand the meaning of a statement we can arrange a meeting for my explanation and a chat about that. A £10 book token will be given to you as an appreciation of your effort and time.

Yours sincerely,

Amir Ghaebi
PhD Student
Department of Information Science
Loughborough University

Instructions: This survey has been designed so that you can respond directly on the survey contained in this e-mail and simply reply to sender to transmit your responses. For each question you are given a series of possible responses. Please put your response directly after the indicated arrow (Answer>) for each question. PLEASE DO NOT ALTER OR EDIT OTHER PORTIONS OF THE SURVEY, INCLUDING THESE INSTRUCTIONS.

1. Perform a search by connecting to Web of Science at: wos.mimas.ac.uk (if you do not have your ATHENS username and password contact to the library for getting one).

2. After formulating your search terms and getting the results, evaluate your search results and choose a number of them as relevant.

IF YOU ARE READY TO START, PLEASE CLICK THE REPLY OPTION AND BEGIN ANSWERING THE QUESTIONS BELOW. IF YOUR REPLY SETTINGS ERASE THE SURVEY, REOPEN THE E-MAIL, AND USE YOUR MOUSE TO COPY AND PASTE THE SURVEY INTO YOUR REPLY SCREEN.

1) How many different search strategies did you use in the session?

ANSWER>

2) How many articles did you get as search results in your final search?

ANSWER>

3) How many articles were completely related to your search topic?

ANSWER>

4) How many articles were partially related to your search topic?

ANSWER>

5) How many articles were not relevant to your search topic?

ANSWER>

6) How much time did you spend for searching?

ANSWER>

The following statements refer to your thought or feeling about the search results. If you dealt with more than one article, provide your overall impression of how you had relevance judgement.

PLEASE RATE HOW STRONGLY YOU AGREE OR DISAGREE WITH THE FOLLOWING STATEMENTS:

7) The articles returned were about my topic.

Response Options: 1=Strongly Agree 2=Agree 3=Unsure 4=Disagree 5=Strongly Disagree

ANSWER>

I selected a number of the articles as relevant because:

8) They were unique or only sources.

Response Options: 1=Strongly Agree 2=Agree 3=Unsure 4=Disagree 5=Strongly Disagree

ANSWER>

9) They contained original ideas or results.

Response Options: 1=Strongly Agree 2=Agree 3=Unsure 4=Disagree 5=Strongly Disagree

ANSWER>

10) They were new to me.

Response Options: 1=Strongly Agree 2=Agree 3=Unsure 4=Disagree 5=Strongly Disagree

ANSWER>

11) They were very informative.

Response Options: 1=Strongly Agree 2=Agree 3=Unsure 4=Disagree 5=Strongly Disagree

ANSWER>

12) They were easy to understand.

Response Options: 1=Strongly Agree 2=Agree 3=Unsure 4=Disagree 5=Strongly Disagree

ANSWER>

13) They were consistent with each other.

Response Options: 1=Strongly Agree 2=Agree 3=Unsure 4=Disagree 5=Strongly Disagree

ANSWER>

14) They were generally/specifically suitable for my needs.

Response Options: 1=Strongly Agree 2=Agree 3=Unsure 4=Disagree 5=Strongly Disagree

ANSWER>

15) I know their authors personally.

Response Options: 1=Strongly Agree 2=Agree 3=Unsure 4=Disagree 5=Strongly Disagree

ANSWER>

16) I know the journals/conferences.

Response Options: 1=Strongly Agree 2=Agree 3=Unsure 4=Disagree 5=Strongly Disagree

ANSWER>

17) The authors are eminent.

Response Options: 1=Strongly Agree 2=Agree 3=Unsure 4=Disagree 5=Strongly Disagree

ANSWER>

18) They described methods/techniques.

Response Options: 1=Strongly Agree 2=Agree 3=Unsure 4=Disagree 5=Strongly Disagree

ANSWER>

19) They were controversial.

Response Options: 1=Strongly Agree 2=Agree 3=Unsure 4=Disagree 5=Strongly Disagree

ANSWER>

20) They provided a good bibliography and links.

Response Options: 1=Strongly Agree 2=Agree 3=Unsure 4=Disagree 5=Strongly Disagree

ANSWER>

21) They provided background or history.

Response Options: 1=Strongly Agree 2=Agree 3=Unsure 4=Disagree 5=Strongly Disagree

ANSWER>

22) They validate my viewpoint.

Response Options: 1=Strongly Agree 2=Agree 3=Unsure 4=Disagree 5=Strongly Disagree

ANSWER>

23) If there were other criteria that caused you selected the articles as relevant, please write it:

ANSWER>

---

24) Gender:

Response Options: 1=Male 2=Female

ANSWER>

25) Age:

ANSWER>

26) Department:

ANSWER>

27) Year of study:

Response Options: 1=1st year 2=2nd year 3=3rd year or above

ANSWER>

28) Research Topic:

ANSWER>

CLICK SEND ON YOUR E-MAIL SOFTWARE TO SEND YOUR RESPONSES BACK TO ME.

***** THANK YOU VERY MUCH FOR HELPING ME ******

Appendix 7: Printed copy of Questionnaire for Evaluation of Information Retrieval Systems Effectiveness

Dear Colleague,

The purpose of this experiment is to evaluate Interactive Information Retrieval Systems, particularly Science Citation Index on Web of Science. The results will be useful to both research academics and information systems providers. As a researcher and user of those systems, your views are very important to this investigation. It would be appreciated if answer the following questions.

Yours sincerely,

Amir Ghaebi
PhD Student
Department of Information Science
Loughborough University
A.Ghaebi@lboro.ac.uk

Instructions:

I) Perform a search by connecting to Web of Science at: wos.mimas.ac.uk (if you do not have your ATHENS username and password contact to the library for getting one). Moreover an online tutorial is available at:

www.isinet.com/tutorials/webofscience.

II) On the first screen select Science Citation Index (Expanded) by clicking check box.

III) After formulating your search terms, evaluate your search results and choose a number of them as relevant.

1. How many different search strategies did you use in the session?
2. How many articles did you get as search results in your final search?
3. How many articles were completely related to your search topic?
4. How many articles were partially related to your search topic?
5. How many articles were not relevant to your search topic?
6. How much time did you spend for searching?

The following statements refer to your thought or feeling about the search results. If you dealt with more than one article, provide your overall impression of how you had relevance assessment.

Please rate how strongly you agree or disagree with the following statements and tick one checkbox for each statement.

	Strongly Agree	Agree	Unsure	Disagree	Strongly Disagree
7. The articles returned were **about my topic?**	☐	☐	☐	☐	☐
I selected a number of the articles as relevant because:					
8. They were **unique or only sources.**	☐	☐	☐	☐	☐
9. They contained **original** ideas or results.	☐	☐	☐	☐	☐
10. They were **new** to me.	☐	☐	☐	☐	☐
11. They were very **informative.**	☐	☐	☐	☐	☐
12. They were **easy to understand.**	☐	☐	☐	☐	☐
13. They were **consistent** with each other.	☐	☐	☐	☐	☐
14. They were **generally/specifically suitable** for my needs.	☐	☐	☐	☐	☐

15. I know their **authors** personally. ❒ ❒ ❒ ❒ ❒

16. I know the **journals/conferences.** ❒ ❒ ❒ ❒ ❒

17. The **authors** are **eminent.** ❒ ❒ ❒ ❒ ❒

18. They described **methods/techniques.** ❒ ❒ ❒ ❒ ❒

19. They were **controversial.** ❒ ❒ ❒ ❒ ❒

20. They provided **a good bibliography and links.** ❒ ❒ ❒ ❒ ❒

21. They provided **background or history.** ❒ ❒ ❒ ❒ ❒

22. They **validate my viewpoint.** ❒ ❒ ❒ ❒ ❒

23. If there were other criteria that caused you selected the articles as relevant, please write it: .....

Finally, please could you provide some information about yourself and your research.

Completed by: Gender: Male❑ Female❑ Age:

Department: Year of study: 1st ❑ 2nd ❑ 3rd or above❑

Research Topic:

Tel: Email: Date:

Thank you very much for helping me.