

ON USER-ORIENTED RELEVANCE JUDGMENT

CHEN ZHIWEI

(B. Eng., Xi'an Jiaotong University, P.R. China)

**A THESIS SUBMITTED
FOR THE DEGREE OF MASTER OF SCIENCE
DEPARTMENT OF INFORMATION SYSTEMS
NATIONAL UNIVERSITY OF SINGAPORE
2004**

Acknowledgment

My foremost and special thanks go to my supervisor Dr. Xu Yunjie. I am indebted to his invaluable guidance, and support throughout this project. His expert advice and great effort in helping me overcome many problems met in my research. This work could not be done without his careful guidance and persistent encouragement.

I would also thank Dr. Kim Hee Woong, Dr. Bock Gee Woo, and Dr. Tan Gek Woo for their kindly help for my data collection.

I would like to thank one of my best friends-Cai Shun for his kindly help for my study and insight comments on my research project. And many thanks go to my labmates for their constructed advices and supports for my survey design.

I also gratefully thank School of Computing of NUS for the rich research resources and financial support.

Content Table

1. Introduction.....	1
2. Literature Review	3
2.1. The Concept of Relevance	3
2.1.1. Objective Relevance	4
2.1.2. Subjective Relevance	6
2.2. Relevance Judgment and Criteria	10
2.2.1. Relevance Judgment	10
2.2.2. Relevance Criteria	12
3. Theory and Research Model.....	18
3.1. Theory	18
3.2. Research Model and Hypothesis.....	26
3.2.1. Topicality	26
3.2.2. Reliability.....	27
3.2.3. Scope.....	29
3.2.4. Understandability.....	29
3.2.5. Novelty.....	31
3.3. Affective Relevance (An Additional Study).....	32
3.4. Alternative Models	35
4. Methodology	37
4.1. Instrument Development.....	37
4.2. Data Collection	39
5. Data Analysis and Result	39
5.1. Pilot study	39
5.2. Main Study.....	41
5.2.1. Measurement Model	41
5.2.2. Structural Model	44
5.3. Testing of Alternative Models	45
6. Discussion and Implications.....	48
6.1. Discussion.....	48
6.1.1. Relevance.....	49
6.1.2. Affective Relevance.....	50
6.1.3. Alternative model	51
6.2. Limitation of This Study.....	52
6.3. Implications	53
6.3.1. Theoretical Implications	53
6.3.2. Practical Implications	55
7. Conclusion	56
References.....	58
Appendix A.....	A
Appendix B	B
Appendix C.....	C

Summary

The rapidly expanding Internet and other digital document depositories have generated a huge amount of textual documents. Searching for relevant information is increasingly being a hard and frustrating task. This phenomenon has brought in the doubt of the effectiveness of mechanical way of relevance definition, and has triggered a resurgence of interest in the concept of relevance. Relevance is regarded as the “fundamental and central concept” in information sciences. As a result of the inadequacy of a system- and algorithm-oriented perspective on relevance, recent studies have adopted a user-oriented and subjective perspective.

How does a user perceive a document as relevant? The literature on relevance has identified numerous factors affecting such judgment. However, there are a few important limitations associated with the prior studies. First, these large numbers of factors are not clearly defined and overlap with each other in meaning. Second, although researchers have suggested a core set of criteria for relevance judgment, few studies have investigated it. Finally, methodologically prior studies are exploratory and data-driven, confirmatory study with hypothesis testing procedure is needed to verify the early conclusions.

Taking a cognitive approach, this study focuses on the criteria users employ in making (situational) relevance judgment. Based on Grice’s communication theory, this paper proposes a theory-driven model and identifies five important relevance criteria: topicality, novelty, reliability, understandability, and scope. In addition to the main objective of this study, affective relevance as another perspective of relevance is also proposed in this paper. Meanwhile, alternative models are proposed to test whether

there are interaction effects among the proposed factors, which have been implied in the literature. A survey study is carried out and data analysis following a psychometric procedure is done to test the proposed model.

Our result shows that topicality and novelty are significant to relevance judgment. This result confirms the early suggestion that topicality is the centre part of relevance. And it also suggests that novelty is the next key relevance criterion beyond topicality. However, the other three criteria are not supported by the test. It is too harsh to conclude that these factors are unimportant in general. The non-significant might be due to the design of survey. A further verification of these hypotheses is needed in other contexts. As an additional test, the result supports our proposition that situational and affective relevance are separated and correlated. For the alternative model test, both linear additive and moderated multiplicative model are equally good in interpreting user's relevance judgment in our test. Further study is needed to investigate in this direction.

On User-Oriented Relevance Judgment

1. Introduction

The rapidly expanding Internet and other digital document depositories have generated a huge amount of textual documents. Information overload has become a pressing issue for users of such systems (Wurman 1989). Searching for relevant information is increasingly being a hard and frustrating task. With the huge amount of textual documents retrieved by typical information retrieval (IR) systems nowadays, most are found irrelevant. This phenomenon has brought in the doubt of the effectiveness of mechanical way of relevance definition such as the famous cosine score in the vector space model, and has triggered a resurgence of interest in the concept of relevance.

Relevance is regarded as the fundamental and central concept in information sciences (Schamber et al. 1990, Saracevic 1975). As a result of the inadequacy of a system- and algorithm-oriented perspective on relevance, recent studies have adopted a user-oriented and subjective perspective. For example, Saracevic (1970, pp.116-120) argues that by the late 1950s there was “official recognition that relevance may not be just a simple system phenomenon related to the effectiveness of matching within a retrieval system, i.e., only the user himself may judge the relevance of the document to him and his uses”. Subjective relevance concepts like psychological relevance and situational relevance are accepted, at least theoretically, as replacements or extensions of the objective and system-determined relevance. In general, relevance is now regarded as a subjective, multidimensional, dynamic, and measurable concept (Schamber et al. 1990).

If relevance is subjective, then what makes a user judge a document as relevant? Many different document attributes have been identified to affect relevance judgment, including novelty, reliability, topicality, among others. Such a list of document attributes can easily contain more than twenty criteria (e.g. Barry and Schamber 1998). However, the extant research suffers a few limitations. First, when the number of factors is so large, it obscures the key factors. Second, although Barry and Schamber (1998) suggest that there is a core set of user criteria cross different situations, no consensus has been reached regarding the set and the definition of these key factors in the set. One factor that seems to be omnipresent is topicality (e.g. Hirsh 1999, Schamber and Bateman 1996, Wang and Soergel 1998, Bateman 1999). In fact, topicality has been identified as the first or basic condition of relevance (Boyce 1982, Greisdorf 2003). In contrast, factors beyond topicality are much divergent. The question remains regarding what they are and how important they are. Finally, methodology wise, past studies are almost exclusively exploratory and data-driven. Exploratory studies are very useful to uncover an unknown phenomenon. However, to further test the validity of the identified factors and weed out the unimportant ones, confirmatory study with hypothesis testing procedure is needed.

With a focus on user's relevance judgment, the purpose of this study is to:

- Identify a set of core factors using a theory-driven approach.
- Test the proposed model with a rigorous psychometric approach.

The paper is organized as follows: In the second section, we first review the literature on relevance and relevance judgment. In third section, a set of core factors is identified based on Grice's (1975, 1989) communication theory, which leads to our research

model and hypotheses, and alternative moderated multiplicative models are also proposed. In the following fourth and fifth section, the empirical study is discussed and the data analysis is reported. In the last section, we discuss the theoretical and practical implications of the findings.

2. Literature Review

2.1. The Concept of Relevance

What is relevance? For more than fifty years, information scientists have attempted to conceptualize this concept, and have defined it in different ways (Saracevic 1975, Schamber 1994). Table 1 summarizes some of these definitions.

Table1. Concept of Relevance

Author	Concept or Definition
Rees & Schultz (1967)	A relation between system responses and request established by a judgment made by the user or his delegate.
Cooper (1971)	The relationship between sentences in queries and sentences in documents as determined by deductive logic.
Wilson (1973)	The relation between an information object and information recipient's individual and personal view of the world and his or her situation in it.
Saracevic (1970, 1975)	Relevance is the (A) gage of relevance of an (B) aspect of relevance existing between an (C) object judged and a (D) frame of relevance as judged by an (E) assessor.
Lancaster (1979)	A relationship between a document and a request as seen by judge.
Bookstein (1979)	The degree to which the user senses that the need bringing him to the system is satisfied.
Swanson (1986)	A mental experience of an individual who has information need.
Sperber and Wilson (1986)	A theoretical concept of psychology, as a relation between assumption (premise) and a context. A necessary condition for an assumption to be relevant in context is for the assumption to have context effect in that context.
Saracevic and Kanter (1988)	Relevance as the aboutness of individual items in the retrieval set.
Schamber (1991)	A concept that expresses a value judgment about the quality of a relationship between information and information need (or in formation

	problem) at a certain time in an information-seeking and use situation.
Harter (1992)	A state of effect that exists when user retrieves information, which suggests new cognitive connections, fruitful analogies, insightful metaphors, or an increase or decrease in the strength of a belief.
Green (1995a, b)	The property of a text's being potentially helpful to a user in the resolution of a need.
Campbell and van Rijsbergen (1996)	The degree to which evidence from the retrieved information objects is representative and indicative of the current information need.
Tang, Shaw, and Vevea (1999)	A user's decision to accept or reject a retrieval information item according to his or her information need at the time of retrieval.
Barry (1994)	Relevance is conceptualized as any connection that exists between the user's information need situation and the information provided by document.
Park (1997)	The relationship between the document and user, which is mentioned by the user for the evaluation of a document
Bateman (1998)	This paper views relevance as relevance, usefulness, value, satisfaction, or some combination of these variables.
Wang and White (1999)	Relevance is a relationship between a need and a document judged by a person.
Hjørland and Christensen (2002)	Relevance is defined as something serving as a tool to a goal. "Tool" is understood in the widest possible sense, including ideas, meanings, theories, and documents as tools.

Although the definitions in Table 1 are brief, they nevertheless reflect different emphases placed on this concept. We will provide a review on the development of the concept of relevance.

2.1.1. Objective Relevance

In the traditional perspective of information science, topical matching- whether the topic of the retrieved document matches the topic of the request, is the common acceptable relevance definition. This system-oriented definition is also conceptualized as (objective) topicality (Schamber et al. 1990), or system or algorithmic relevance (Saracevic 1996). However, Schamber, Eisenberg, and Nilan (1990) argue that the limitation of the concept of system relevance lies in the inadequate premise that the

subject term (user's query) can present meaning and system can use these concrete representations of meaning to match document to the user's information need. However, the user's information need (real meaning) is not concrete or observable. Thus, the system-oriented relevance definition is insufficient and incomplete.

Not satisfied with the linear, mechanistic, and static of traditional definition and model, information scientists shift their focus from the rationalistic traditional perspective to alternative perspective (e.g., Dervin 1983a, Dervin & Nilan 1986, Ingwersen 1984, Katzer 1987, Saracevic 1970, 1975, Wilson 1984), which allows more room for consideration of internal value stemming for the relevance judges themselves. Table 2 shows the comparisons of these two major perspectives in information science.

Table 2. Two Major Perspectives in Information Science

	Traditional	Alternative
General	Rationalistic, logical-empirical; assumes users and systems exist as separate entities	Interpretive, situational, contextual; assumes perceptions of systems exist within users
User	Rational, orderly, passive receiver of information	Interpretive, complex, active constructor of information
Information	Objective, communicable representation of knowledge	Subjective concept relating to some change in knowledge user's cognitive state or knowledge base
Meaning	Exists outside of user and is commonly knowable at a group level; is constant; can be indicated topically (e.g., by subject terms) and structurally (e.g., by syntax)	Exists within user, based on his/her interpretations (including group understandings) that are complex but knowable; is constantly reconstructed; can be indicated non-topically (e.g., by context, format) as well as topically and structurally
Information need	Can be expressed accurately and completely by user in form of representation (e.g., formal request) presented to IR system; can be resolved by substantive, topical content; remains	Involves user's perception of gap or anomaly in his/her knowledge base that he/she may not be able to express adequately to IR system; involves user's values, expectations, and perceptions of situations and cannot be entirely resolved by topical content;

	stable during IR process	changes constantly during IR process
Relevance	Judgment of effective contact between system and user	Judgment of quality of relationship between information and user's information need
Information behavior models	Clearcut, linear, trans-situational; based on formal logics; assume stability, predictability in user-system interactions (e.g., Source-to-Destination communication model)	Complex, inter-subjective, situational; based on interactive human behaviors; assume constant change, with predictability in certain situational environments (e.g., Sense-Making model)
Research	Concerned with improving ability of IR systems to deliver relevant information items to users; focuses on distinct user-system interactions; assumes only external states and behaviors of users are systematic, measurable; involves variables based on readily observable, quantifiable features (e.g., demographics)	Concerned with understanding how information (from formal or informal, external or internal sources) resolves needs and serves purposes of users; focuses on users' dynamic information problem situations; assumes internal cognitive states and behaviors of users are systematic, measurable; involves variables based on qualitative aspects of users related to situations and environments (e.g., information tasks, uses).

(Source: Schamber, Eisenberg, and Nilan 1990, p.768)

Schamber, Eisenberg, and Nilan (1990) conclude that the alternative perspective is not against the traditional one, but extends the conceptualizations and approaches and brings the psychology, linguistics, communication, and social science disciplines into information science research. With the shift of perspective, researchers gradually turn to user-oriented relevance study.

2.1.2. Subjective Relevance

Since 1960's, there has been a general trend that relevance is increasingly regarded as a subjective concept as oppose to an algorithm-determined one (e.g., Saracevic 1975, Schamber 1994, Cosijn and Ingwersen 2000, Borlund 2003, and Mizzaro 1997). The

term subjective relevance is used as an umbrella to cover the concept of subjective topicality and situational relevance, in which subjective topicality is the aboutness of a document with regarding to the user's information need as perceived by the user (e.g. Schamber et al. 1990, Hjørland and Christensen 2002, Mizzaro 1997) and situational relevance is the usefulness, value, utility, pragmatic application, or pertinence of a document in related to the fulfillment of interests, tasks or problematic situations intrinsic to user (e.g., Saracevic 1975, Cosijn and Ingwersen 2000, Hjørland and Christensen 2002, Park 1997, Mizzaro 1997).

The topical relevance extends system-determined query-document match which is known as system or objective relevance. However, it is different from the system relevance. While system relevance is judged by mechanical criteria such as the cosine similarity in the vector space model, topical relevance is the relatedness between a document and the topic area of interest as judged by the user. Although topicality is important to understand the concept of relevance, "relevance is not necessary the same as topicality" as indicated by Bookstein (1979, p. 270). He explains that a document on topic may not be judged relevance, if a user is already familiar with the document's content, or is only interested in an aspect of topic treaded in the document. Boyce (1982) further argues that merely hitting on the topic area is insufficient; users are looking for informativeness beyond topicality. Hersh (1994)'s study in medical field also calls for situational factors in defining what is relevance. In 1990's, more researchers turned to the cognitive and situational aspects of this concept (e.g., Harter 1992, Barry 1994, Park 1997).

Situational relevance takes a pragmatic perspective and defines relevance as the utility of a document to the fulfillment of user's task or problematic situation. In this view, if a document contributes to the problem-solving, it is relevant; otherwise irrelevant. Wilson (1973, p.458) first introduces the concept of situational relevance and describes it as "the actual uses and actual effects of information: how people do use information, how their views actually change or fail to change consequent on the receipt of information." Saracevic (1975, 1996) regards the utility perspective of relevance as a cost-benefit trade-off. Saracevic (1975, p.334) indicates that "it is fine for IR systems to provide relevant information, and the true role is to provide information that has utility-information that helps to directly resolve given problem, that directly bears on given actions, and /or that directly fit into given concern and interests." Borlund (2003, p.922) conceptualizes situational relevance as a user-centered, empirically based, realistic, and potentially dynamic type of relevance.

Both subjective topicality and situational relevance take a psychological perspective. Although the concept of psychological relevance has been proposed (e.g., Harter 1992), it is not another type parallel to the former two. Rather, it addresses the common psychological nature of them. Such psychological nature can be summarized into follow characteristics. First, the situational or task requirement from the external environment is translated into a cognitive state associated with uncertainty, which creates a desire to know the unknown (Saracevic 1975) in the user. Such internalized external requirement constitutes the information need (Saracevic 1975, Borlund 2003, Schamber et al. 1990) which both motivates an information seeking behavior and establishes relevance judgment criteria. Second, a document is "consumed" and the physical attributes of it (e.g. publication date, writing style, content) are internalized

into a set of psychological perceptions such as topicality, novelty, understandability, and so on. A rich array of research on user relevance judgment criteria (e.g., Hirsh 1999, Schamber and Bateman 1996, Wang and Soergel 1998, Bateman 1999) has uncovered a large set of both physical/objective document attributes and psychological perceptions of them. Third, the psychological perception of a document is judged against the information need to form a relevance judgment of the document. Forth, according to Sperber and Wilson’s (1986) relevance theory, the psychological perception of a document also changes the cognitive state of the user and his information need. Such interaction effect is known as context effect of the user’s cognition (Harter 1992). The context effect is the theoretical foundation of relevance being a dynamic concept. Schematically, the relationship can be summarized as in Figure 1.

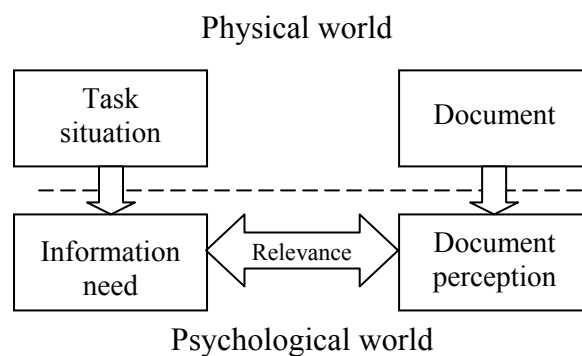


Figure 1. The Psychological Nature of Relevance

Between topicality and situational relevance, topicality is viewed as a basic requirement while situational relevance is viewed as a “higher” requirement as it corresponds more directly to user’s judgment in real situation (Borlund 2003). In this sense, situational relevance subsumes topicality. In this study, we adopt a situational definition of relevance and define it as the perceived utility of a document to the

fulfillment of the underlying information need of user's task or problematic situation. We term relevance refers to situational relevance hereafter.

2.2. Relevance Judgment and Criteria

2.2.1. Relevance Judgment

User-oriented relevance judgment research has been considered as an important foundation and approach toward understanding the concept and building model for relevance. Saracevic (1970, p.122) suggests that "relevance based on human judgment is complex but measurable, with relevance judgment falling into patterns that can lead to successful theorizing and subsequent predictions." Froehlich (1994, p.129) summaries the existing relevance judgment studies and states that "the absence of a unified definition of relevance does not mean that information scientist can not determine the diverse criteria that people brings to systems by which to judge its output." Lancaster and Warner (1993) emphasize the importance of end-user in making relevance judgment. They discuss that document-oriented judgment which is shared among a group of judges (e.g., indexer, intermediaries) and is somewhat objective. However, problem or user-oriented judgment can be made only by the user, which is related to their personal situations and is highly subjective. Greisdorf (2003) indicates that relevance as a judgmental process leading to evaluative measurement is both problem solving and decision making exercise involving facilitative cognitive processing.

How people make the relevance judgment? In prior studies, researchers use information behavior models to help explain how people account for or make sense out of unfamiliar aspects of their need, and suggest user's relevance judgment as a

cognitive and dynamic process. For example, Ree (1966, p.138) states that “it is our belief that judgment of relevance by a user is the expression of the user’s opinion as to how the information conveyed by a document matches, overlaps, complements, and/or is useful to his concept framework or previous knowledge.” Artandi (1973) views relevant information as a means of reducing uncertainty, and argues that “to reduce uncertainty information must be ‘relevant’ in the sense that it can be integrated and evaluated by the individual in terms of prior experience (his existing state) and his possible future states and activities.” Belkin et al (1982) views user’s information need as anomalous states of knowledge (ASK). They assume users as being in state of ambiguities in their mind, and an IR system is designed to resolve the user’s ASK by helping user articulate her need. Dervin’s (1983a, 1983b) sense making model concentrates on how users bridge cognitive gaps or uncertainties in a problem domain to make senses of information. Harter (1992) treats psychological relevance as the effect of document on user’s cognitive state based on communication theory. Harter’s study strongly supports relevance judgment as a cognitive and dynamic process because user’s cognitive state is inevitably changed by each document encountered, whether it is relevant or not.

Kuhlthau (1993) suggests the shift of user’s judgment from topical relevance to situational relevance in his/her information searching process. She further discusses that at the early stage of information seeking user’s document evaluation is to be a rather general standard of inclusion, whereas in the later stage of evaluation, users are found to be more discriminative and exclusive. Wang and Soergel (1998) bring the decision rules to investigate user’s document judgment, their results show that users

often use elimination rule (a salient criteria) to reject a document, and apply multi-criteria (several criteria) to accept a document.

2.2.2. Relevance Criteria

While relevance in general is conceptualized as a user's judgment of the strength of relationship between a document and information need (Saracevic 1975), a question that follows naturally is the criteria that users employ in such judgment. Schamber, Eisenberg, and Nilan (1990, p.771) notice the importance of relevance criteria studies and suggest that "an understanding of relevance criteria, or the reasons underlying relevance judgment, as observed from the user's perspective, may contribute to a more complete and useful understanding of the dimensions of relevance". As early as in 1960's, researchers have attempted to identify the criteria for relevance judgment. For example, Cuadra and Katter (1967a, b) find that relevance judgment is affected by 38 factors, such as style, specificity and level of difficulty of document. They suggest that user's relevance responds vary in relation to the characteristics of document and the "state" of user-his need, attitude, and knowledge toward the textual document judged (p.12). Ree and Schulz (1967) identify 40 variables that would affect relevance judgment and indicate the more information is given to user, the more stringent relevance judgment will be. Since 1990, many empirical studies were carried out to discover such criteria or factors in different problem domains (e.g., Schamber 1991, Park 1993, Barry 1994, Hirsh 1999). Table 3 summarizes some of these studies.

Table 3. Relevance Criteria in Past Research

Reference	Context	Subject (Sample size)	No. of criterion	Criteria	
Schamber (1991)	Weather information	Working people (30)	10	Presentation quality, Currency, Reliability, Verifiability, Geographic proximity, Specificity, Dynamism, Accessibility, Accuracy, Clarity	
Su (1993)	Assigned essay	Students (40)	5	Completeness, Precision, Relevance, Expectancy, Coverage	
Park (1993)	Academic research	Academic staffs, and graduate students (10)	22	Interpretation of a citation	Title, Style of the title, Author name, Journal name and document type, Abstract, Interconnections among elements in a citation
				Internal (Experience) Context	User's previous experience and perception, User's level of experience in the problem area, User's previous research experience, User's education (or training)
				External (Search) Context	Perception of the search quality, Purpose of search (or search goal), Perception about the availability of information, Priority of information needs, Stage of research, End product of the research
				Problem (Content) Context	Same (similar) problem for definition, Same (similar) problem as background, Similar problem off the target, Different problem for the methodology, Different problem for the framework, Different problem as background, Different problem not of interest, New information in the problem context, Old (that is , repetitive) information in the problem context, Insufficient information in the problem context
Barry (1994)	Online free search for information	Students (18)	23	Information content of document	Depth and Scope, Objective accuracy / validity, Clarity, Recency, Tangibility, Effectiveness
				Source of document	Source quality, Source reputation / visibility
				Document as a physical entity	Obtainability / available, Cost
				Other information and source	Consensus within the field, external verification, Available within environment, Personal available

				User's Situation	Time constraints, Relationship with author
				User's belief and preference	Subjective accuracy / validity, Affectiveness, Background / experience,
				User's background	Ability to understand, Content novelty, Source novelty, Document novelty
Park (1997)	Academic problem and need	Graduate students (24)	12	Applicable, good, helpful, important, interesting, need, new, related, relevant, similar, studied, useful.	
Wang and Soergel (1998)	Research project	Graduate students (25)	11	Topicality, Novelty, Subject area, Recognition, Quality, Orientation / level, Recency, Available, Special requisites, Authority, Relation / origin	
Tang and Solomon (1998)	Term paper	Graduate students (1)	10	Topical related, types of article, similar topical focus, duplicates, recency, length, depth/breadth, language, geographic focus, version of article (repetitiveness)	
Hirsh (1999)	Research paper on any sports	Primary students (10)	11	Textual material	Authority, Convenience / accessibility, Interesting, Language, Novelty, Peer interest, Quality, Recency / Temporal issues, Topicality
				Graphic material	Authority, Clarity/ completeness, Interesting, Peer interest, Expediency,
Fitzger and Galloway (2001)	Academic task	Under-graduates (10)	32	Relevance related reasoning	Interest, Specific idea, Useful or helpful, Specific use, Banned idea, Divergent, Specificity, Background, More is better, Essential, Serendipity, Prior knowledge
				Evaluation related reasoning	Good, Context, Methodology, Perspective, Insufficient, Author, Currency, Wrong methodology, Obvious, Strange, Disagree, Authority
				Affect related reasoning	Funny, Like or dislike, Disturbing, Want, Sad, Annoy, Happy, Fun
Maglaughlin and Sonnewald (2002)	Academic need for research paper or thesis	Graduate students (12)	29	Abstract	Citability, Informativeness
				Author	Author novelty, Discipline, Institutional affiliation, Perceived status, Accuracy - validity, Background
				Content	Content novelty, Contract, Depth-scope, Domain, Citations, Links to other information, Relevant to other interests, Rarity, Subject matter, Though catalyst

				Full text document	Audience, Document novelty, Type, Possible content, Utility, Recency
				Journal or Publisher	Journal novelty, Main focus, Perceived quality
				Participant	Competition, Time requirements
Choi and Rasumuseen (2002)	Images in American history	Students (38)	9	Topicality, accuracy, time frame, suggestiveness, novelty, completeness, accessibility, appeal of information, technical attributes of images	

Through these empirical studies, a large number of criteria/factors have been identified under different situations or tasks. These studies provide a rather comprehensive view of relevance criteria. However, there are a few important limitations associated with prior studies. First, the number of factors is very large. If a predictive model is to be built eventually in an IR system, asking user to provide feedback on all these factors or automatically measuring all of them is surely impractical. Second, the terminology is confusing. Same criterion (according to its definitions in papers) was named differently by authors and users (e.g., reliability, accuracy, validity), which calls for a combination of concepts (Greisdorf 2003, Schamber and Bateman 1996). Third, factors overlap with each other in meaning (e.g. accuracy and reliability, utility and usefulness). Fourth, the judgment of an IR system and the judgment of document need to be distinguished. For example, accessibility of a document is more a property of an IR system (whether it carries a certain document or not) rather than that of the document content per se. The relevance of a document should be based on its content rather than its physical property such as availability. Five, document attributes and relevance evaluations are treated at the same level. Variables like utility, usefulness, and helpfulness should be treated as a certain aspect of relevance judgment itself, i.e., the dependent variables, rather than independent variables (i.e., the criteria). Document attributes, whether objectively observed (e.g., date of publish) or subjectively perceived (e.g., novelty) should be the independent variables.

A special case of the overall evaluation of a document is its hedonic consequence such as enjoyment and happiness. Just like consumption of product can be for utilitarian or hedonic purpose, so is the consumption of information (Wang et al 1998). This study mainly focuses on the utilitarian perspective of relevance, and the hedonic perspective

of relevance is also discussed and tested for research completeness. Finally, as mentioned above, methodologically these studies are exploratory rather than confirmatory. After so many explorations, a theory-driven and confirmatory study is called for to verify these results.

Some of the above-mentioned problems have been identified by prior research as well. For example, Schamber (1994) notices that the content-related criteria seem to confirm and illustrate the multidimensionality of relevance. And certain features of document presentations (e.g., author, index term, and format) may serve as clues to content for user's judgment. Green (1995a) points out that the subject content of a text is the major factors underlying the relevance of that text to the user's context. Hertzum et al (2002) explain that the criteria related to efficient such as cost, availability is not for relevance, because these criteria are not directly concerned with the user's information need.

Some further studies are carried out based on the criteria identified by prior exploratory studies. Barry and Schamber (1998) compare the results of their two studies under totally different situations: academic and weather media, and find that a considerable overlap of relevance criteria. Table 3 also suggests that many studies share some common criteria. Bateman (1998) carries out a longitudinal study and find that the important criteria remain fairly stable throughout the whole process. In contrary, based on the process model proposed by Kulhthau (1993), Vakkari (2000) find different criteria used in information assessment at different stages of seeking. Borlund (2003) suggests that difference of Batman and Vakkari's results may due to the criteria have been assessed for partial and highly relevance at the same time and

formulation of information need. Borlund (2003, p.918) explains that “relevance criteria (may) change as the information need develops and matures, but also that an information need may be composed of several subfoci, which, consequently, are represented by different relevance criteria. The possible existence of subfoci, and priorities of subfoci, may explain the assessment behavior in cases where information objects are assessed as either partially relevant; or only parts of the object are assessed (highly) relevant.”

Summarizing from the prior literature, it seems that there is a set of core relevance judgment criteria that most users would follow. However, the importance of a particular criterion might change depending on the context and the stage of a user’s information behavior. For example, Barry and Schamber (1998) attribute the a few criteria not overlapping in their two studies to the differences in the situational environments. The questions remain: What are the set of core relevance criteria and how should we conceptualize them? This study attempts to address this question.

3. Theory and Research Model

3.1. Theory

Departing from the extant research which adopts an inductive and grounded exploratory methodology, we adopt a theory-driven and deductive approach. A theory-driven approach has a certain advantages. First, established theories, such as those from psychology and communication, have been tested in many different contexts and have good generalizability. It is desirable to build our study on such intellectual foundation as a particular application. Second, established theory typically identifies

the most important factors in a general domain. When applying to a particular problem, these factors serve as general guidance to identify domain-specific factors. Finally, because established theory suggests the general relationships among relating factors, hypotheses can be built based on the relationships.

To identify the core relevance criteria, we propose that Grice's (1975, 1989) maxims on human communication can serve as a theoretical foundation of relevance judgment. Not only does Grice's framework of maxims address the human communication in general (in which IR can be regarded as an indirect form of human communication), it is also consistent with many empirical studies in the IR area as we will discuss briefly. Her work established the foundation of the inferential model in human communication which is more general than Shannon's code model of communication (Sperber and Wilson 1986). Grice (1975, 1989) posits that the essential feature of human communication, both verbal and non-verbal, is the recognition of speaker's intention. In this model, a hearer infers the speaker's meaning or thought on the basis of the words or information provided. A communication is successful when both parties are cooperative in making their meanings clear (i.e., the principle of cooperation). What kind of communication is cooperative? Grice further describes the hearer's expectation of the speaker's message in term of the following conversational maxims: quantity, quality, relation, and manner.

The maxim of quantity has two sub-maxims. In Grice's words, contributing appropriate amount of information to communication is to "make your contribution to the conversation as informative as is required," and "do not make your contribution to the conversation more informative than is required." While Grice has a focus on

conversational communication, a more appropriate term in written communication via documents would be “scope”. We identify “scope” as one relevant criterion. The maxim of quality also has two sub-maxims: “do not say what you believe to be false,” and “do not say that for which you lack adequate evidence.” We use the term “reliability” because “quality” implies more than what Grice means in IR. The maxim of relation is defined as to “be relevant.” However, the term “relevant” is in its daily sense -- whether a response is on topic or the other party abruptly starts to talk something else. In that sense, it is the “topicality” that we discussed above. Finally, the maxim of manner is to “avoid obscurity of expression,” “avoid ambiguity,” “be brief,” and “be orderly.” The purpose of this maxim is that conversation should be perspicuous hence reduce the cognitive load on the hearer. We term it “understandability” in the context of written document. In summary, based on Grice’s maxims, we identify four relevance criteria. We add to the four a fifth which is nevertheless implied by Grice’s maxim: novelty. Although Grice’s maxim of quantity focuses on the amount of information, it suggests that new information should be supplied; therefore the conversation is “informative.” Wang and Soergel (1998) suggest that novelty and the resultant epistemic value are implied in functional value of a document.

Grice’s theory plays a significant role in human communication and pragmatics studies (e.g., Bacht and Harnish 1979, Neale 1990). The communication maxims have been widely applied in other fields, such as optimality theory (Atlas and Levinson 1981), cooperative answering system (Gaasterland et al. 1992), spoken dialogue systems (Dybkjaer et al. 1998), etc. Most noticeably, in the communication study, Sperber and Wilson (1986) extend Grice’s work and develop the theory of relevance,

in which all Grice's maxims are reduced to the "principle of relevance," i.e., to conform to the maxims is to be relevant. Unfortunately, Sperber and Wilson (1986) focus on how a hearer adjusts cognitive context to make senses out of a message rather than on the perceptions of a message that makes it relevant. In comparison, Grice's maxims directly address this issue.

Applying Grice's theory and maxims to IR is appropriate (Hjørland and Christensen 2002). First, analogically, IR can be regarded as an asymmetric written communication between an author and the readers. The IR system can be seen as an intermediary that "speaks" for the authors. The iterative process of query and document matching is the process of "conversation". Users expect the system to be cooperative and the retrieved document to obey the maxims.

Second, the five criteria identified based on Grice's maxims correspond very well to the empirical findings in relevance research. Table 4 summarizes a representative list of such studies. As shown in table 4, many factors identified in prior literature tap directly on these six criteria (enjoyment as another perspective of relevance is also included in the table) or serve as antecedents of them. For example, accuracy (e.g., Barry 1994) is an aspect of reliability, while recency (e.g., Hirsh 1999) of a document leads to its novelty. Nevertheless, recency and novelty are different as recency suggests a physical publication time while novelty is a subjective perception of the content. A document can be novel to a reader, yet published long ago. Certain criteria identified previously are related to yet can not be classified exactly into one of the six categories. One of such variables might be prior knowledge (e.g. Fitzger and Galloway 2001). It is possible that prior knowledge on a topic increases understandability of

document as well as reducing its novelty. However, prior knowledge is so far the only variable that is identified to be “double-loaded”. Still other criteria include accessibility, relationship to the author, document format, link to other documents etc. Such criteria are not based on document content, but some peripheral attributes. We shall confine our relevance to the document content and only and safely ignore such peripheral attributes. In short, the five criteria (ignoring enjoyment) based Grice’s maxims are comprehensive enough to cover most criteria identified in prior user studies, which in return testifies the generalizability of the theory.

Table 4. The Main Factors with Prior Results

Reference	Topicality	Reliability	Understand-ability	Novelty	Scope	Enjoyment	Relevance	Others
Schamber (1991)	Geographic proximity	Accuracy, Reliability, Verifiability	Clarity	Currency, Dynamism	Specificity	----	Presentation quality	Presentation quality, Accessibility
Su (1993)	(Search result) Precision, Completeness	----	----	----	Coverage	Affect	Relevance	Time/Cost, subject, Knowledge on Accessibility
Barry (1994)	Assumed	Subjective and Objective accuracy / validity, Tangibility, Source quality, Source reputation, Consensus within the field, External verification	Ability to understand	Content novelty, Source novelty, Document novelty, Recency, personal availability, Background/ experience	Depth / Scope	Affectiveness	Effectiveness	Cost Obtainability / availability, Time constraints, Relationship with author, Availability with environment
Park (1997)	Related	----	----	New, Similar	----	Interesting	Applicable, Good, Help, important, Need, Useful, Relevance	----

Reference	Topicality	Reliability	Understand-ability	Novelty	Scope	Enjoyment	Relevance	Others
Bateman (1998)	About my topic	Accurate, Credible, Consistent, prominent, Reputation, provides proof, Describe method / technique, Statistical approach, Controversial	Well-written, Understand-able, Provides example, graphic, presentation of information, Provides background	Unique or the only source, original, New to me, Familiar, Current	Focused, Suitable general or specific, Suitable length, Detail, Introductory, Completeness Overview, Comprehensive	I like it, Validates my view point, Interesting Enjoyment	----	Easy to obtain, Free or inexpensive, , Format of source, Interactive, Provides bibliography or link
Spink et al (1998)	Include my search terms, all the concept searching for were included, Related to current problem, Wrong meaning of search term	Authoritative source	Wrong language. Don't understand context	Identifies a different, but related concept (new term), The chronology, Duplicate information	Specific to my query. On target, but too technical / narrow, dealt only partially with the subject Contained multiple concepts, too brief	Excited me	Answer my question, My personal image of what I perceived to be a relevant document, Could be help, Not useful	Not on money, List good resource/ reference
Wang and Soergel (1998)	Topicality	Quality	Special requisites	Novelty, Recency,	Orientation, subject area	----		Recognition, Availability, Time

Reference	Topicality	Reliability	Understand-ability	Novelty	Scope	Enjoyment	Relevance	Others
Tang and Solomon (1998)	Topical relatedness, Similar topical focus, Geographical focus	----	language	Duplicates, Recency, Version of article or repetitiveness	Depth / Breadth, Length	----	----	Types of articles,
Hirsh (1999)	Topicality	Authority, Quality	Clarity, Language	Novelty, Recency/ Temporal issue	----	Interesting, Peer interest	Expediency	Accessibility
Fitzger and Galloway (2001)	Context, Banned idea	Disagree, Authority, Wrong Methodology	Background	Divergent, Strange, Perspective, Currency	Specificity, In sufficient, More is better, Specific idea	Interest, Funny, Like or dislike	Useful or helpful, Essential, Good	Prior Knowledge
Choi and Rasumusse n (2002)	Topicality	Accuracy	----	Novelty	Completeness,	Appeal of information	Suggestive ness	Time frame, Accessibility, Technical attributes
Maglaughli n and Sonnewald (2002)	Subject matter, Domain, Relevant to other interests, Possible content	Accuracy- validity, Citations Perceived status, Perceived quality	Background	Content, document and journal Novelty, Rarity, Recency	Depth-scope, discipline, Main focus	----	Informative -ness, Thought catalyst, Utility	Institutional affiliation, Link to other information, Audience, Type, Competition, Time requirement

In summary, we adopt the situational definition of relevance, and defined it as the perceived utility of a document to the fulfillment of the underlying information need of user's task or problematic situation. Quite naturally following from Grice's theory, favorable judgment of the five relevance criteria should lead to favorable relevance judgment. We shall further justify each criterion in the next section. Figure 2 summarizes our proposed research model.

3.2. Research Model and Hypothesis

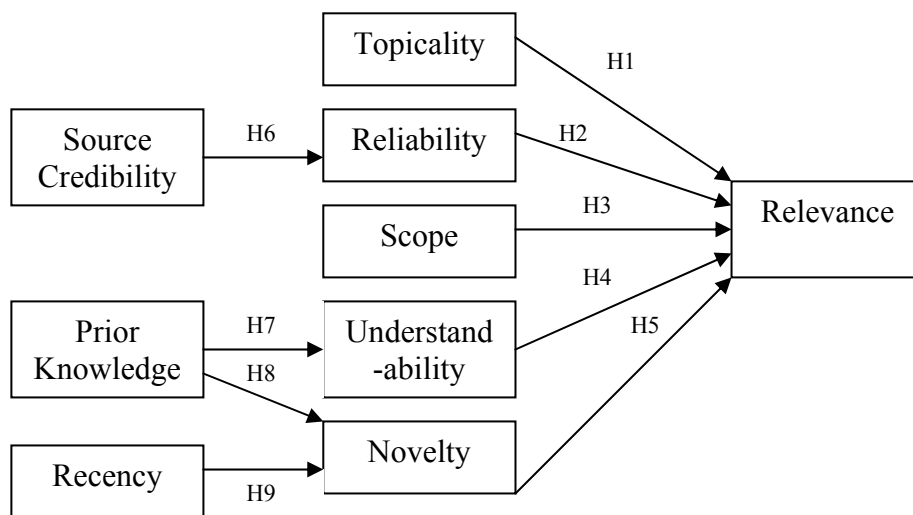


Figure 2. Research Model

3.2.1. Topicality

Topicality is the essence of Grice's maxims of relation. If a conversation is to be successful, the violation of this maxim is rare, if not impossible (Grice 1989). The concept of subjective topicality has been interpreted in variety ways as aboutness or subjective about (Maron 1977), topic related (Tang and Solomon 1998), topical relevance (Saracevic 1996), or intellectual topicality (Borlund 2003). Greisdorf and O'Connor (2003)'s study finds that user's major perception of "on topic" are relate to

the user's problem, contain the user's request, and describe, explain or expand user's topic.

Regardless of different terms, the importance of topicality is widely recognized in relevance literature. Maron (1977) points out that aboutness is the heart of indexing and subjective about is better in determining the conceptual relatedness for user's problem than system relevance. Boyce (1982) indicates that users first judge the topicality of document, and then think about other factors for their relevance judgment. Harter (1992) treats topicality as a weak level of relevance. Froehlich (1994) summarizes the early studies and notes the nuclear role of topicality for relevance. Greisdorf (2003) also acknowledges topicality as the first or basic condition of relevance.

We adopt a subjective view and define topicality as the extent to which the retrieved document is related to a user's current topic of interest as perceived by the user. Because of its foundational role in situational relevance, in consistent with almost all prior exploratory studies, we hypothesize:

H1: Document topicality is positively associated with relevance.

3.2.2. Reliability

Intuitively, people accept information that is perceived to be accurate. Grice (1989) observes that "quality" is the prerequisite for other maxims to operate. Ultimately, if a document is to be relevant by reducing uncertainty in the mind of the user, it must be reliable in itself first. Many different disciplines testify the importance of reliability. In

data quality management, accuracy is acknowledged as the (if not the only) key dimension of data quality (Wang et al. 1996). When evaluating output of database, without accuracy, user will dismiss its usefulness immediately. In persuasion literature of psychology, Petty and Cacioppo (1986) regard reliability as a critical aspect of the acceptance of the message and indicate that a message receiver first judges the reliability of information, and then decide whether to adopt it. In accounting research, Johnson et al. (1981) also show that reliability is the key criterion to evaluate the quality of data for acceptance. In IR literature, the importance of reliability has been noticed by some researchers. Taylor (1986) proposes the value-added model attempting to improve system performance and identifies reliability and accuracy as “value” that users bring or feedback to the systems in their document judgment.

How does user judge reliability of a document? Reliability is first and foremost determined by the document content. And making judgment of reliability of document retrieved from IR systems should be a difficult task for most users, because there is no quality control mechanism for Web and such systems. However, in addition to that, the credibility of the source can be regarded as an external cue of document reliability. Petty et al. (1994, p.103) note that “source status, by influencing perceptions of source credibility, competence, or trustworthiness, can provide message recipients with a simple rule as to whether or not to agree with the message.” Information from an expert is perceived more reliable than the one from a source without credential (Petty and Cacioppo 1986).

We define reliability as the degree that the content of a retrieved document is perceived to be true, accurate, or believable. Similar concepts in the literature are

accuracy (Schamber 1991), validity (Barry 1994), authority (Hirsh 1999), and agree/disagree (Fitzger and Galloway 2002). We hypothesize:

H2: Document reliability is positively associated with relevance.

H6: Source credibility is positively associated with document reliability.

3.2.3. Scope

Grice's (1989) maxim of quantity posits that adequate amount of information is what the hearer prefer. The concept of scope can be described in term of two components: breadth and depth (Miranda 2003, Belardo and Pazer 1985). Levitin and Redman (1995) suggest the scope and level of detail to be two important dimensions of data quality. They argue that a user needs the data to be broad enough to satisfy all the intended use and, at the same time, not to include the unnecessary information. For the level of detail, they further show that the detailed information may be used as quality safeguard, while too detailed information is an annoyance.

We define scope as the extent to which the topic or content covered in a retrieved document is appropriate to user's need, i.e., both the breadth and depth of the document are suitable. This definition represents similar concepts of specificity (Schamber 1991, Cool 1993, Fitzgerald and Galloway 2001), depth/scope (Barry 1994), depth/breadth (Tang and Solomon 1998) etc. We hypothesize:

H3: Document scope is positively associated with relevance.

3.2.4. Understandability

Understandability corresponds to Grice's maxim that a message should be perspicuous. Researches in communication and education show that the use of jargon or technical

language may reduce the clarity of a message (Dwyer 1999) and lead to significantly lower evaluation than a jargon-free message. Both expert and non-expert are sensitive to the use of jargon in report (Brown, Braskamp and Newman 1978, Thompson, Brown, and Furgason 1981). Similarly, in accounting research, understandability is also a measurement of the effectiveness of accounting reports to decision makers (Adelberg 1979). In a client-professional exchange, the use of sophisticated language may affect the acceptance of the professional's advice (Elsbach and Eloffson 2000).

How does user interpret his/her understandability of document? McNamara et al. (1996) suggest that stock of prior knowledge can be regarded as a user-related factor affecting the perceived understandability of given document content. They discuss that mental representation comparing to just read allows for a deeper understanding of the text, which is linked to the reader's long term memory and knowledge. Möller et al. (2000)'s study shows difference in text comprehension between readers with low and high prior knowledge. In hypertext reading test, Potelle and Rouet (2003) find that those with less prior knowledge in a topic area can be more sensitive to the presentation of the message and find it more difficult to understand.

We define understandability as the extent to which the content of a retrieved document is easy to read and understand as perceived by user. It unifies the similar concepts like clarity (Schamber 1991), ability to language use (Tang and Solomon 1998, Hirsh 1999), and special requisites (Wang and Soergel 1998). We hypothesize:

H4: Document understandability is positively associated with relevance.

H7: User's prior knowledge is positively associated with document understandability.

3.2.5. Novelty

Psychological researchers define novelty as a stimulus that has not been previously presented or observed and thus unfamiliar to the subject. In psychological literature, the novelty seeking behavior is regarded as an internal drive or motivation force of human being (Acker and McReynolds 1967, Cattell 1975, Farley and Farley 1967). Seeking new and potentially discrepant information may help people “create a ‘bank’ of potentially useful knowledge” and further “improve people’s problem-solving skills” (Hirschman 1980b, p.284). In human conversation study, Dessalles (1998) indicates the first way for relevant conversation is to bring the new information and novelty or prior improbability is one condition for informativeness. Lancaster (1968) first introduces the concept of novelty into IR research, and defines it as the retrieval of citations previously unknown to requester. Harter (1992, p.608) notices that normally “a citation corresponding to an article already known to the requester could not be psychology relevant” because it will not produce cognitive change in the subject. However, it may serve as a reminder. Therefore, novelty should be regarded as a matter of degree. Recent exploratory studies acknowledge novelty as an important factor affects relevance (e.g., Barry 1994, Choi and Rasumussen 2002).

What factors influence user’s perception of novelty of a document? Prior knowledge may serve as one clue for novelty. In human intelligence research, Raahein (1974) conceives of intelligence in novel problem solving situations as the search for how the present problem situation fits into some previously understood series of situations. He further argues as the individual’s range of past experience increases, people are more able to cope with further novelty situation, and the less novelty will be perceived by user in the similar situation. In Chase and Simon (1973)’s chess experiment test, they

find that chess master's ability to recall and recognize the regular position of chess superior to the novices. In marketing literature, some researches indicate that consumer often use information already known to learn about new product. Heit (1997) finds that existing knowledge may facilitate learning of new information and reduce people's complexity perception of information objects. Recency of the publishing date of documents maybe considered as another factor that affect user's novelty perception. Hirschman (1980) implies that the documents like magazine and newspaper are valuable to consumer, because every issue subscription presents a commitment to the consumer to provide new data. In relevance literature, a few prior studies (e.g., Tang and Solomon 1998, Cool 1993) mentioned the recency of a document for relevance judgment. However, when a user comments on recency, novelty is implicitly referred to. Thus, recency can be regarded as one possible way of ensuring novelty.

We define novelty as the extent to which the content of a retrieved document is new to the user or different from what the user has known before. It unifies the similar concepts such as content novelty (Barry 1994), new content (Park 1997), divergent and strange content (Fitzger and Galloway 2001), etc. We hypothesize:

H5: Document novelty is positively associated with relevance.

H8: User's prior knowledge is negatively associated with document novelty.

H9: Document recency is positively associated with document novelty.

3.3. Affective Relevance (An Additional Study)

The concept of affective relevance as another perspective of relevance has been commonly mentioned in the literature. Saracevic (1996, p.214) defines this type of

relevance as affective relevance and describes it as “the relation between the intents, goals and motivation of a user, and text retrieved by a system”. Wang and Soergel (1998) indicate the emotional value of document for relevance judgment. Although early exploratory studies have identified the affective criteria that users contribute to interpret relevance (see table 4), whether this affective relevance should be classed as a separate category or as the ultimate subjective relevance on a relevance scale is still not clear (Schamber 1994). In addition to the main purpose of this study, this paper makes a pilot study to attempt to address the question: What is the relationship between affective relevance and situational relevance?

In psychology literature, researchers have revealed that the people’s (reading) attitude can be divided into cognitive and affective aspects (Fishbein and Ajzen 1975, Lewis and Teale 1980, Greaney and Neuman 1990). The cognitive component is typically a specific set of beliefs about the perceived utilitarian or functional consequences of the act to one’s goal. The affective component is typically affective feelings or emotions that people have in related to their goal (McGuire 1969). Batra and Ahtola (1990) conceptualize utilitarian and hedonic consequences as the bi-dimensional of an attitude. In marketing research, utilitarian and hedonic consequences have been well studied and conceptualized them as two major perspectives of product consumption (e.g., Hirschman 1980a, Holbrook 1980). This framework is also suitable for document reading and evaluation, Stockman (1999) applies the attitude theory into reading behavior study and suggests the utilitarian and hedonic consequences as two aspects of reading attitude which finally affect people’s reading behavior.

What influence user's enjoyment of document? Novelty may be one factor. Robert South, one of famous English divines in 17th century, writes the saying that "novelty is the great parent of pleasure". In reading literature, some studies also suggest that user's perception of novelty or surprise within the content is usually associated with his/her pleasure in reading. For example, Brewer and Lichtenstein (1982) find that people would take pleasure in reading when the book arouses curiosity and resolves feelings of suspense. In another side, reader's ability to understand the text of document also influences their reading pleasure. Britton et al. (1978) indicate that readers are more fully absorbed by the easy reading text than the difficult one according to their reading capacity. For example, people are more likely to read the popular fictions and they often feel annoyance in reading incomprehensible books. Situational relevance can affect affective relevance. Based on cognition theory, utilitarian and hedonic aspects are normally positively correlated (Osgood et al 1957). For example, getting of an attractive product is like to increase the chances of a favorable emotional as well as a favorable functional response. Zajonc and Markus (1982) treat utilitarian and hedonic consequences as the antecedence of preference, and suggest that utilitarian would affect hedonic consequences for object evaluation.

Based on the psychological literature and early relevance studies, we regard situational relevance as utilitarian perspective of relevance and enjoyment as hedonic perspective of relevance. We define enjoyment as user's enjoyable or pleasure feelings of a document in relation to his need or goal. To explore the hedonic perspective of relevance, we identify the factors (maybe not all) that would affect the user's enjoyment judgment of document based on psychology and cognition literature. We hypothesize:

H10: Relevance is positively associated with enjoyment

H11: Document understandability is positively associated with enjoyment

H12: Document novelty is positively associated with enjoyment

Figure 2.1 summarizes the research model including enjoyment.

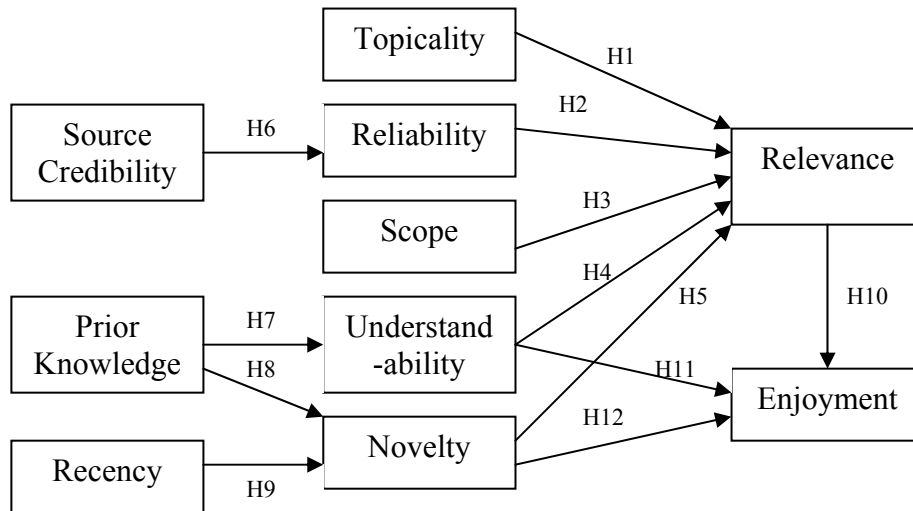


Figure 2.1. Research Model (including Enjoyment)

3.4. Alternative Models

When making relevance judgment, are users following an additive or multiplicative computation? The model in Figure 2 assumes an additive relationship of the criteria (i.e. relevance is a weighted sum of each criteria), while prior studies also suggest the possible multiplicative model, i.e. relevance judgment is made in different stages/levels. For example, Boyce (1982) indicates that the judgment of topicality is the first stage for relevance, and user employ other criteria in further judgment only when topicality is fulfilled. Greisdorf (2003) also finds that when topicality is not met, a document almost always judged as not relevant or partially relevant at most. Therefore, topicality is a necessary but not sufficient criterion for relevance. Similarly, Grice (1989) suggests that reliability of a message is a prerequisite maxim, and all

other maxims including topicality come into operation only on the assumption that this maxim is satisfied. Therefore, reliability is also a necessary but not sufficient criterion. Similarly, an argument can be posted that understandability is a prerequisite of other criteria. Underlying such argument is a multiplicative or interaction model, in which a low evaluation of one of the three would nullify the rest. Ignoring the external variables (i.e. source credibility, prior knowledge, document recency, and enjoyment), the model in Figure 2 and the multiplicative model suggested here can be represented as:

Model 1: $\text{Relevance} = \sum w_i c_i$, $c_i \in \{\text{topicality, reliability, scope, understandability, novelty}\}$

Model 2: $\text{Relevance} = \text{Topicality} + \text{Topicality} \times \sum w_i c_i$, $c_i \in \{\text{reliability, scope, understandability, novelty}\}$

Model 3: $\text{Relevance} = \text{Reliability} + \text{Reliability} \times \sum w_i c_i$, $c_i \in \{\text{topicality, scope, understandability, novelty,}\}$

Model 4: $\text{Relevance} = \text{Understandability} + \text{Understandability} \times \sum w_i c_i$, $c_i \in \{\text{topicality, reliability, scope, novelty}\}$

where w_i is the weight of each criterion.

In this set of models, we ignore the higher order of interactions, although such interactions are theoretically possible. The pursuit for higher order interaction is warranted only if we could find the first order interaction effect.

When would such interaction model be valid? It seems reasonable to expect when a document violates one criterion bluntly, an elimination by aspect process will be initiated and a document being dismissed immediately. If the three prerequisites are at

lease marginally met, the interaction effect might be much weaker. However, such contemplation is subject to empirical test.

4. Methodology

In order to test the proposed models, a survey method was used followed by rigorous psychometric analysis. Structural equation modeling as a psychometric analysis method is a well established and dominant quantitative data analysis method in psychology, sociology, education, marketing, information systems research, education, and many other disciplines. It is particularly suitable for studying relationship among psychological perceptions which are not directly observable to researchers. Since no prior relevance research follows such methodology, we will briefly introduce the methodology and point to key references when appropriate.

4.1. Instrument Development

In designing a survey study for psychometric analysis, each construct which is typically a psychological perception (e.g., topicality, relevance) is measured by multiple questions which are known as items or instrument. Such questions are typically conceptually similar yet different and reflect different ways of the manifestation of the unobservable construct. The purpose of using multiple items to measure a construct is to extract the latent meaning underlying all these measurable items (refer to Nunnally and Bernstein 1994 for detailed treatment). Because this is the first study in relevance research (to our knowledge) that uses psychometric instrument, all items are self-developed based on the definition of these concepts and the real user's comments reported in the literature. Items are constructed as 7-point Likert

scale (Tang et al 1999). For example, one question to measure relevance is “this document is helpful to solve my problem at hand” (1—strongly disagree, 7—strongly agree).

To ensure that items do reflect the intended construct, the face and content validity is checked first. Face validity is the degree that a question for a construct appears to measure what it is proposed to measure. And content validity is the degree that questions for a construct have a representative coverage of manifestations for the intended construct. The questions we used are to a large degree the rephrasing of similar concepts proposed in the literature. This provides the basis for content validity. Moreover, two staffs and four Ph.D. students in the related researcher field are invited to discuss the phrasing of the questions to ensure that at least the questions are valid on face and content (Nunnally and Bernstein 1994, chapter 3). Minor changes were made based on the feedback.

Questions designed to measure a construct should not be measuring another construct. Item sorting is such a method to ensure the pertinence of each question to its own construct (refer to Moore and Benbasat 1991). Item sorting has three rounds. In the first and second round, eight judges in two groups were invited to sort the questions into as many groups as they deem appropriate. No construct definitions or construct-question relationships were known to the judges. In the last round, four judges were asked to match each question to a construct definition which is now known to them. The inter-judge agreement was measured with Kappa score. The Kappa score of the three group's sorting are all above 0.7 which is above the suggested level. We therefore concluded that our questions are good enough for the following survey. Questions for this study are listed in Appendix A.

4.2. Data Collection

The survey was carried out in two steps: a pilot study and a main study. The purpose of the pilot test is to quantitatively test the questionnaire quality and construct validity on small-scale data. Both the pilot study and the main study were carried out in a computer lab. Subjects are undergraduate and graduate students in a major university in Southeast Asia, who were invited by announcement in several courses. The data collection lasted for three weeks. In the survey, subjects were asked to search documents on an assigned topic of “the health and safety of using mobile phone”, which is common debated healthcare topic in relation to people’s life. They were asked to provide their demographics, their prior knowledge on the topic, and then search the internet and list at least five documents that are at least marginally relevant after reading. Then they evaluated two documents which were randomly assigned by the research. Subjects generally took 30 to 60 minutes to finish the whole process and SGD\$10 were given out as a reward. Both the pilot and the main study were done in this fashion.

5. Data Analysis and Result

5.1. Pilot study

In the pilot study, 76 valid questionnaires were collected with a sample of 38 students. Exploratory factor analysis (EFA, also see Nannally and Bernstein 1994) was conducted to test the convergent and discriminant validity of instrument. Convergent validity means that all questions intended to measure a construct do reflect that construct. Discriminant validity means that a question does not reflect an unintended construct. For pilot study, exploratory factor analysis with principal component

analysis was used to extract the number of factors that naturally emerge from the data.

Minor principal component with eigenvalue less than 1 are ignored as a convention.

Table 5 reports the principal component analysis result with Varimax rotation using SPSS10. The recency is not included as it is measured with a single item. The item of Scope4 was dropped because of loading problem. The reason may due to that this item was described as reverse meaning. The rest items loaded on the intended construct with loading value greater than 0.5, and the loading on unintended construct was less than 0.4 (refer to Hair et al. 1998). Thus, the remaining items showed appropriate validity.

Table 5. Factor Loading Table

	Component								
	1	2	3	4	5	6	7	8	9
TOPIC1	.577	.244	.263	.038	.244	.369	.081	.314	.199
TOPIC2	.688	.118	.104	.078	.112	.367	.163	.283	.093
TOPIC3	.815	.122	.077	.243	-.061	.240	.131	.219	.000
TOPIC4	.749	.261	.009	.094	-.010	.286	.186	-.075	.040
RELIA1	.315	.612	.086	.127	-.091	.398	-.109	.281	.074
RELIA2	.160	.846	.015	.160	-.077	.274	-.016	.222	.065
RELIA3	.133	.854	-.029	.291	-.040	.175	.015	.247	.043
RELIA4	.128	.836	-.020	.301	.049	.207	.033	.217	.077
SCOPE1	.034	.052	.889	-.133	-.031	.040	.129	-.145	-.024
SCOPE2	.156	.017	.874	-.075	-.109	-.061	.174	-.009	.071
SCOPE3	-.010	-.111	.749	.002	-.147	.327	.012	.239	.132
UNDER1	.135	.083	-.070	.915	.013	-.028	.029	.114	.099
UNDER2	-.036	.181	-.037	.860	-.060	.291	.047	-.083	-.004
UNDER3	.135	.313	-.097	.865	-.006	.108	.099	.098	-.012
UNDER4	.274	.310	-.066	.538	-.010	.254	.021	.253	.103
NOVEL1	.166	.064	.129	-.018	.685	.324	.182	-.097	-.238
NOVEL2	.077	.088	-.033	-.079	.837	.226	.188	-.169	-.106
NOVEL3	.047	-.163	-.145	-.011	.850	.020	-.058	.167	-.025
NOVEL4	-.169	-.049	-.192	.047	.771	-.178	.040	.105	-.182
RELEV1	.145	.170	.098	.172	.202	.851	.090	.167	.099
RELEV2	.207	.108	.069	.060	.077	.856	.249	.217	.037
RELEV3	.119	.185	.055	.129	-.008	.850	.065	.189	.155
RELEV4	.343	.272	-.062	.042	-.019	.765	.089	.057	-.016
RELEV5	.232	.219	.120	.190	.095	.780	.159	.119	-.079
ENJOY1	.051	.014	.075	.028	.093	.101	.881	.040	.081
ENJOY2	.083	.008	.138	.066	.028	.001	.858	.027	.172
ENJOY3	.038	-.016	.112	.070	.045	.112	.888	-.013	-.045
ENJOY4	.086	.039	.040	-.004	.049	.027	.925	.095	.000
ENJOY5	.160	-.039	-.043	.021	.045	.315	.810	.071	-.155

SOURC1	.195	.388	-.074	.152	-.016	.237	.082	.800	.039
SOURC2	.169	.288	-.153	.117	-.046	.262	.138	.810	.012
SOURC3	.102	.302	.255	.027	.140	.247	.048	.740	-.137
KNOWE1	.168	.194	-.075	.048	-.282	.049	.054	.026	.796
KNOWE2	.067	.053	.066	.050	-.164	.050	.012	.016	.897
KNOWE3	-.056	-.041	.124	.030	-.010	.055	.021	-.062	.871
Eigenvalue	10.770	4.385	3.550	2.755	1.924	1.722	1.522	1.260	1.141
Variance %	30.772	12.529	10.143	7.872	5.498	4.920	4.350	3.599	3.261
Cumulative Variance %	30.772	43.301	53.444	61.316	66.814	71.734	76.083	79.683	82.944

5.2. Main Study

In the main study, 162 valid questionnaires (with 81 students) were collected. The demographics of subjects are reported in Table 6.

Table 6. Demographics in Main Study

Gender	Male=100 (61.7%)		Female=62(38.3%)	
Education	Undergraduate=98 (60.5%)		Postgraduate=64 (39.5%)	
Age	Mean=24.3		S.D.=2.9	
Search engine	Google=93.8%	Yahoo=12.3%	Library=12.3%	Others=8.6%

5.2.1. Measurement Model

Following the methodological suggestion of Anderson and Gerbing (1989), before hypothesis testing, the first step of structural equation modelling is the measurement model which is used to further ensure the instrument quality. Unlike EFA, the measurement model analysis pre-specifies the construct-question correspondence but leave the constructs to freely correlate. Questions are expected to be highly correlated with the intended constructs only. Measurement model is analyzed with confirmatory factor analysis (CFA) using statistical package LISREL v8.51. Convergent validity is verified by the average variance extracted (AVE) of each item by the intended

construct, the composite factor reliability (CFR), and Cronbach's alphas (α) (Hair et al, 1998). The latter two measures how consistently questions of a construct correlate with each other. Table 7 reports the results of our measurement model.

According to Fornell and Larcker (1981), an AVE score above 0.5 indicates an acceptable level of convergent validity. Chin (1998) recommends the minimal requirement for alpha and CFR should be above 0.7. As shown in table 7, these criteria are all satisfied. Thus, the convergent validity is ensured.

Table 7. Measurement Model

Construct	Item	Std. Loading	T-value	AVE	CFR	α
Topicality	TOPIC1	0.89	14.13	0.685	0.900	0.900
	TOPIC2	0.92	15.14			
	TOPIC3	0.79	11.89			
	TOPIC4	0.69	9.69			
Reliability	RELIA1	0.74	10.69	0.722	0.912	0.908
	RELIA2	0.86	13.49			
	RELIA3	0.90	14.35			
	RELIA4	0.89	14.23			
Scope	SCOPE1	0.61	7.63	0.528	0.767	0.768
	SCOPE2	0.70	9.01			
	SCOPE3	0.85	11.12			
Understand-ability	UNDER1	0.90	14.55	0.771	0.930	0.927
	UNDER2	0.92	15.10			
	UNDER3	0.93	15.58			
	UNDER4	0.75	11.11			
Novelty	NOVEL1	0.76	10.95	0.532	0.814	0.808
	NOVEL2	0.93	14.60			
	NOVEL3	0.57	7.36			
	NOVEL4	0.60	7.56			
Relevance	RELEV1	0.92	15.65	0.751	0.938	0.934
	RELEV2	0.86	13.62			
	RELEV3	0.86	13.42			
	RELEV4	0.87	13.91			
	RELEV5	0.82	12.48			
Enjoyment	ENJOY1	0.91	14.83	0.759	0.940	0.931
	ENJOY2	0.84	13.01			
	ENJOY3	0.90	14.34			
	ENJOY4	0.91	14.95			
	ENJOY5	0.79	11.90			

Source Credibility	SOURC1	0.96	16.29	0.819	0.931	0.928
	SOURC2	0.92	15.08			
	SOURC3	0.83	12.92			
Prior Knowledge	KNOWE1	0.69	8.77	0.552	0.786	0.777
	KNOWE2	0.83	10.55			
	KNOWE3	0.70	8.95			

For discriminant validity, we compared the χ^2 between the original model and a constrained model which sets the correlation of two constructs to 1.0. A significant change in χ^2 indicates that the two constructs are different. Pair-wise constrained test was done for every pair of constructs and the results show the significant difference in chi-square. Hence, the discriminant validity is confirmed. Testing results are listed in Appendix B.

Another way of checking discriminant validity is that the inter-factor correlation should be less than the square root of AVE (Fornell and Larcker 1981). The correlation among constructs is reported in table 8.

Table 8. Construct Correlation Table

	KNO	REC	SOU	TOP	RELI	SCO	UND	NOV	RELE	ENJ
KNO	0.74									
REC	0.13	1.00								
SOU	0.03	0.27	0.91							
TOP	-0.03	0.03	0.37	0.82						
RELI	0.08	0.19	0.70	0.37	0.85					
SCO	0.02	0.04	0.11	0.50	0.08	0.73				
UND	0.15	0.13	0.02	0.23	0.09	0.27	0.88			
NOV	-0.17	0.03	0.32	0.58	0.28	0.27	-0.16	0.73		
RELE	-0.03	0.11	0.28	0.78	0.32	0.47	0.21	0.56	0.87	
ENJ	-0.08	0.15	0.22	0.47	0.18	0.24	0.47	0.39	0.53	0.87

Table 9 reports the overall measurement model fit. Most of the indices are above the recommended standard, except NFI and GFI and AGFI (see Nummally and Bersyein 1994, McKnight et al. 2002, Chin and Todd 1995, and Segars and Grover 1993).

Table 9. Overall Measurement Model Fit

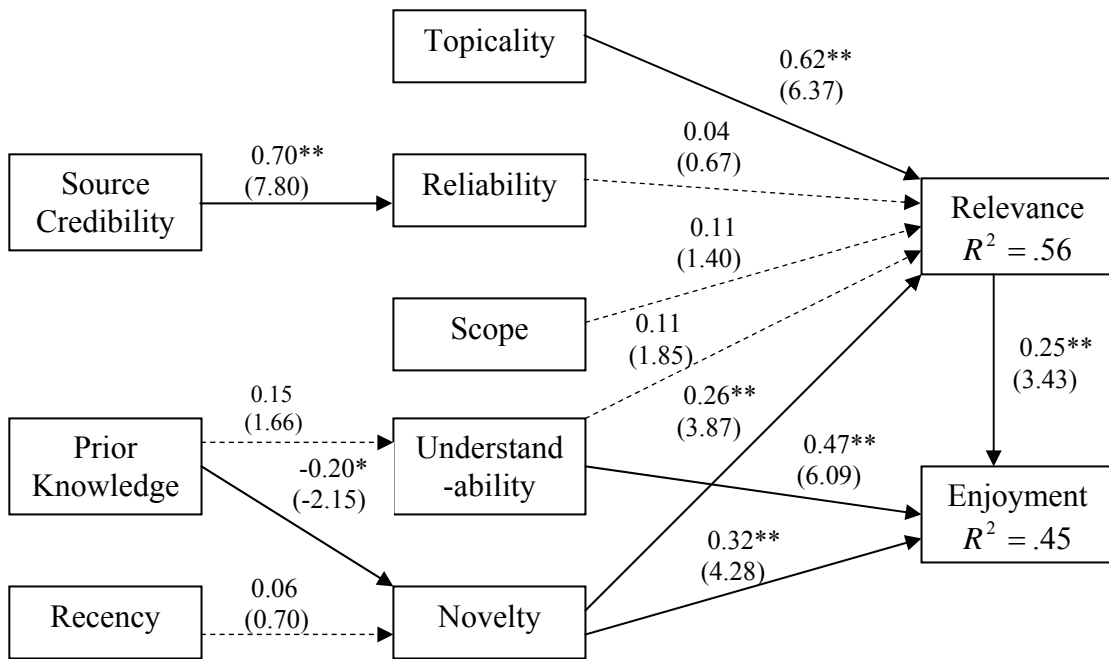
Fit Index	Proposed Model
Chi-Square (χ^2)	818.68 (p=0.00)
Degree of Freedom (dof)	550
χ^2 / dof	1.49
Root Mean Square Error of Approximation (RMSEA)	0.055
Normed Fit Index (NFI)	0.82
Non-normed Fit Index (NNFI)	0.91
Comparative Fit Index (CFI)	0.92
Incremental Fit Index (IFI)	0.92
Relative Fit Index (RFI)	0.80
Goodness of Fit Index (GFI)	0.78
Adjusted Goodness of Fit Index (AGFI)	0.73

5.2.2. Structural Model

Since the measurement model was solid, we proceeded to hypothesis testing. Hypothesis testing is done by creating a structural equation model in LISREL, which specifies both item-construct correspondence and construct-construct causal relationship. The coefficients are then solved with maximum likelihood estimation. The result is summarized in Figure 2. Before we draw conclusion on the hypotheses, the modelling fitting should be checked first. The result indicates low yet acceptable model fit. GFI, RFI, and NFI, though low, should be considered acceptable for newly developed instrument. The rest indices are all above the recommended level (refer to Nunnaly 1994 for a detailed treatment on methodology).

Because model fitting is acceptable, we can interpret the result as following. In the relevance test, we find that topicality and novelty are significant. However, the other three are not supported by our data set. In the enjoyment (affective relevance) test, all

the proposed hypotheses are supported. For the important criteria's antecedent test, source credibility is significantly related to document reliability and prior knowledge is also significantly negative associated with document novelty, while other two are not support in our test. We will discuss the results in next section.



$\chi^2 = 913.04$, $dof = 573$, $p = 0.0000$, $RMSEA = 0.061$, $NFI = 0.81$, $NNFI = 0.90$, $CFI = 0.91$, $IFI = 0.91$, $RFI = 0.80$, $GFI = 0.77$, $AGFI = 0.73$.

* $p < 0.05$, ** $p < 0.01$, → Supported, -- > Not supported

Figure 3. Standardized LISREL Solution

5.3. Testing of Alternative Models

Because LISREL does not support the test of interaction effect, we resort to moderated multiple regression (MMR) for alternative models. Multiple questions are averaged to produce the factor score. Such factor scores are then normalized. Multiplication terms

(e.g. Topicality*Novelty) are produced as the product of the normalized scores. The data analysis is done by both SPSS 10 and Minitab 4.0. As SPSS can not produce the formal normality test result, we use Minitab to report this result.

If the linear additive model is adequate, we expect to see normal residuals and decent R^2 . If the alternative models are better, we expect that the corresponding R^2 will significantly larger than the linear additive model (Model 1). Moreover, if the interaction effects do exist, we expect the interaction terms to be significant. The following report the results of four models, and the normal plot of residual is list in Appendix C.

Model1: $Relevance = \sum w_i c_i$, $c_i \in \{topicality, reliability, scope, understandability, novelty\}$, where w_i is the weight of each criterion.

Model 1 Linear Additive Model

Hypothesis	Coefficients	T-value	P-value	VIF	R^2	Normality of Residual P-value
Topicality \rightarrow Relevance	0.607	8.491	0.000	1.775	58.9%	0.295
Reliability \rightarrow Relevance	0.076	1.066	0.288	1.200		
Scope \rightarrow Relevance	0.114	1.931	0.550	1.206		
Understandability \rightarrow Relevance	0.092	1.380	0.170	1.202		
Novelty \rightarrow Relevance	0.178	2.822	0.005	1.440		

Model 2: $Relevance = Topicality + Topicality \times \sum w_i c_i$, $c_i \in \{ reliability, scope, understandability, novelty, \}$, where w_i is the weight of each criterion.

Model 2 Moderated Multiplicative Model

(Topicality is the prerequisite criterion)

Hypothesis	Coefficients	T-value	P-value	VIF	R^2	Normality of Residual P-value
Topicality \rightarrow Relevance	0.226	1.356	0.177	9.834	59.7%	0.363
Topicality*Reliability \rightarrow Relevance	0.017	1.248	0.214	4.029		

Topicality* Scope → Relevance	0.016	1.504	0.135	2.484		
Topicality* Understandability → Relevance	0.015	1.106	0.271	4.074		
Topicality* Novelty → Relevance	0.037	3.256	0.001	3.035		

Model 3: $Relevance = Reliability + Reliability \times \sum w_i c_i$, $c_i \in \{ topicality, understandability, novelty, scope$, where w_i is the weight of each criterion.

Model 3 Moderated Multiplicative Model
(Reliability is the prerequisite criterion)

Hypothesis	Coefficients	T-value	P-value	VIF	R^2	Normality of Residual P-value
Reliability → Relevance	-0.829	-6.619	.000	3.729	59.7%	0.286
Reliability * Topicality → Relevance	0.119	8.696	.000	4.234		
Reliability* Scope → Relevance	0.012	1.192	.235	1.731		
Reliability* Understandability → Relevance	0.018	1.441	.152	2.661		
Reliability* Novelty → Relevance	0.032	2.774	.006	2.485		

Model 4: $Relevance = Understandability + Understandability \times \sum w_i c_i$, $c_i \in \{ topicality, reliability, novelty, scope \}$, where w_i is the weight of each criterion.

Model 4 Moderated Multiplicative Model
(Understandability is the prerequisite criterion)

Hypothesis	Coefficients	T-value	P-value	VIF	R^2	Normality of Residual P-value
Understandability → Relevance	-0.638	-6.029	0.000	2.783	55.4%	0.590
Understandability * Topicality → Relevance	0.010	7.671	0.000	3.327		
Understandability * Reliability → Relevance	0.016	1.253	0.212	2.658		
Understandability * Scope → Relevance	0.014	1.441	0.152	1.803		
Understandability * Novelty → Relevance	0.033	2.978	0.003	1.680		

6. Discussion and Implications

6.1. Discussion

The object of this study is to identify and confirm a set of key relevance judgement criteria. Five such criteria were identified based on Grice's maxims and prior literature. And some related variables were also investigated, namely recency, prior knowledge, and source credibility. We believe such variables serve as antecedents of the relevance criteria.

Based on the EFA of pilot data and the measurement model of the main study data, we show that these constructs do have discriminant validity, i.e., they are distinct concepts. For each construct, different terms with minor difference in meaning should be unified. For example, Relevance can be termed value, satisfaction, pertinence, helpfulness, or intention to use, yet all such terms loaded on a single factor, indicating the shared meaning. Both exploratory and confirmatory factor analysis offers the way to reduce the vast number of criteria identified in prior literature.

In addition, enjoyment as another perspective of relevance is proposed and tested. Meanwhile, alternative models of relevance judgment are also proposed to explore and test whether interaction effects exist among the factors and which model is better to interpret user's relevance judgment in our data. Table 10 summarizes the hypothesis testing results.

Table 10. Hypothesis Testing Result

Hypothesis	Result
H1: Topicality → Relevance (+)	Supported
H2: Reliability → Relevance (+)	Not supported
H3: Scope → Relevance (+)	Not supported

H4: Understandability → Relevance (+)	Not supported
H5: Novelty → Relevance (+)	Supported
H6: Source Credibility → Reliability (+)	Supported
H7: Prior Knowledge → Understandability (+)	Not supported
H8: Prior Knowledge → Novelty (-)	Supported
H9: Recency → Novelty (+)	Not supported
H10: Relevance → Enjoyment (+)	Supported
H11: Understandability → Enjoyment (+)	Supported
H12: Novelty → Enjoyment (+)	Supported

6.1.1. Relevance

The result shows 56% of relevance variance is explained by all criteria, and topicality and novelty are statistically significant to relevance judgment. The standardized coefficient (0.62) shows that topicality is the major criteria in relevance judgment, which confirms the early suggestion that topicality is centre part of relevance in IR (e.g. Froehlich, 1994). The results also show that novelty is the second important relevance criteria (0.26). We conclude that novelty is the next key relevance criterion beyond topicality for relevance judgment. Novel and new document is potentially relevant in solving user's current problem, while already known document is not, it may only serve as a tool for reminder. This result further confirms early observation of the insufficient role of topicality and reveals the involvement of other important criteria beyond topicality in making relevance judgment (e.g., Boyce, 1982).

The result also shows that the other three proposed factors (reliability, understandability, and scope) are not supported by the data. It is too hasty to conclude that these factors are really unimportant. Relevance has been regarded as context-dependent (Vakkari 2000, Bateman 1998, Barry and Schamber 1998). The non-significance might be due to the design of the survey and the artificial context. A plausible explanation is that this survey asked participant to list documents that they

perceive as least marginally related, and then evaluate two of them. Such procedure, when looking back, introduces bias in the evaluated document. The evaluated document could be expected to be largely reliable and understandable. The average score of reliability and understandability is about 5.5 and 5.8 respectively. The high reliability and understandability makes these two variables less a concern in relevance judgement, hence nonsignificant. As a limitation of this study, the significance of these two constructs should be further tested with an improved survey procedure. Construct scope has lower average (about 4.0), its nonsignificance is less likely a result of survey design. It seems that readers consider scope an optional premium in relevance judgment.

For those “peripheral factors” which are not properties of the document content, we found that source credibility is very important to content reliability, and prior knowledge affects novelty. However, recency does not affect novelty in this context, presumably because the information on safety of hand phone use is not time-dependent. Prior Knowledge does not affect understandability because the understandability of documents evaluated is very high, and the significance might change as the task context changes.

6.1.2. Affective Relevance

In addition to the main purpose of this study, this paper measures and tests the hedonic perspective of relevance, and attempts to clarify the relationship between affective relevance and situational relevance. The result shows that document understandability, novelty, and situational relevance are significant to enjoyment as we hypothesized. The significant effect of situational relevance on enjoyment is

consistent with the conclusion of product consumption in marketing and psychology literature. Thus, this study provides theory and data support to the proposition that affective relevance (enjoyment) should be treated as a separate and independent construct, which is influenced by situational relevance. In this study, large enjoyment variance are explained ($R^2=45\%$). Although as a pilot study we do not conclude that the affective relevance is only affected by our proposed three variables, this study provides a first and useful clue for user's affective relevance judgment research. Further study is needed to identify other criteria (if so) in this field.

6.1.3. Alternative model

In the alternative models test, we can see that the proposed three moderated multiplicative models are no better than the linear additive model (model1). First, comparing the R^2 of our proposed four model, the linear additive model accounts for 58.9%, and the other three moderated multiplicative are 59.7% (model2), 59.7% (model3), and 55.4% (model4) respectively. Hence, there is no significant difference in R^2 ($\Delta R^2 < 5\%$), and the model 4 is a slight lower than others. Second, from normal probability plot of residual (see Appendix C), we can see that all residuals of the four models are following the normal distribution, which suggests that all four model are qualified to interpret the judgment. Although in multiplicative models the interaction effects do exist among the factors, not all moderate effects are significant. In model 2, only one construct (Topicality* Novelty) is supported (T-value=3.256). In model 3, three constructs (Reliability, Reliability* Topicality and Reliability* Novelty) are supported, while the coefficient of the relationship between Reliability and Relevance is significantly negative, which is not our expected. In model 4, three constructs (Understandability, Understandability* Topicality and Understandability *

Novelty) are supported by our data. However, the coefficient of the relationship between Understandability and Relevance is also significantly negative.

Since there is no significant difference between additive and multiplicative model in interpreting relevance judgment in our data, we keep that the original model (the linear additive model) as our research model for relevance judgment in this study. And our proposed alternative models (moderated multiplicative model) are needed to further investigate and test under other situations.

6.2. Limitation of This Study

As the first confirmatory study in this area following a psychometric procedure, we shall point out the key limitations before we draw any implication. First, the bias in the selection of document for relevance evaluation is a critical limitation in the hypothesis testing. The conclusions and implications drawn from this study are applicable only to documents that bear minimum topicality already.

Second, we ignore the order effect among the documents in user's document reading and questionnaire answering process, which is theoretical possible suggested by prior study (Eisenberg and Barry, 1988). Thus, it may undermine the reliability of results of the testing results.

Third, the model fit is not good enough, which suggests the questionnaire quality need to be improved. These limitations serve as the directions for future study.

Finally, since relevance is context-dependent, a typology of different contexts and the importance of relevance criteria in those contexts are interesting future research topic.

6.3. Implications

6.3.1. Theoretical Implications

The theoretical implication of this study is multi-fold. First of all, based on the Grice's theory of communication, this study proposes a theory-based model and identifies five most important relevance criteria for relevance judgment: topicality, novelty, reliability, understandability and scope. These five criteria and their antecedents have covered most of factors that would affect relevance in prior literature. Thus, this study provides the first theory-driven model for user-oriented relevance judgment.

Second, as this is the first study that uses psychometric instrument in relevance research, all instruments are self-developed based on the definitions of the concepts and the real user's comments reported in the literature. Based on the EFA and measurement model test, the result shows all instruments are qualified to model test. Thus, this study provides the first set of instruments for further test in this field.

Third, this study makes the first attempt to use confirmatory study to verify the important relevance criteria. In this pursue, it confirms early observation that topicality is the centre part of relevance in IR (e.g. Froehlich, 1994). In addition, it suggests novelty as the next most important criteria in relevance judgement.

If we assume that reliability, understandability, and scope are not the major concerns when a reader consumes a document, topicality and novelty become the central concern. The prominent role of topicality and novelty provides insight to the concept of relevance. Based on them, the concept of relevance can be depicted with four quadrants delineated by topicality and novelty (Figure 4). The low-to-high classification of a document on each dimension is of course a simplification. In real world, it is more likely a continuum. Nevertheless, this tabulation offers insights in our understanding of relevance.

		Novelty	
		Low	High
Topicality	Low	Irrelevant	Potentially relevant
	High	Tool	Informative

Figure 4. Relevance quadrants

In the low topicality - low novelty quadrant, a document is neither on topic, nor new to the user. It is thus most likely to be dismissed as *irrelevant*. In the high topicality – low novelty quadrant, a document is on topic but already known to the user. Imagine if we are going to write another paper to address the limitations of this study, reference (Saracevic 1975) is a classical paper on the topic of subjective relevance and has topicality. However, the authors are familiar with the content already. We may still treat it as relevant because we need to reference to it or to check some concepts defined, or to quote some sentences. Such a document is useful and relevant to our research, yet it is used as a *tool*. The low topicality – high novelty quadrant deals with documents that are unclear in topicality, yet provides certain new information that attracts the user’s attention. As Harter (1992) points out, there is no absolutely fixed information need in a search process. Information need can be multiple and vague. The interaction of new information in a document and the current cognitive state helps to

clarify the information need and create future search topic. Consequently, a document might be regarded as *potentially relevant* because the user anticipates its future value rather than the current value. Finally, the high topicality – high novelty quadrant possesses the ideal documents. They might help the user clarify information need, offer new problem solution or new evaluation method for different problem solutions. In each case, they are *informative*.

Forth, this study attempts to clarify the relationship between affective relevance and situational relevance. Based on the psychological literature, this study conceptualized situational relevance as utilitarian perspective of relevance and affective relevance as hedonic perspective of relevance. The test result is consistent with the conclusion of marketing and psychology research. Thus, this study provides theory and data support that affective relevance should be a separate and independent concept from situational relevance, and it is influenced by situational relevance.

Finally, this study provides two kinds of models: linear additive model and moderated multiplicative models in interpreting relevance judgment for further study.

6.3.2. Practical Implications

This study is also useful to information retrieval systems design. Decades of research effort have been made to better capture topicality. What this study suggests that the next power house of IR systems design might be the quantification of novelty. How to capture a reader's cognitive state before document evaluation? How to measure the novelty of a document against such cognitive state? How to combine novelty and

topicality into an overall relevance score? While this study does not offer any answer to these questions, we do suggest that effort in this direction will be rewarding.

7. Conclusion

Relevance has been acknowledged as a fundamental and centre concept in information science. In order to better understand the meaning of relevance, many exploratory studies have done to identify the criteria that would affect user's relevance judgment. However, few study attempts to verify these criteria and no confirmatory study is available to test the relationships between relevance and the criteria.

Based on Grice's communication theory, this study proposes a conceptual model and identifies five most important criteria: topicality, novelty, reliability, understandability, and scope for relevance judgment. And these criteria and their antecedent cover most of factors that would affect relevance in prior literature. Therefore, we provide the first theory-driven model for relevance judgment research. Based on our model, we verify the various factors identified in prior studies, and attempt to provide a theoretical foundation for them. As an additional study, affective relevance as another perspective of relevance is also proposed in this model, which is independent from situational relevance. Meanwhile, three alternative models are proposed to examine the interaction effect among the criteria.

Our result shows that topicality and novelty are supported. This result confirms the early suggestion that topicality is the centre part of relevance. And it also reveals novelty is the next key relevance criteria beyond topicality. However, the other three

criteria are not significant in our data. It is too harsh to conclude that these factors are unimportant in general. The non-significant might be due to the design of survey. Further test is needed to test the importance of each of criterion under certain situations. As an additional test, the result supports our proposition that situational and affective relevance are separated and correlated. For the alternative model test, both linear additive and moderated multiplicative model are equally good (comparing their R^2 and normality of residual) in interpreting user's relevance judgment in this test. Further study is called for to investigate and test under other situations.

References

- [1] Acker, M. and P. McReynold (1967), "The Need for Novelty: A Comparison of Six Instruments," *Psychological Record*, 17, pp.177-182.
- [2] Adelberg, A. H. (1979), "A Methodology for Measuring the Understandability of Financial Report Message," *Journal of Accounting Research*, 17(2), pp.565-592
- [3] Anderson, J. C. and D. W. Gerbing (1998), "Structure Equation Modeling in Practice: A Review and Recommended Two-step Approach," *Psychological Bulletin*, 103(3), pp. 411-423
- [4] Aritandi, S. (1973), "Information Concepts and Their Utility," *Journal of the American Society for Information Science*, 24, pp. 242-245.
- [5] Atlas, J. and S. Levinson (1981), "It-Clefts, Informativeness and Logical Form," In: P. Cole (ed.), *Radical Pragmatics*, New York, AP.
- [6] Bach, K. and R. Harnish (1979), *Linguistic Communication and Speech Acts*, MIT Press.
- [7] Barry, C. L. (1994), "User-defined Relevance Criteria: An Exploratory Study," *Journal of the American Society for Information Science*, 45(3), pp.149-159.
- [8] Barry, C., and L. Schamber (1998), "Users' Criteria for Relevance Evaluation: A Cross-Situational Comparison," *Information Processing & Management*, 34, pp. 219-236.
- [9] Bateman, J. (1998), "Changes in Relevance Criteria: A Longitudinal Study," *Proceedings of the 61st Annual Meeting of the American Society for Information Science*, 35, pp. 23-32.
- [10] Batra, R. and O.T. Ahtola (1990), "Measuring the Hedonic and Utilitarian Sources of Consumer Attitude," *Marketing Letters*, 2(2), pp.159-170.

- [11] Belkin, N.J., R.N. Oddy, and H.M. Brooks (1982), "ASK for Information Retrieval," *Journal of Documentation*, 38(2), pp.61-71
- [12] Bookstein, A. (1979), "Relevance," *Journal of the American Society for Information Science*, 30(5), pp. 269-273.
- [13] Borlund, P. (2003), "The Concept of Relevance in IR," *Journal of the American Society for information Science and Technology*, 54(10), pp.913-925
- [14] Boyce, B. (1982), "Beyond Topicality: A Two Stage View of Relevance and the Retrieval Process," *Information Processing and Management*, 18(3), pp. 105-109.
- [15] Britton, B.K., A. Piha, J. Davis, and E. Wehausen (1978), "Reading and Cognitive Capacity Usage: Adjunct Question Effects," *Memory and Cognition*, 6, pp.266-273.
- [16] Brown, R.D., L.A. Braskamp, & D.L.Newman (1978), "Evaluator Credibility and Acceptance as a Function of Report Styles: Do Jargon and Data Make a Difference?" *Evaluation Quarterly*, 2(2), pp.331-341.
- [17] Campbell, I., and K. van Rijsbergen, (1996), "The Ostensive Model of Developing Information Needs," In P. Ingwersen & N.O. Pors (Eds.), *Proceedings of CoLIS 2, second international conference on conceptions of library and information science: Integration in perspective*; Copenhagen, pp. 251–268.
- [18] Cattell, R. B. (1975), *Personality and Motivation: Structure and Measurement*, New York: Harcourt, Brace & World. Inc.
- [19] Chase, W.G. and Herbert, A. S. (1973), "Perception in Chess," *Cognitive Psychology*, 4, pp. 55-81
- [20] Chin, W. W. (1998), "The Partial Least Squares Approach to Structural Equation Modeling," In *Modern Methods for Business Research*, G. A. Marcoulides (ed.), Mahwah, NJ: Lawrence Erlbaum Associate, pp.295-336.

- [21] Chin, W. W., and P. A. Todd (1995), "On the Use, Usefulness, and Ease of Use of Structural Equation Modeling in MIS Research: A Note of Caution," *MIS Quarterly* 19(2), pp. 237-246.
- [22] Choi, Y. and E. M. Rasmussen (2002), "Users' Relevance Criteria in Image Retrieval in American History," *Information Processing and Management*, 38, pp.695–726
- [23] Cool, C., N. J. Belkin, and P. B. Kantor (1993), "Characteristics of Texts Affecting Relevance Judgments," In M.E, Williams (Ed.), *Proceedings of the 14th National Online Meeting*, pp.77-84, Medford, N J: Learned Information.
- [24] Cooper, W.S. (1971), "A Definition of Relevance for Information Retrieval," *Information Storage and Retrieval*, 7(1), pp.19-37
- [25] Cosjin, E. and Ingwersen, P. (2000), "Dimensions of Relevance", *Information Processing & Management*, 36(4), pp. 533-550.
- [26] Cuadra, C.A., and Katter, R.V. (1967a), *Experimental Studies of Relevance Judgments: Final Report. I: Project Summary* (NSF Report NO. TM-3520/001/00). Santa Monica, CA: System Development Corporation.
- [27] Cuadra, C.A., and Katter, R.V. (1967b), "Opening the Black Box of 'Relevance'," *Journal of Documentation*, 23(4), pp.291-303.
- [28] Dervin, B. (1983a), "Information as a User Construct: The Relevance of Perceived Information Needs to Synthesis and Interpretation," In S.A. Ward & L. J. Reed (Eds.), *Knowledge structures and use: Implications for synthesis and interpretation*, pp.153-184, Philadelphia: Temple University Press.
- [29] Dervin, B. (1983b), *An Overview of Sense-making Research: Concepts, Methods and Results to Date*. Paper presented at the International Communication Association Annual Meeting, Dallas, TX.

- [30] Dervin, B., and M. Nilan (1986), "Information Needs and Uses," *Annual Review of Information Science and Technology*, 21, pp.3-33.
- [31] Dessalles, J. L. (1998), "Altruism, Status, and the Origin of Relevance," In J. R. Hurford, M. Studdert-Kennedy & C. Knight, *Approaches to the Evolution of Language: Social and Cognitive Bases*, Cambridge University Press.
- [32] Dwyer, J. (1999), *Communication in Business: Strategies and Skills*, Prentice Hall, Sydney.
- [33] Dybkjaer, L., N. Bernsen, and H. Dybkjaer (1998), "A Methodology for Diagnostic Evaluation of Spoken Human Machine Dialogue," *International journal of human- computer studies*, 48(5), pp.605-626.
- [34] Eisenberg, M., and C. Barry (1986), "Order Effects: A preliminary Study of the Possible Influence of Presentation Order on User Judgments of Document Relevance," *Proceedings of the 49th Annual Meeting of the American Society for Information Science*, 23, pp.80-86.
- [35] Elsbach, K. D. and G. Eloffson, (2000), "How the packaging of decision explanations affects perceptions of trustworthiness," *Academy of Management Journal*, 43(1), pp. 83-89
- [36] Farley, F, and S. V. Farley (1967), "Extroversion and Stimulus-Seeking Motivation," *Journal of Consulting Psychology*, 31, pp. 215-216.
- [37] Fishbein, M. and I. Ajzen, (1975), *Beliefs, Attitudes, Intention and Behaviour: An Introduction to Theory and Research*, Reading, MA: Addison-Wesley.
- [38] Fitzgerald, M. A. and C. Galloway (2001), "Relevance Judging, Evaluation, and Decision Making in Virtual Library: A Descriptive Study," *Journal of the American Society for Information Science and Technology*, 52(12), pp. 989-1010.

- [39] Fornell, C and D. F. Larcker (1981), "Structure Equation Models: LISREL and PLS Applied to Customer Exist-Voice Theory," *Journal of Marketing Research*, 18(2), pp. 39-50
- [40] Froehlich, T.J. (1994), "Relevance Reconsidered: Towards an Agenda for the 21st Century: Introduction to Special Topic Issue on Relevance Research," *Journal of the American Society for Information Science*, 45, pp.124 –134.
- [41] Gaasterland, T., P. Godfrey, and J. Minker (1992), "An Overview of Cooperative Answering," *Journal of Intelligent Information Systems*, 1(2), pp.123-157
- [42] Greaney, V. and S.B. Neuman, (1990), *The Functions of Reading: A Cross-Cultural Perspective. Reading Research Quarterly*, 25, pp.172-195.
- [43] Green, R. (1995a), "Topical Relevance Relationships: I. Why Topic Matching Fails," *Journal of the American Society for Information Science*, 46, pp. 646–653.
- [44] Green, R. (1995b), "Topical relevance relationships: II. An Exploratory Study and Preliminary Typology," *Journal of the American Society for Information Science*, 46, pp.654-662
- [45] Greisdorf, H. (2003), "Relevance Thresholds: A Multi-stage Predictive Model of How Users Evaluate Information," *Information Processing & Management*, 39, pp. 403-423
- [46] Greisdorf, H., and B. O'Connor, (2003), "Nodes of Topicality: Modeling User Notions of On Topic Documents," *Journal of the American Society for Information Science and Technology*, 54(14), pp.1296–1304.
- [47] Grice, H. P. (1975), "Logic and Conversation," in P. Cole and J. Morgan, eds., *Syntax and Semantics*, 3, Academic Press, pp. 41-58.
- [48] Grice, H. P (1989), *Studies in the Way of Words*, Harvard University Press.

- [49] Hair, J F., R. E., Anderson, R. L. Tatham, and W. C. Black (1995), *Multivariate Data Analysis with Reading* (4th ed), Englewood Cliffs, NJ: Prentice Hall.
- [50] Harter, S. P. (1992), “Psychological Relevance and Information Science,” *Journal of the American Society for information Science*, 43(9), pp.602-615
- [51] Heit, E. (1997), “Knowledge and Concept Learning,”_In K. Lamberts & D. Shanks (Eds.), *Knowledge, Concepts, and Categories*, pp.7-41. Hove: Psychology Press.
- [52] Hersh, W (1994), “Relevance and Retrieval Evaluation: Perspective from Medicine”, *Journal of the American Society for Information Science*, , 45(3), pp. 201-206.
- [53] Hertzum, M., H.H.K. Andersen, V. Andersen, and C.B. Hansen (2002), “Trust in Information Sources: Seeking Information from People, Documents, and Virtual Agents,” *Interacting with Computers*, 14(5), pp.575–599.
- [54] Hirschman, E. C. (1980a). “Attributes of Attributes and Layers of Meaning,” In *Advances on Consumer Research*, 7, J.C. Olson, ed., Ann Arbor: Association for Consumer Research.
- [55] Hirschman, E. C. (1980b), “Innovativeness, Novelty Seeking, and Consumer Creativity,” *Journal of Consumer Research*, 7, pp.283-295
- [56] Hirsh S.G. (1999), “Children’s Relevance Criteria and Information Seeking on Electronic Resources,” *Journal of the American Society for information Science*, 50(14), pp.1265-1283
- [57] HjØrland, B. and F. S. Christensen (2002), “Work Tasks and Socio-Cognitive Relevance: A Specific Example,” *Journal of the American Society for Information Science and Technology*, 53(11), pp.960–965.

- [58] Holbrook, M. B. (1980), "Some Preliminary Notes on Research in Consumer Esthetics," in *Advances on Consumer Research*, 7, J.C. Olson, ed., Ann Arbor: Association for Consumer Research.
- [59] Ingwersen, P. (1984), "Psychological Aspects of Information Retrieval," *Social Science Information Studies*, 4(2&3), pp.83-95.
- [60] Johnson, J. R., R. A. Leitch, and J.Neter (1981), "Characteristics of Errors in Accounting Receivable and Inventory Audits," *Accounting Review*, 1(2), pp. 270-293.
- [61] Katzer, J. (1987), "User Studies, Information Science, and Communication," *Canadian Journal of Information Science*, 12(3/4), pp.15-30.
- [62] Kuhlthau, C.C. (1993), *Seeking Meaning: A Process Approach to Library and Information Science*, Norwood, NJ: Ablex Publishing.
- [63] Lancaster, F. W. (1968), *Information Retrieval Systems: Characteristics, Testing, and Evaluation*, New York: Wiley.
- [64] Lancaster, F. W (1979), *Information Retrieval Systems: Characteristics, Testing, and Evaluation*, New York: Wiley.
- [65] Lancaster, F. W., and A. J. Warner (1993), *Information Retrieval Today*, Arlington, VA: Information Resources Press.
- [66] Levitin, A. and T. Redman (1995), "Quality Dimensions of a Conceptual View," *Information Processing & Management*, 31(1), pp.81-88
- [67] Lewis, R. and W.H. Teale (1980), "Another Look at Secondary School Students' Attitude toward Reading," *Journal of Reading Behaviour*, 12, pp.187-201.
- [68] Maglaughlin, K. L. and H. Sonnewald (2002), "User Perspective on Relevance Criteria: A Comparison among Relevance, Partially Relevance, and Not-

- Relevance,” *Journal of the American Society for Information Science and Technology*, 53(5), pp. 327-342.
- [69] Maron, M.E. (1977), “On Indexing, Retrieval and the Meaning of About,” *Journal of the American Society for Information Science*, 28(1), pp.38-43
- [70] McGuire, W.J. (1969), *The Nature of Attitudes and Attitude Change*, In: G. Lindzey and E. Aronson (eds.), *The handbook of social psychology*, 108-141. New York: Rinehart and Winston.
- [71] McKenna, M.C., B.D. Stratton, M.C. Grindler, and S.J. Jenkins (1995), “Differential Effects of Whole Language and Traditional Instruction on Reading Attitudes,” *Journal of Reading Behaviour*, 27(1), pp.19-44.
- [72] McKinney, V. M., K., Yoon, and F. Zahedi (2002), “The Measurement of Web-customer Satisfaction: An Expectation and Disconfirmation Approach,” *Information System Research*, 13(3), pp. 296-315.
- [73] McNamara, D.S., E. Kintsch, N.B. Songer, W. Kintsch (1996), “Are Good Texts Always Better? Interactions of Text Coherence Background, Knowledge, and Levels of Understanding in Learning from Text,” *Cognition and Instruction*, 14 (1), pp. 1-43.
- [74] Miranda, S. and C. S. Saunders (2003), “The Social Construction of Meaning: An Alternative Perspective on Information Sharing,” *Information Systems Research*, 14(1), pp.87-108
- [75] Mizzaro, S. (1997), “Relevance: The Whole History,” *Journal of the American Society for Information Science*, 48(9), pp.810–832.
- [76] Möller, J., T. Müller-Kalthoff (2000), “Learning with Hypertext: the Impact of Navigational Aids and Prior Knowledge,” *German Journal of Educational Psychology*, 14 (2/3), pp.116-123.

- [77] Moore, G. C. and I. Benbasat (1991), "Development of an Instrument to Measure the perception adoption an information technology innovation," *Information Systems Research*, 2(3), pp.192-222
- [78] Neale, S. (1990), *Descriptions*, MIT Press.
- [79] Nummally, J.C., I.H .Bernstein (1994), *Psychometric Theory*, McGraw-Hill, New York.
- [80] Osgood, C. E., J.S. George, and H.T. Percy (1957), *The Measurement of meaning*, Urbana, Illinois: University if Illinois Press.
- [81] Park, H. (1997), "Relevance of Science Information: Origins and Dimensions of Relevance and Their Implications to Information Retrieval," *Information Processing & Management*, 33(3), pp.339-352
- [82] Park, T.K. (1993), "The Nature of Relevance in Information Retrieval: An Empirical Study," *Library Quarterly*, 63, pp.318–351.
- [83] Petty, R. E. and J. T. Cacioppo (1986), "The Elaboration Likelihood Model of Persuasion," In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology*, 19, pp.123-205. New York: Academic.
- [84] Petty, R., J. Priester, and D. Wegender (1994), "Cognitive Processes in Attitude Change," In R. Wyer & T. Srull (Eds.), *Handbook of social cognition*, pp.69-142, Hillsdale, NJ: Erlbaum.
- [85] Potelle, H. and J.F. Rouet (2003), "Effects of Content Representation and Readers' Prior Knowledge on the Comprehension of Hypertext," *International Journal of Human-Computer Studies*, 58, pp.327–345.
- [86] Raaheim, K. (1974), *Problem Solving and Intelligence*, Oslo: Universitetsforlaget.
- [87] Rees, A. M., and T. Saracevic, (1966), "The Measurability of Relevance," *Proceedings of the American Documentation Institute*, 3, pp.225-234.

- [88] Rees, A.M., and D.G. Schultz (1967), *A Field Experimental Approach to the Study of Relevance Assessments in Relation to Document Searching, I: Final report* (NSF Contract No. C-423), Cleveland: Case Western Reserve University.
- [89] Rooy, R.V. (2002), "Utility, Informativity, and Protocols," In: Bonanno et al (eds.), *Proceedings of LOFT 5: Logic and the Foundations of the Theory of Games and Decisions*, Torino.
- [90] Saracevic, T. (1970), "The Concept of 'relevance' in information science: A Historical Review," In Saracevic, T., *Introduction to information Science*, pp. 111-151, New York: R.R. Bowker.
- [91] 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(6), pp. 321-343.
- [92] Saracevic, T., "Relevance Reconsidered '96", In P. Ingwersen & N. Ole Pots (Eds.) *CoLIS2. 2nd International Conference on Conceptions of Library and Information Science*, 1996, pp. 201-218, Copenhagen, Denmark: Royal School of Librarianship.
- [93] Saracevic, T., and Kantor, P., "A Study of Information Seeking and Retrieving. III. Searchers, Searches, and Overlap", *Journal of the American Society for Information Science*, 39(3), 1988, pp.197-216
- [94] Schamber, L. (1991), "Users' Criteria for Evaluation in a Multimedia Environment," In J.-M. Griffiths (Ed.) *Proceedings of the 54th Annual Meeting of the American Society for Information Science*, 28, pp. 126-133.
- [95] Schamber, L. (1994), "Relevance and Information Behavior," In M. E. Williams (Eds.) *Annual Review of Information Science and Technology*, 29, pp. 33-48.

- [96] Schamber, L., and J. Bateman, (1996), "User Criteria in Relevance Evaluation: Toward Development of a Measurement Scale," *Proceedings of the 59th Annual Meeting of the American Society for Information Science*, 33, pp. 218-225.
- [97] Schamber, L., M. B. Eisenberg, and M. S. Nilan (1990), "A Re-examination of Relevance: Toward a Dynamic, Situational Definition," *Information Processing and Management*, 26(6), pp.755-776.
- [98] Segars, A.H. and V. Grover (1993), "Re-examining Perceived Ease of Use and Usefulness: A Confirmatory Factor Analysis," *MIS Quarterly*, 17(4), pp.517-525.
- [99] Sperber, D., and T. D. Wilson (1986), *Relevance: Communication and Cognition*. Cambridge, MA: Harvard University Press.
- [100] Spink, A., H. Greisdorf, and J. Bateman (1998), "From Highly Relevant to not Relevant: Examining Different Regions of Relevance," *Information Processing & Management*, 34, pp. 599–621.
- [101] Su, L. T. (1993), "Is Relevance an Adequate Criterion for Retrieval System Evaluation: an Empirical Study into the User's Evaluation," In S. Bonzi (Ed.) *Proceedings of the 56th Annual Meeting of the American Society for information Science*, 30, pp. 93-103.
- [102] Stokmans, M.J. (1999), "Reading Attitude and its Effect on Leisure Time Reading," *Poetics*, 26, pp.245-261
- [103] Swanson, D.R. (1986), "Subjective versus Objective relevance in Bibliographic Retrieval Systems," *Library Quarterly*, 56, 1986, pp. 189-398
- [104] Tang, R., W.M. Shaw Jr, and J.L. Vevea (1999), "Towards the Identification of the Optimal Number of Relevance Categories," *Journal of the American Society for Information Science*, 50, pp.254–264

- [105] Tang, R., and P. Solomon (1998) "Towards an Understanding of the Dynamics of Relevance Judgments: An Analysis of One Person's Search Behavior," *Information Processing & Management*, 34, pp.237–256.
- [106] Taylor, R.S. (1986), *Value-added Processes in Information Systems*, Norwood, NJ: Ablex.
- [107] Thompson, P.A., R.D. Brown, and J. Furgason (1981), "Jargon and Data do Make a Difference," *Evaluation Review*, 5(2), pp. 269-279.
- [108] Vakkari, P., and N. Hakala (2000), "Changes in Relevance Criteria and Problem Stages in Task Performance," *Journal of Documentation*, 56, pp.540–562
- [109] Wang, P. and 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), pp.115–133
- [110] Wang, P. and M. D. White (1999), "A Cognitive Model of Document Use during a Research Project. Study II. Decisions at the Reading and Citing Stages," *Journal of the American Society for Information Science*, 50(2), pp.98-144
- [111] Wang, R. Y. and D. M. Strong (1996), "Beyond Accuracy: What Data Quality Means to Data Consumer," *Journal of Management Information Systems*, 12(4), pp.5-34.
- [112] Wang, R. Y., and Y. L. Lee, L. Pipino, and D. M. Strong (1998), "Manage Your Information as a Product," *Sloan Management Review*, 39(4), pp. 95-105.
- [113] Wilson, P., "Situational Relevance", *Information Storage and Retrieval*, 9(8), 1973, pp.457-471
- [114] Wilson, T. D. (1984), "The Cognitive Approach to Information-seeking Behavior and Information Use," *Social Science Information Studies*, 4(2&3), pp.197-204

[115] Wurman, R. (1989), *Information Anxiety*. Doubleday, New York, NY.

[116] Zajonc, R. B., and H. Markus (1982), "Affective and Cognitive Factors in Preference," *Journal of Marketing*, 9, pp.123-131.

Appendix A

Instrument

Construct	Item	Description	Mean	S.D.
Topicality	TOP1	This document has a substantially amount of information about my current topic of interest.	4.86	1.56
	TOP2	The content of this document is substantially about my current topic of interest.	5.07	1.41
	TOP3	The topic of this document is substantially related to my current topic of interest.	5.44	1.34
	TOP4	The topic of this document is within the domain of my current topic of interest.	5.68	1.17
Reliability	RELI1	I think the content of this document would be accurate.	5.44	1.18
	RELI2	I think the content of this document would be consistent with the fact.	5.47	1.07
	RELI3	I think the content of this document would be true.	5.49	1.08
	RELI4	I think the content of this document would be reliable.	5.56	1.11
Scope	SCO1	The content of document is either too general or too specific for me.	4.09	1.33
	SCO2	The coverage of this document is either too abroad or too narrow for me.	3.92	1.42
	SCO3	This document gives either too many or too few details than what I expected.	4.22	1.60
	SCO4	The breadth and depth of this document is inappropriate for me. (dropped)	--	--
Understand-ability	UND1	Readers of my type should find this document very easy to read.	5.68	1.34
	UND2	I am able to follow the content of this document with little effort.	5.86	1.12
	UND3	The content of this document is easy to understand.	5.82	1.15
	UND4	After reading it, I am very clear about the main content of this document.	5.94	1.05
Novelty	NOV1	This document has a substantial amount of new information to me.	4.80	1.45
	NOV2	This document has a substantial amount of unique information that I come across for the first time.	4.46	1.63
	NOV3	The content of this document is different from what I have read before.	3.98	1.53
	NOV4	I have not read the content similar to this document before.	3.81	1.45
Relevance	RELE1	This document has a great value in meeting my need.	4.59	1.40
	RELE2	This document is satisfactory in meeting my need.	4.73	1.28
	RELE3	This document is very pertinent to my need.	4.65	1.30
	RELE4	This document is helpful to solve my problem at hand.	4.76	1.51
	RELE5	I would make use of this document.	5.04	1.56
Enjoyment	ENJ1	Reading this document gives me pleasure.	4.13	1.33
	ENJ2	Reading this document is a very nice time out.	4.34	1.33
	ENJ3	The content of this document is very interesting.	4.11	1.23
	ENJ4	Reading this document is truly enjoyable.	3.95	1.31
	ENJ5	I am absorbed by the content of this document.	4.16	1.44
Source Credibility	CRE1	Regarding the source of this document (e.g. author, publishing organization, or journal etc),	5.80	1.16

		I think the source is trustworthy.		
	CRE2	I think the source is reputable.	5.83	1.20
	CRE3	I think the source is authoritative.	5.55	1.35
Prior Knowledge	KNOW1	I know this topic very well.	3.64	1.12
	KNOW2	I am able to tell other much about this topic.	3.22	1.06
	KNOW3	I would consider myself an expert in this topic.	2.31	1.26
Recency	REN1	The publication date of this document is considered recent in this topic.	4.62	1.67

Appendix B

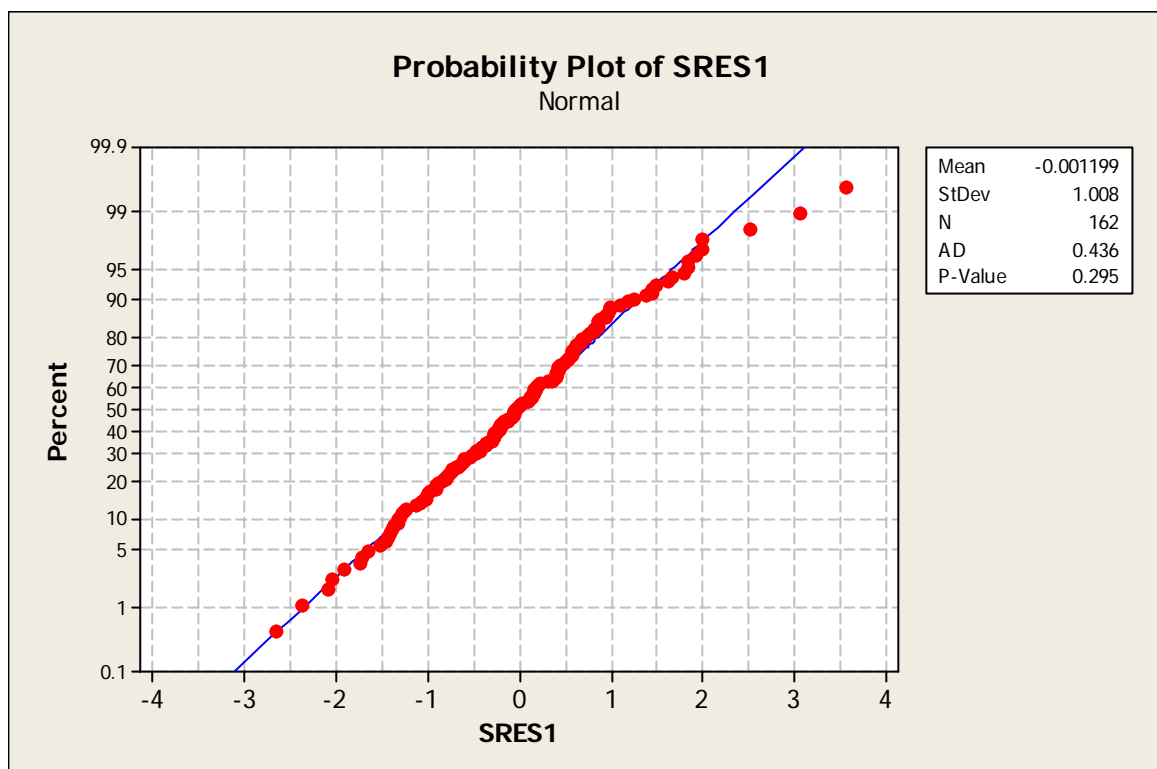
Discriminant Validity Test

Construct1	Constuct2	Model1- Constrained		Model2 - Free		$\Delta \chi^2$	Δdf	P- value
		χ^2	df	χ^2	df			
Relevance	Enjoyment	896.27	35	82.44	34	813.83	1	0.0000
Relevance	Topicality	254.31	27	86.52	26	167.79	1	0.0000
Relevance	Reliability	1010.07	27	58.10	26	951.97	1	0.0000
Relevance	Scope	150.65	20	37.84	19	112.81	1	0.0000
Relevance	Understand- ability	590.28	27	38.42	26	551.86	1	0.0000
Relevance	Novelty	260.04	27	73.33	26	186.71	1	0.0000
Relevance	Credibility	382.09	20	46.45	19	335.64	1	0.0000
Relevance	Knowledge	939.73	20	37.93	19	901.80	1	0.0000
Relevance	Recency	895.91	10	17.95	9	877.96	1	0.0000
Enjoyment	Topicality	459.56	27	79.38	26	380.18	1	0.0000
Enjoyment	Reliability	552.62	27	76.40	26	476.22	1	0.0000
Enjoyment	Scope	165.80	20	41.88	19	123.92	1	0.0000
Enjoyment	Understand- ability	603.49	27	60.79	26	542.70	1	0.0000
Enjoyment	Novelty	314.57	27	72.71	26	241.86	1	0.0000
Enjoyment	Credibility	365.60	20	42.43	19	323.17	1	0.0000
Enjoyment	Knowledge	193.09	20	46.49	19	146.60	1	0.0000
Enjoyment	Recency	916.43	10	28.80	9	887.63	1	0.0000
Topicality	Reliability	518.63	20	85.09	19	433.54	1	0.0000
Topicality	Scope	141.39	14	41.07	13	100.32	1	0.0000
Topicality	Understand- ability	487.82	20	61.42	19	426.40	1	0.0000
Topicality	Novelty	262.58	20	85.89	19	176.69	1	0.0000
Topicality	Credibility	367.41	14	44.56	13	322.85	1	0.0000
Topicality	Knowledge	188.19	14	46.83	13	141.36	1	0.0000
Topicality	Recency	459.85	6	36.96	5	422.89	1	0.0000
Reliability	Scope	152.74	14	18.99	13	133.75	1	0.0000
Reliability	Understand- ability	591.51	20	24.22	19	567.29	1	0.0000
Reliability	Novelty	296.04	20	65.00	19	231.04	1	0.0000
Reliability	Credibility	341.58	14	39.59	13	301.99	1	0.0000
Reliability	Knowledge	161.55	14	27.17	13	134.38	1	0.0000
Reliability	Recency	487.28	6	10.19	5	477.09	1	0.0000
Scope	Understand- ability	143.93	14	22.99	13	120.94	1	0.0000

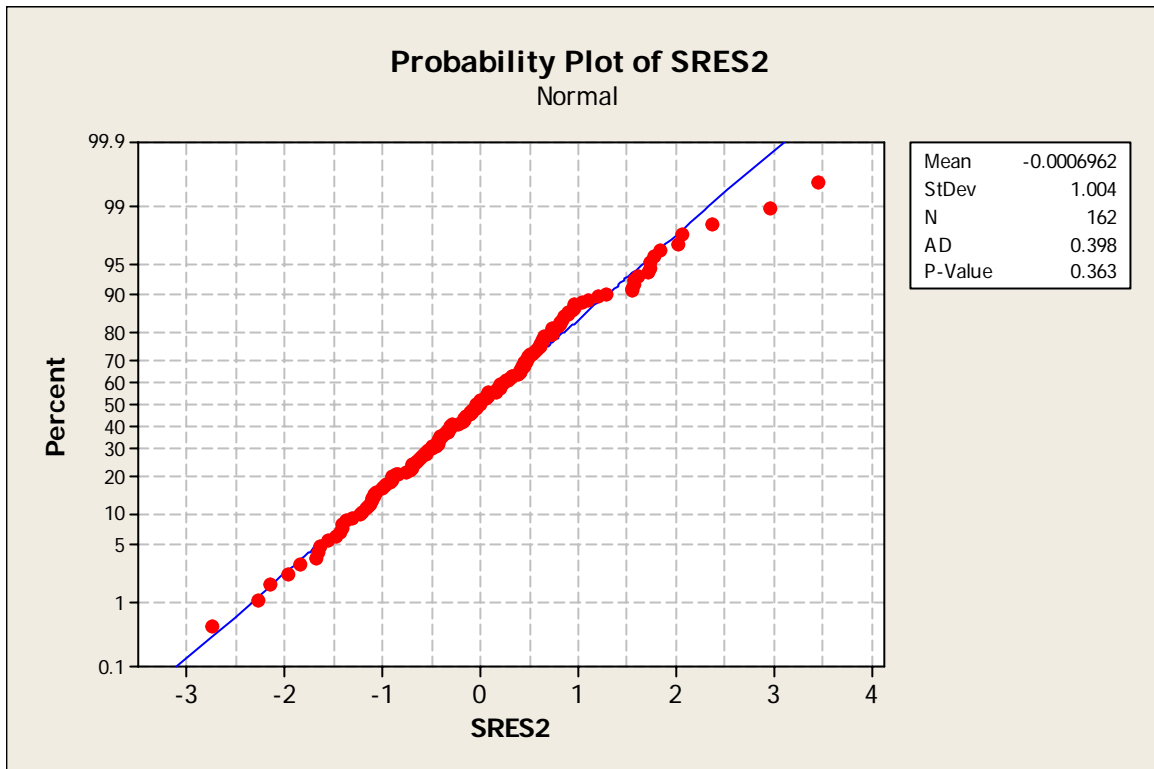
	ability							
Scope	Novelty	156.43	14	23.86	13	132.57	1	0.0000
Scope	Credibility	135.95	9	3.68	8	132.27	1	0.0000
Scope	Knowledge	149.13	9	5.54	8	143.59	1	0.0000
Scope	Recency	131.70	3	0.62	2	131.08	1	0.0000
Understand-ability	Novelty	271.55	20	40.72	19	230.83	1	0.0000
Understand-ability	Credibility	342.71	14	18.05	13	324.66	1	0.0000
Understand-ability	Knowledge	154.05	14	16.23	13	137.82	1	0.0000
Understand-ability	Recency	567.83	6	6.95	5	560.88	1	0.0000
Novelty	Credibility	245.43	14	20.44	13	224.99	1	0.0000
Novelty	Knowledge	160.26	14	20.78	13	139.48	1	0.0000
Novelty	Recency	259.57	6	15.72	5	243.85	1	0.0000
Credibility	Knowledge	340.75	9	8.67	8	332.08	1	0.0000
Credibility	Recency	317.28	3	0.40	2	316.88	1	0.0000
Knowledge	Recency	143.11	3	4.48	2	138.63	1	0.0000

Appendix C

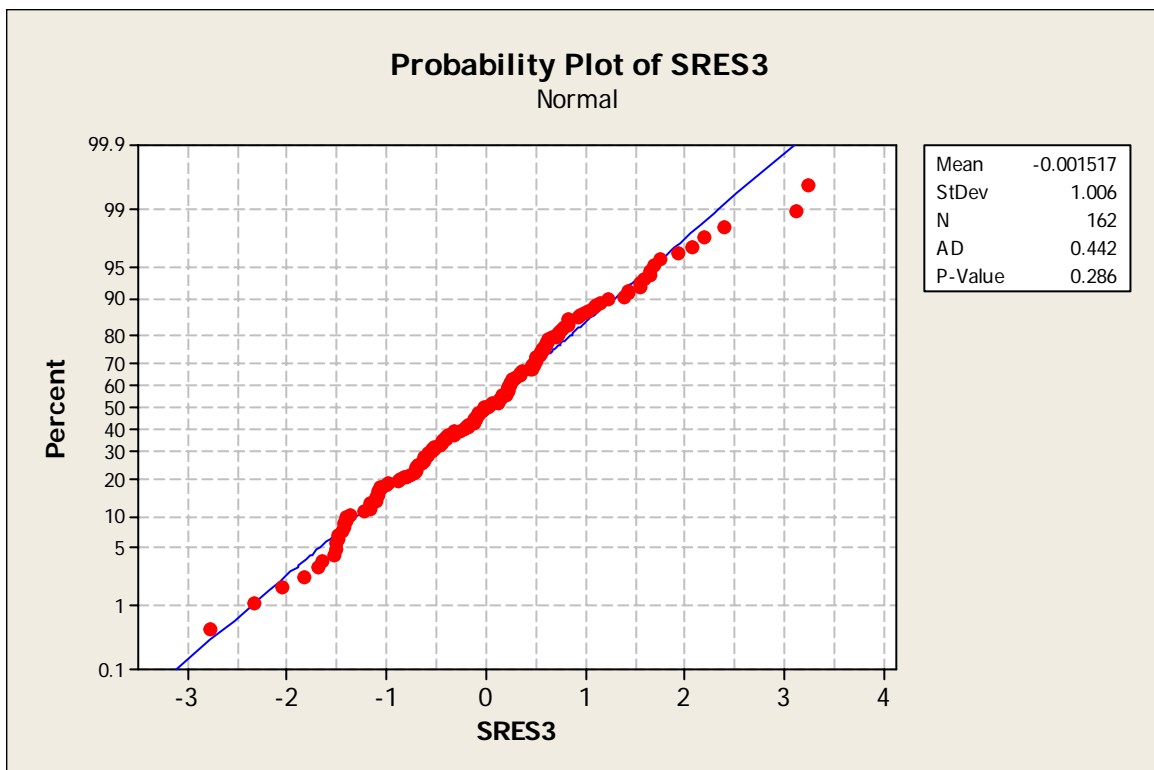
Model 1



Model2



Model 3



Model 4

