A Study of Information Seeking and Retrieving. I. Background and Methodology*

Tefko Saracevic
School of Communication, Information and Library Studies, Rutgers, The State University of New Jersey, 4 Huntington St., New Brunswick, N. J. 08903

Paul Kantor
Tantalus Inc. and Department of Operations Research, Weatherhead School of Management, Case Western Reserve University, Cleveland, Ohio 44106

Alice Y. Chamis† and Donna Trivison‡
Matthew A. Baxter School of Library and Information Science, Case Western Reserve University, Cleveland, Ohio 44106

The objectives of the study were to conduct a series of observations and experiments under as real-life a situation as possible related to: (i) user context of questions in information retrieval; (ii) the structure and classification of questions; (iii) cognitive traits and decision making of searchers; and (iv) different searches of the same question. The study is presented in three parts: Part I presents the background ot the study and describes the models, measures, methods, procedures, and statistical analyses used. Part II is devoted to results related to users, questions, and effectiveness measures, and Part III to results related to searchers, searches, and overlap studies. A concluding summary of all results is presented in Part III.

Introduction

Problem, Motivation, Significance

Users and their questions are fundamental to all kinds of information systems, and human decisions and human-system interactions are by far the most important variables in processes dealing with searching for and retrieval of information. These statements are true to the point of being trite. Nevertheless, it is nothing but short of amazing how relatively little knowledge and understanding in a scientific sense we have about these factors. Information retrieval systems, expert systems, management and decision information systems, reference services and so on, are instituted to answer questions by users — this is their reason for existence and their basic objective, and this is (or at least should be) the overriding feature in their design. Yet, by and large and with very few exceptions (see ref. 1) the basis for their design is little more than assumptions based on common sense and interpretation of anecdotal evidence. A similar situation exists with online searching of databases. While the activity is growing annually by millions of searches it is still a professional art based on a rather loosely stated set of principles (see ref. 2) and experience. While there is nothing inherently wrong with common sense, professional art, and principles derived from experience or by reasoning, our knowledge and understanding and with them our practice would be on much more solid ground if they were confirmed or refuted, elaborated, cumulated, and taught on the basis of scientific evidence.

Since 1980 a number of comprehensive critical literature reviews have appeared on various topics of information seeking and retrieving, among them reviews of research on

- interactions in information systems, by Belkin and Vickery [3]
- information needs and uses, by Dervin and Niles [4]
- psychological research in human computer interaction, by Borgman [5]
- design of menu selection systems, by Shneiderman [6]
- online searching of databases, by Fenichel [7] and Bellardo [8].

It is most indicative that an identical conclusion appears in every one of these reviews despite different orientation of

*Work done under the NSF grant IST85-05411 and a DIALOG grant for search time.

†Present address: Kent State University, Kent, Ohio

‡Present address: Dyke College, Cleveland, Ohio

Received February 5, 1987; accepted March 19, 1987.

the review and different backgrounds of the reviewers. They all conclude that research has been inadequate and that more research is needed. In the words of Belkin and Vickery: "... research has not yet provided a satisfactory solution to the problem of interfacing between end-user and large scale databases." Despite a relatively large amount of literature about the subject, the research in information seeking and retrieving is in its infancy. It is still in an exploratory stage.

Yet, the future success or failure of the evolving next generation of information systems (expert systems, intelligent front-ends, etc.) based on built-in intelligence in human-system interactions depends on greatly increasing our knowledge and understanding of what is really going on in human information seeking and retrieving. The key to the future of information systems and searching processes (and by extension, of information science and artificial intelligence from where the systems and processes are emerging) lies not in increased sophistication of technology, but in increased understanding of human involvement with information.

These conclusions form the motivation and rationale for the study reported here and describe our interpretation of the significance of research in this area in general.

Background

The work reported here is the second phase of a larger long-term effort whose collective aim is to contribute to the formal characterization of the elements involved in information seeking and retrieving, particularly in relation to the cognitive context and human decisions and interactions involved in these processes. The first phase, conducted from 1981 to 1983 (under NSF grant IST80-15335) was devoted to development of appropriate methodology; the study resulted in a number of articles discussing underlying concepts, describing models and measures, and reporting on pilot tests [9–18]. These articles explain the methodological background for the second phase.

The second phase is reported here. It was a study conducted from 1985 to 1987 (under grants listed under title) devoted to making quantified observations on a number of variables as described below. To our knowledge this is the largest and most comprehensive study in this area conducted to date. Still and by necessity (due to the meager state of knowledge and observations on the variables involved) this is an exploratory study with all the ensuing limitations. The results are really reflective of the circumstances of the experiments alone. While at the end generalizations are made, they should be actually treated as hypotheses ready for verification, refutation, and/or elaboration.

The third phase (planned for 1987 to 1989) will have as its objective to study in depth the nature, relations and patterns of some of the critical variables observed here. In this way we are trying to proceed (or inch) along the classic steps of scientific inquiry.

Organization of Reporting

A comprehensive final report to NSF deposited with ERIC and NTIS [19] provides a detailed description of models, methods, procedures, and results; a large appendix to the report contains the "raw" data and forms and flowcharts for procedures used. Thus, for those interested there is a detailed record of the study, particularly in respect to procedures and data.

In this journal, the study is reported in three articles or parts. This first part is devoted to description of models, measures, variables and procedures involved and relates the study to other works. The second part, subtitled "Users and Questions" presents the results of classes or variables that are more closely associated with information seeking; included in Part II are also results related to effectiveness measures. The third part, subtitled "Searchers and Searches" is devoted to the classes of variables that are more closely associated with information retrieving. A summary of conclusions from the study as a whole is also presented in Part III.

Objectives and Approach

As mentioned, the aim of the study was to contribute to the formal, scientific characterization of the elements involved in information seeking and retrieving, particularly in relation to the cognitive context and human decisions and interactions involved. The objectives were to conduct the observations under as real-life conditions as possible related to: (1) user *context* of questions in information retrieval; (2) the structure and classification of *questions;* (3) cognitive traits and decision-making of *searchers;* and (4) different *searches* of the same question.

The following aspects of information seeking and retrieving were studied as grouped in five general classes of the entities involved:

1. *User:* effects of the context of questions and constraints placed on questions.
2. *Question:* structure and classification as assigned by different judges and the effect of various classes on retrieval.
3. *Searcher:* effects of cognitive traits and frequency of online experience.
4. *Search:* effects of different types of searches; overlap between searches for the same question in selection of search terms and items retrieved; efficiehcy and effectiveness of searches.
5. *Items retrieved:* magnitude of retrieval of relevant and nonrelevant items; effects of other variables on the chances that retrieved items were relevant.

The approach was as real-life as possible (rather than laboratory) in the following sense:

- users posed questions related to their research or work and evaluated the answers accordingly; they were not paid for their time, but received a free search
- searchers were professionals, i.e. searching is part of their regular job; they were paid for their time
- searching was done on existing databases on DIALOG. There were no time restrictions
- items retrieved (i.e. answers) were full records as available from the given database.

The control was that all the searching was done under the same conditions. However, for control purposes there were two major restrictions departing from real-life situations: (i) only one and the same database was used for searching the same question by different searchers; in real-life more than one may be used; and (ii) searchers did not have access to users for interviews, they all received the same written question as elaborated by the user. These restrictions were posed because there was no way that we could control the observations otherwise. However, similar restrictions are not that uncommon in real-life searching.

Related Works

The enumerated reviews [3–8] provide an extensive coverage of works related to this study, particularly the review by Belkin and Vickery [3], thus only a brief overview is provided here.

Models of Information Seeking Context

An exhaustive list of variables by Fidel and Soergel [20] illustrates the complexity of the context and processes in online searching: they listed over 200 variables grouped into 8 broad categories. Other models in which some or other of these variables were highlighted greatly depended on a given view of the information seeking context. For a long time the predominant concept around which models revolved was the concept of information need; we shall mention Taylor's work [21] as representative of this school of thought. Slowly, modeling changed to that of problem orientation, viewing the problem behind the question, rather than information need as central to information seeking context. The work by Belkin and colleagues [1, 22, 23] is representative of the problem oriented school of thought, which has increasingly borrowed notions and approaches from cognitive science. The study reported here belongs in this problem oriented category, greatly affected by cognitive science.

Models of Questions

The nature of questions, as reviewed by Graesser and Black [24], has been a subject of study in a number of fields from philosophy and logic to computer science and artificial intelligence. Librarianship also has many works on classification of questions, some going back over 50 years [25]. More recently, the whole area of questions and questioning has become an intensive area of study in artificial intelligence because of its importance to natural language processing, question-answering systems, and expert systems. The book by Graesser and Black is representative of work in this area. So is the pioneering work by Lehnert [26]. Among other things, she provided a novel classification scheme for questions. The work on questions in artificial intelligence is innovative, but it also demonstrates that the progress in this area is slow and incremental. The study reported here in regard to structure and classification of questions is complementary to this work in artificial intelligence.

Models of Search Processes

A number of works in information science have been devoted to modeling and description of the search process. These range from simple flowcharts to complex analysis of the elements and steps involved. Here are some representative models that deal with

- elements and tactics in question analysis and search strategy, by Bates [27,28]
- types of search strategies, by Markey and Atherton [29]
- definition and principles of user interviews and search processes, by Soergel [30]
- identification of heuristics and tactics that are applicable to a wide range of search problems, by Harter and Peters [31]

Most of the descriptions in these studies have been inferred from observations of professional practice or describe desires to improve practice and make it more standardized. Remarkably few models have been put to a scientific test.

Empirical Studies

The factors affecting online searching and human-system interface have been studied in a number of experiments in which data were collected under (more or less) controlled conditions. Here is a list of representative topics in such studies (for others, see ref. 3)

- differences in searching and in search results as affected by various degrees of searching experience, by Fenichel [32]
- relationship between some given cognitive characteristics or educational level of searchers and type of searching and/or search results, by Bridle [33], Bellardo [34], and Woelfl [35]
- types of elements, sequences, and modifications in the search process, by Penniman [36,37], Fidel [38], and Oldroyd and Cetroen [39]

- effects of the type of training received by searchers, by Borgman [40]
- effects of various types of search questions and various user goals on searching, by Rouse and Rouse [41]
- observation of end user search behavior in an operational setting, by Sewell and Teitelbaum [42]
- conceptualization and test of the search process as an imperfect decision-making task, by Fischhoff and MacGregor [43]

The study reported here is closely related by type to the empirical studies reviewed above and as a result it builds on these studies.

General Model

Complex systems such as information retrieval systems (or information systems in general) can be modeled and studied in a number of ways. In the 1960s and 1970s the emphasis was on study and evaluation of input processes and components, such as various representations of documents (or texts) and subsequent retrieval effectiveness. In the 1980s a shift occurred toward study of output processes, users and interactions. In either case, a part of the larger system or a microsystem was isolated, modeled and studied.

Information seeking and retrieving is viewed here as such a microsystem of a larger information system. It is the microsystem that involves the user and interacts with the user and whose role it is to help the user in obtaining appropriate responses.

Figure 1 presents a general model of information seeking and retrieving describing the major events with the accompanying classes of variables. A similar model is found in Belkin and Vickery [3]. The interactive nature of this microsystem is its primary characteristic and one should envision arrows between all events and variables.

The model provides an overview of all events and variables involved in information seeking and retrieving. We follow this general model by presenting next, in greater detail, the model and measures for each class of variables selected for study in this project.

Users and the Context of Information Seeking

There is more to a question than words expressing it. This is a well known truism examined from various viewpoints in psychology and cognitive science, philosophy, linguistics, artifical intelligence, librarianship, and information science [24]. We assume that the context of a question is a governing force describable by a set of variables affecting all events in information seeking and retrieving. The context can be considered as to its external or environmental aspects, or internal or cognitive aspects. Here we are concerned with the latter involving the following

1. *Problem* underlying the question (or more accurately, perception of the problem by the user).

Event	Class of Variables
User (information seeker) has a problem which needs to be resolved	* User Characteristics * Problem statement
User seeks to resolve the problem by formulating a question and starting an interaction with an information system	* Question statement * Question characteristics
Presearch interaction with a searcher i.e. a human or computer intermediary	* Searcher characteristics * Question analysis
Formulation of a search	* Search strategy * Search characteristics
Searching activity and interactions	* Searching
(Possible: initial evaluation of results and reiterative searching)	(* Adjusted search)
Delivery of responses to user	* Items retrieved * Formats delivered
Evaluation of responses by user	* Relevance * Utility

FIG. 1. A general model of information seeking and retrieving.

2. *Intent* for use of the information by the user.
3. *Internal knowledge state* of the user in respect to the problem at hand.
4. *Public knowledge expectations* or estimate by the user.

Problem

Within the framework of information seeking, a problem is defined as an unknown in a work or situation of a potential user of an information system. A problem signifies that which causes difficulty in finding or working out a solution. We assume that information is necessary to solve problems, make decisions, or improve understanding. Such information can be obtained in many ways. One of them is to obtain or deduce it from the existing body of public knowledge, such as in organized retrieval systems, expert systems and the like.

In problem solving research in cognitive science a problem is said to exist when (a) it is at a given state, (b) it is desired to be at another state, and (c) there is no clear way to get from (a) to (b) [44,45]. Either of the two states could be well defined or poorly defined, leading to four obvious categories: both well defined, both poorly defined, first well defined and the other poorly defined, and vice versa.

Borrowing from these notions we have concentrated on observing the effects of the degree of how well the problem is defined, as perceived by the user. We also compared the user perception with that of a searcher.

Internal Knowledge State

People ask questions because they don't know something or they want to confirm something, or, in the words of

Belkin [1], because they have an anomalous state of knowledge in respect to a problem. Internal knowledge state refers to the degree of knowledge an information seeker has about the problem at hand and/or the question arising from the problem.

Internal knowledge state involves many aspects related to cognitive structures and processes: how knowledge is stored, organized, associated, retrieved, and changed in one's mind. A considerable amount of research in cognitive science is devoted to these questions (see ref. 46). Recognizing the great complexity of internal knowledge states, we have concentrated on a rather simple aspect in this study: the effects of the degree of internal knowledge about the problem or question at hand, as perceived by the user about his/her own knowledge. We have also compared the user indication of internal knowledge with that of the searcher.

Intent

An information seeker inevitably has some purpose in mind for the use of requested information. In the framework of information retrieval, the intent is defined as a planned or prospective use of information, including constraints, if any, on that information. In other words, the users have some preconceived ideas about

- the use of information in respect to the problem
- the amount of time and effort they are willing to spend in absorption of or deduction from the provided information
- the desirable characteristics of responses to the question, such as completeness, precision, reliability, timeliness, etc.
- the form characteristics of responses deemed most desirable, such as to the language, source, etc.
- the economic value they attach to responses.

The intent in information retrieval and the goal state in problem solving are related but not identical. They are treated separately because the information seeking intent can be a very specific aspect of problem solving, exclusively devoted to the use of supplied information within a broader context of a goal in problem solving.

Information seeking intent can be focused (where a type of use is more specified), and unfocused (where the use is less specified). We concentrated on the effects of the degree of how well the intent is defined, as perceived by the user. We also compared the user perception with that of the searcher.

Public Knowledge Estimate

Public Knowledge is the recorded knowledge on a subject in the public domain; in the context of information retrieval it refers to the records or literature on a subject.

People ask questions within the framework of public knowledge. This involves a number of aspects such as their perception of what is (or what is not) there, how is it organized, what can they expect to get, etc. Thus, a user's estimate of public knowledge defines his/her expectations, which in turn affects the evaluation.

We concentrated on the effects of the estimate of public knowledge by the user. We also compared the user estimate with that of the searcher.

Measures of Information Seeking Context

Four Likert-type scales have been used to obtain an indication of the information seeking context first from users and then also as perceived by searchers.

1. PROBLEM DEFINITION SCALE
   "In your opinion, and on a scale from 1 to 5, would you describe your problem as weakly defined or clearly defined, with 1 being weakly defined and 5 being clearly defined?"

   : ____ : ____ : ____ : ____ : ____ :
   1 2 3 4 5
   weakly defined clearly defined

2. INTENT SCALE
   "On a scale from 1 to 5, would you say that your use of this information will be open to many avenues, or for a specifically defined purpose, with 1 representing open to many avenues and 5 representing a specifically defined purpose?"

   : ____ : ____ : ____ : ____ : ____ :
   1 2 3 4 5
   open to many avenues purpose is narrowly defined

3. INTERNAL KNOWLEDGE SCALE
   "On a scale from 1 to 5, how would you rank the amount of knowledge you possess in relation to the problem which motivated this request?"

   : ____ : ____ : ____ : ____ : ____ :
   1 2 3 4 5
   little personal knowledge considerable personal knowledge

4. PROBLEM-PUBLIC KNOWLEDGE SCALE
   "On a scale from 1 to 5, how would you rank the probability that information about the problem which motivated this research question may be found in the literature?"

   : ____ : ____ : ____ : ____ : ____ :
   1 2 3 4 5
   highly improbable that it exists highly probable that it exists

Structure and Characteristics of Questions

Questions are a whole class of linguistic expressions, uttered with the intention of inducing a response. In addition to grammar, there is also a logic and pragmatic context of questions (e.g. refs. 47 and 48). Study of questions and questioning dates at least from Aristotle and encompasses a number of fields as reviewed by Kearsley [49]. As mentioned, this is a particular active area of research in artificial intelligence and cognitive science because of the significant implications for design of intelligent systems [24,26].

While questions can have a number of functions, of interest here are those that have a role in problem solving. Within the context of information seeking, questions are defined as statements that are verbalized (written or oral) representations of a problem at hand; their function is to elicit a response. We have concentrated on

1. Observing the effects of constraints on questions as indicated by users.
2. Describing and testing a structure of questions in information retrieval.
3. Developing and testing a classification scheme for questions in information retrieval oriented toward grouping of questions by several characteristics above and beyond their subject contents.
4. Observing the effects of different classes of questions.

Constraints on Questions

The user was asked to indicate for the question the following aspects considered as constraints

1. Do you want a broad or precise search?
2. What is the type of application of this research or work (undergraduate study; graduate study; faculty research; industrial; general; other)?
3. Do you want to place restrictions on the language of publication of the articles retrieved (English only; any language)?
4. Do you want to restrict the years of publication of the article retrieved (last 5 years; no limits; specify years: 19__ to 19__)?

Structure of Questions

As a rule, questions in information retrieval consist of three parts: a lead-in, a subject, and a query.

The lead-in is not directly searchable. It may consist of phrases such as: "I want information about . . . ." However, at times lead-in may have implication for searching; for instance "what is . . . " implies request for definition, "where is . . . " implies request for location, or "when was . . . " implies involvement of a time element. In such cases, lead-ins are important for recognizing presuppositions in a question, i.e. implications not directly expressed (see discussion below).

The subject of the question is the central concept of the question. It is a concept around which all other aspects of the question revolve and relate. A question can have more than one subject.

The query is the specific aspect asked about the question's subject. It is an attribute, characteristic, component, or part of the subject about which information is desired. There can be more than one query about a subject.

Example:

What are	the advertising expenditures	of the automobile industry?
LEAD-IN	QUERY	SUBJECT

In addition, questions may have constraints geared toward restricting or orienting the type of desired responses.

Classification of Questions

It is not uncommon to hear searchers describe given questions as: "complex," "specific," "very general," "difficult," "unsearchable," "unclear," and the like. In such cases searchers are (possibly even unwillingly) applying certain general attributes to classify questions. While no generally accepted classification of questions in information retrieval exists, it is of practical interest to specify certain attributes which could be used to classify or describe questions in information retrieval.

The scheme described here has been developed on the basis of criteria that might relate categorizations of questions to approaches to and outcomes of searching. Five categories are used: domain, clarity, specificity, complexity and presupposition.

1. **Domain:** classifies a question in the appropriate general subject or topic area in which it falls. This can be done on the basis of a general subject classification. In our case, we have chosen the list of DIALOG subject categories of their files as the subject classification scheme.
   Measure: for a question the classifier indicates the number or name associated with the given DIALOG category best fit for the question. More than one DIALOG category can be used, as necessary.
2. **Clarity:** classifies a question on the basis of degree of clarity in respect to (a) semantics (meaning of terms) and (b) syntax (relation and logic between terms).
   Measure: for a question the classifier indicates on a scale from 1 to 5 the degree of clarity, where 1 means "unclear" and 5 means "clear". Two scales are used:
   (a) For semantic clarity (meaning of terms).
   (b) For syntactic clarity (relation and logic between terms).
3. **Specificity:** classifies a question on the basis of degree of specificity of (a) query terms and (b) subject terms. Specificity ranges from very general or broad (e.g. as found in a thesaurus under BT-Broader terms) to very specific or narrow (e.g. as found in a thesaurus under NT-Narrower terms).
   Measure: for a question the classifier indicates on a scale from 1 to 5 the degree of specificity, where 1

means "broad" and 5 means "narrow". Two scales are used

(a) For terms associated with the query part of the question.

(b) For terms associated with subject part.

4. **Complexity:** classifies a question on the basis of complexity for searching as related to number for search concepts involved. The search concepts are those that are used as search terms or will be translated into one or more search terms (i.e. a search concept can have more than one search term as its synonyms or near synonyms). The search concepts in both the query part and subject part of the question are added together.

   Measure: for a question the classifier indicates two aspects

   (a) On a scale from 1 to 5 the degree of complexity, where 1 means low complexity and 5 means high complexity.

   (b) The number of search concepts to be used as or translated into search terms.

5. **Presupposition:** classifies a question on the basis of presence or absence of implied (not explicitly stated) concepts derived from sharing of common linguistic and world knowledge. Of interest are those implied concepts that could be used as search terms or search constraints. Most commonly presuppositions are expressed by phrases such as "What is . . . ", "Where is . . . ", etc. implying existence or verification, identity or definition, quality, relation, number, location, or time.

   Measure: for a question the classifer indicates two aspects:

   (a) On a scale from 1 to 5 the degree of presence of presuppositions, where 1 means "no presupposition" and 5 means "many presuppositions".

   (b) The number of presuppositions that can be translated into search terms or search constraints.

Searchers

A large number of environmental (or external) and cognitive (or internal) factors, e.g. organizational setting, economic constraints, affect searcher's decisions and thus retrieval effectiveness and efficiency. While recognizing the external factors, we have concentrated on a limited number of cognitive traits of searchers:

1. Language ability or the ability to make inductive inferences through word association, as measured by a standard test called Remote Associates Test.
2. Logical ability or the ability to make deductive inferences as measured by a standard test called Symbolic Reasoning Test.
3. Preferred style or mode of learning as measured by a standard test called Learning Style Inventory.
4. Online experience as derived from a questionnaire given to searchers.

Remote Associates Test (RAT)

RAT is a test of semantic associations. It claims to test the ability to make inductive inferences. The test instrument was developed by Mednick and Mednick [50] and it has been widely applied and tested for fifteen years. The test presents the subjects with sets of three stimulus words and asks them to infer the fourth word that is related (or has something in common with) all three stimulus words, as in the following examples:

			Prescribed right answer
cookies	sixteen	heart _____	(sweet)
poke	go	molasses _____	(slow)
surprise	line	birthday _____	(party)
skunk	kings	boiled _____	(cabbage)

The test consists of thirty such matches to be made in twenty minutes. The score is a straight count of right (predetermined) answers out of thirty.

Symbolic Reasoning Test (SRT)

SRT is one of the tests in the Employee Aptitude Survey (EAS). The survey is a battery of 10 tests developed by Ruck and Ruck [51], widely used in business and industry for personnel selection. The Symbolic Reasoning Test, based as the name implies on symbols, claims to measure deductive inferences. This is a thirty item test done in five minutes. Each item specifies a relationship of A to B to C and requires a "true", "false", or "don't know" answer as in the following example:

A > B > C therefore A < C: true, false, don't know

The test is scored on a straight count of correct answers.

Learning Style Inventory (LSI)

LSI is based on a theory describing learning as an integrated, four stage process that involves the use of four different cognitive modes as described by Kolb [52]: (i) Concrete Experience (CE); (ii) Reflective Observation (RO); (iii) Abstract Conceptualization (AC); (iv) Active Experimentation (AE). LSI claims to measure individual preferences for each of the above four basic modes for learning and places an individual in a composite category indicative of his/her learning style. The test has been widely applied and tested for over ten years. The respondent is asked to rank order from 1 to 4 a series of four statements in response to a question on how he/she learns. There are twelve sets to rank within twenty minutes; here are two examples

When I learn:

_ I like to deal with my feelings	_ I like to watch and listen	_ I like to think about ideas	_ I like to be doing things

I learn best from:

_ Personal relations	_ Observation	_ Rational theories	_ A chance to try out and practice

Each column represents a separate learning mode arranged from left to right: CE, RO, AC, and AE. The raw scoring is done by adding all the rank numbers given by the respondent in a column. Two combination scores are obtained by subtracting Concrete Experience score from Abstract Conceptualization score (AC − CE), and the Reflective Observation score from Active Experimentation score (AE − RO). The combination scores are used to indicate an individual's learning style—the extent to which he/she emphasizes abstractness over concreteness (AC − CE) and action over reflection (AE − RO). These scores are also used to make a graph with AC − CE on the *Y* axis and AE-RO on the *X* axis. An individual is placed in one of the four quadrants characterizing a person as to style of learning as: (i) converger (lower left); (ii) diverger (upper right); (iii) assimilator (lower right); or (iv) accommodator (upper left). A person with a zero score on either AC − CE or AE − RO is considered indeterminate.

Searcher's Experience

The amount and type of experience a searcher had in online searching was obtained from a questionnaire. Since DIALOG was used for searching, the questions asked were specific to DIALOG system:

1. "How often do you search DIALOG?: daily, twice a week, once per week, twice per month, less."
2. "Indicate seven DIALOG databases you search most often in order of decreasing use."
3. "Indicate thesaurus most important to you when you search."

The answer to the first question has been used as a searcher's variable. The answers to the last two questions have been used to match the searchers with databases they are to search in the study, so that their background and the databases match as closely as possible.

Search

The elements and subprocesses in a search have been modeled and categorized in a number of ways [27 to 31]. While it is difficult to make a clear separation, a common way is to make a distinction between:

1. *Question Analysis:* procedures that deal with decisions on semantic and pragmatic (contextual) aspects of a question in preparation of a search statement, including determination of appropriate information sources, and a possible interview with a user.
2. *Formulation of Search Strategy:* procedures that deal with decisions on syntactic and logical aspects of the search statement, including incorporation of constraints, if any, and selection of tactics related to a desired level of effectiveness and efficiency.
3. *Searching:* procedures used in the conduct of a search, including the use of protocols and capabilities of given information systems and obtaining the responses.

In this study we have concentrated on several specific variables within Question Analysis and Search Strategy. We kept searching constant, meaning that the environment, hardware, software, procedures, and information system for all searches were the same.

Question Analysis

We have concentrated on two aspects:

1. Degree of overlap or agreement in selection of search terms by different searchers searching the same question based on the written question statement submitted by the user.
2. Differences in results among search statements derived from four different sources for search term selection, namely:

Type 1. From a statement about the problem at hand as tape recorded by the user, but without recourse to the written question.

Type 2. From the tape problem statement and the written question submitted by the user.

Type 3. From the written question using only the words in the question as search terms without any further elaboration.

Type 4. From the written question plus terms from an appropriate thesaurus for elaboration.

These four types of searches we labeled "project searches."

Search Strategy

We have concentrated on three aspects of the search statement as a whole:

1. Degree of overlap or agreement in retrieved items by search statements done by different searchers searching the same question based on the written question statement submitted by the user.
2. Differences in search effectiveness as expressed by measures based on relevance and utility judgements of users.
3. Differences in search tactics and efficiency as used by different searchers searching the same question; these characteristics are based on the level of effort used in a search as expressed by the tactics and efficiency measures described below.

Measure of Overlap in Search Terms

This measure indicates for each pair of searchers the degree of agreement or overlap in selection of search terms in searching of the same question. However, because there may be differences in how Searcher 1 agrees with Searcher 2, from how Searcher 2 agrees with Searcher 1 the measure is asymmetrical (e.g., in a search statement Searcher 1 uses two terms, and Searcher 2 uses six terms; the two terms of S1 are also used by S2, but S2 has four more terms, thus S1 is in 100% agreement with S2, but S2 is only in 33% agreement with S1). The overlap or agreement measures for search terms are:

$$S_{1,2} = \frac{|S_1 \cap S_2|}{|S_1|} = \frac{\text{no. of search terms in common}}{\text{total no. of terms used by Searcher 1}}$$

$$S_{2,1} = \frac{|S_1 \cap S_2|}{|S_2|} = \frac{\text{no. of search terms in common}}{\text{total no. of terms used by Searcher 2}}$$

Measure for Overlap in Output

This measure indicates for each pair of searchers the degree of overlap in retrieved items for the search of the same question. The overlap measure for output parallels the measure for the overlap or degree of agreement in search terms. It is calculated in the same way except that either the total number of retrieved items or else the number of relevant items retrieved is substituted for the number of search terms. Since the formula is the same as shown above, it is not repeated here. Both overlap measures are asymmetrical and both can be used for arranging data into a matrix and/or a histogram to study the distribution of overlap from a group as a whole.

Effectiveness Measures

Two sets of measures for evaluating the effectiveness of a search have been used, based on the two most often used criteria:

1. *Relevance:* the degree of fit between the question and the retrieved item. The criteria of *"aboutness"* is used.
2. *Utility:* the degree of actual usefulness of answers to an information seeker. The criteria used is the *value* to the information seeker.

In this study both relevance of items and the utility of the entire retrieved set have been established by users.

Definition of relevance. The following definitions have been provided to users for judging the answers (i.e. abstracts):

"Each abstract should be evaluated according to its degree of relevance to the question you submitted for searching. The degree of relevance should be determined using the following three point scale:

RELEVANT—Any document which on the basis of the information it conveys is considered to be related to your question, even if the information is outdated or already familiar to you.
PARTIALLY RELEVANT—Any document which on the basis of the information it conveys is considered only somewhat or in some part related to your question or to any part of your question.
NONRELEVANT—Any document which on the basis of the information it conveys is not at all related to your question."

Recall and Precision. These are measures based on the relevance judgement of users where:

Precision = probability that a retrieved item is relevant

Recall = probability that a relevant item in the file is retrieved

These probabilities were estimated as follows for a given search:

$$\text{Precision} = \frac{\text{No. of relevant items retrieved by the search}}{\text{Total no. of items retrieved by the search}}$$

$$\text{Recall} = \frac{\text{No. of relevant items retrieved by the search}}{\text{Total no. of relevant items in the union of items retrieved by all searchers for that question}}$$

Precision is easy to establish directly from the output of evaluated items for a search. Recall is not easy to establish, because it is never apparent how many items in a file are relevant to the question. Each question was searched by a number of searches and types of searches. A union of retrieved items from all searches for the question was established (i.e., by merging all the outputs and eliminating duplicates) and sent to the user for evaluation. In this way the evaluated items from the union served as the benchmark of individual search recall. This presents a *comparative* rather than absolute measure of recall performance for any given search.

Utility Measures. These are measures based on users' expression of degree of satisfaction and value of the retrieved items as a whole. Recall and precision are universally used measures. Unfortunately, there are no such universally used utility measures, thus we had to establish our own. The following questions were posed to the users which reflect utility based measures:

1. How much time did you spend reviewing these abstracts? ____
2. In an overall sense, if you were asked to assign a dollar value to the usefulness of this entire set of abstracts to you, what would that dollar value be?
   $_____ _____ I cannot assign a dollar value
3. Could you rate your participation in this project and the information which resulted as:

   5 Worth much more than the time it has taken
   4 Worth somewhat more than the time it has taken
   3 Worth about as much as the time it has taken
   2 Worth less than the time it has taken
   1 Practically worthless

4. Problem Resolution Scale—On a scale of 1 to 5, what contribution has this information made to the resolution of the problem which motivated your question?

: ___ : ___ : ___ : ___ : ___ :
1 2 3 4 5
nothing contributed — substantial contribution

5. Satisfaction Scale—On a scale of 1 to 5, how satisfied were you with the results of the search?

: ___ : ___ : ___ : ___ : ___ :
1 2 3 4 5
dissatisfied — satisfied

Tactics and Efficiency Measures

The following measures have been used for online searches describing search characteristics as to:

1. Number of *commands* used by a search.
2. Number of command *cycles* used by a searcher. A cycle is a set of commands in sequence from those used to select, combine, and/or expand terms to a command used to type (or view) the results. A cycle ends with display (type, print) of intermediate or final items retrieved for a set of preceding commands.
3. Number of *search terms* used by a searcher searching a question.
4. *Preparation time* used by a searcher in preparing a search for a question.
5. Online *connect time* used by a searcher in searching a question.
6. *Total time* used by a searcher (connect time plus preparation time).

The first three measures reflect the search tactics and the last three the efficiency or costs associated with the search and searching.

Items Retrieved

Output from a search may be called by a number of names: references, answers, documents, abstracts, displays, etc. We have selected a neutral label, "items retrieved," to designate the output for a search, that is, the individual records retrieved from a database in response to a question. While databases provide different format options for each item retrieved, we have chosen to use the full database record of each and every item retrieved. Thus, "items retrieved" were *full records*. In all databases searched for the project, full records included bibliographic information, index terms and/or classification codes and an abstract.

We have concentrated on two aspects.

1. Observing the distributions of items retrieved.
2. Analyzing the chances or odds that an item retrieved will be relevant as affected by other variables.

The second aspect is actually a major point of the study.

Measures for Items Retrieved

The following measures of quantity were used:

1. Total number of items retrieved for a question calculated in two ways: (i) as a sum for all searches for the question *including* duplicates retrieved by different searches, and (ii) as a union of all searches for the question counting the distinct numbers only by *excluding* or eliminating the duplicates.
2. Total number of items evaluated by a user for a question again calculated in two ways. (The items evaluated for a number of question were smaller than items retrieved as described below under Procedures.)
3. Total number of items judged by users as relevant, partially relevant and not relevant for a question, again calculated two ways.

The quantities from the last measure have been used to calculate the chances of retrieval of a relevant item as a function of some other variable by a statistical method called Logarithmic Cross Ratio Analysis (described under Analysis Methods). These measures of performance when correlated with variables associated with users, questions, searchers and searches revealed effects which cannot be seen using the traditional measures of recall and precision.

Procedures

Users

On the basis of advertising about the project forty users volunteered to participate. The forty users that started with the project completed all their tasks from start to finish, thus, user response was 100%. Each of the forty users:

- filled out a short questionnaire about his/her background
- submitted one question for searching together with information on desired question constraints; thus, 40 questions were used in the study
- participated in a taped interview describing the problem underlying the question and the intent in use of information and indicated the context measures
- evaluated items retrieved in response to his/her question as to relevance and indicated the utility of the search as a whole.

The forms, instructions, and procedures used in these and all other tasks in the study, as well as the full statement of questions are assembled in the appendix to the Final Report for the project [19].

Searchers

On the basis of advertising and presentations at professional meetings we received an indication of interest from about forty information professionals, thirty-six of these eventually participated from beginning to end. These became known as "outside searchers." The outside searchers were paid $100 for their time. Each of the searchers:

- was tested on three cognitive tests described above and filled out a questionnaire on their search experience
- received instructions on procedures for searching in a presentation and in writing
- received five or six questions for searching (each question was searched by 5 different searchers, but since there were forty questions and thirty-six searchers some searched six questions); the questions they received were copies of written statements by the users, together with users' indication of constraints
- prepared a preliminary search strategy
- conducted the search and recorded the results. (The whole search was recorded on a disc and a printout).

No restrictions, such as time limits, were placed on searchers in preparation for and conduct of the search.

In addition, three searchers from the full time staff of the project were engaged in searching so called project searches, as described below. These became known as "project searchers". Thus, a total of thirty-nine searchers participated in the study.

Searching

Searching was done on DIALOG, the largest vendor in the world as to the number of databases and frequency of use. More searchers have more professional experience with DIALOG than with any other system, which was the reason for selecting it.

Each question was searched on one database only. The database was selected by the project team on the basis of closeness of fit between the question and the subject of the database. The searchers were assigned questions which matched as closely as possible their own database experience. Searching was done on microcomputers with preprogrammed log-on and downloading protocols. Appropriate thesauri and manuals were assembled and made available to all searchers. All searching was done in the same room and environment, and under the same conditions.

Project Searches

As mentioned, each question was searched by five outside searchers. In addition, there were four searches done in-house by the project staff. These additional searches were labeled "project searches", and the staff searchers as "project searchers." Thus, nine searches were done for each question: five by outside searchers and four by project searchers.

As mentioned, the objective of four project searches was to study various types of searches resulting from different sources for the search strategy. The sources for each type are enumerated above under "Question Analysis."

The project searches were done by three project searchers. They also conducted the taped interview with the users. However, the project searches were arranged so that they were done by the project searcher who did *not* do the interview. Thus, each project searcher heard the taped interview for the first time when the first project search was done. The project searchers took all the same cognitive tests as the outside searchers.

Question Structure and Classification

This part of the study had two separate objectives:

1. To test the suggested model of question structure and the scheme for question classification.
2. To observe the effects of different question classes.

For the first objective a separate experiment was conducted. It consisted of testing the consistency or degree of agreement in assigning question structure and classification classes by a number of judges. On the basis of advertising a group of twenty-one information professionals was assembled to assign the question structure and classification. These were a different group from the searchers and had nothing to do with searching. They are called "classification judges." The experiment consisted of each judge:

- receiving twenty questions for judging; these were selected at random from the forty questions submitted by users
- assessing the question structure on an appropriate form
- classifying each question as to domain, clarity, specificity, complexity, and presuppositions.

The results of classification were calculated for agreement. The classes (and questions in these classes), with significant agreement, were then used to analyze chances of retrieval of relevant answers, i.e. to address the second objective of this part of the study.

Prior to conducting the experiment with twenty-one judges, a pilot experiment was carried out with two other judges to observe if the procedures worked. They did. In the Final Report [19] we report results only on the two judges in the pilot study and not on the twenty-one judges in the experiment proper (the report was written before this part of the study was completed). Here we are reporting the results on twenty-one judges only, i.e., we are disregarding the pilot study.

Evaluation by Users

Here is a summary of steps involved in evaluation:

- the end result of each search was a list of accession numbers of items retrieved

- the accession numbers for retrieved items from nine searches for each question were merged and a union set created, i.e., the duplicates were eliminated, thus this union set consisted of all and only distinct items retrieved for the question
- if the union of retrieved items exceeded 150, only the first 150 items were designated to be sent to the user for evaluation. This was done to make evaluation manageable for the user, who, if presented with an overwhelming output might have considered the task unreasonable, and rightly so, and may not have finished. (However, for three questions that slightly exceeded the 150 limit we sent all retrieved items). Since all searchers had an equal chance to contribute to the retrieved set, including the first 150 items, there was no bias toward any searcher. DIALOG databases are organized on a last in/first out principle, thus the first 150 items represent the most recent additions to the databases and the literature
- the full record of each item retrieved in the union set to be sent to user was downloaded from DIALOG onto floppy disk and then printed
- to each item retrieved a line for evaluation was added:
  ___ Relevant ___ Partially Relevant ___ Nonrelevant
- the printout and a carbon copy were sent to the user for evaluation. The user also indicated the utility of the search as a whole. Finally, the user returned the original with his/her evaluation.

The relevance evaluations were recorded with accession numbers of each item retrieved to create a benchmark file against which the output of all searches was compared. In turn, a large master file was created containing values for all variables involved, including user evaluations. The master file served as a basis for statistical analysis.

Summary

The study involved the following:

- forty users, each providing one question and a taped interview on the problem at hand
- thirty-six outside searchers and three project searchers for a total of thirty-nine searchers
- for each of the questions five different outside searches and four project searches, for a total of nine searches per question
- all together for the forty questions 360 searches, consisting of 200 outside and 160 project searches.

In addition, a separate question structure and classification experiment involved twenty-one judges.

Methods of Statistical Analysis

Approach

The basis for statistical analysis was user evaluation of retrieved items. Every evaluation involved five distinct entities: (i) user, (ii) question, (iii) searcher, (iv) search, and (v) retrieved item. The analysis proceeded by examining these entities first one by one and then at several different levels of aggregation.

Any measured variable (using measures described above) describes one or more of the five entities. For instance, the cognitive characteristics of the searcher describe only a single entity, the searcher; the number of commands or the number of search terms used describe only another single entity, the search. On the other hand, user evaluation is a description of the relevance of a retrieved item by the user; and it relates two entities: user and retrieved item. The overall retrieval or precision scores for a given question combines several searches by several searchers and so are descriptive of three entities: the question, the searches, and the searchers together. And so on.

The data has been examined at each of such different levels of aggregation. Some of the levels of aggregation are more familiar in everyday practice, while others are more powerful in the search for possible explanatory relations. A data file was formed for each level of aggregation by retaining those variables that are meaningful at that level and ignoring the others. These files were used to investigate various statistical relationships.

The relationships (as well as other statistics) were analyzed and displayed by using BMDP and SPSSX statistical packages. BMDP is widely used in biomedical research and SPSSX in social science research. The BMDP Manual [53] provides a detailed description of statistical techniques used.

We distinguished between findings that are statistically significant and those that are also important, i.e. those findings that provided a substantial explanation of the relevance of retrieved items. When a statistically significant relationship is found it can be assigned some measure of association. This is a measure of the extent to which one of the variables in question (X—the one presumed independent) determines the other (Y—the one taken to be dependent). The fact that a relationship is significant does not mean that it is important.

We regarded a relationship as important if the independent variable explains a substantial amount of the observed variation in the dependent variable. The measures of importance that have been used in this study are the R-squared measure for regression analysis and the t-value for analysis of the log cross ratio. Below we review the two. The R-squared measure is reviewed only briefly because it is used widely in information science research, but the log cross ratio is reviewed in some detail, because to our knowledge this is its first application in research in this area. It is a powerful technique widely used in biomedicine and, as we found, a powerful tool for explanation of factors that affect chances of retrieved items being relevant.

Macro (Search-wise) and Micro (Item-wise) Analysis

Two levels of statistical analysis were used in this study: macro, or search-wise, and micro, or item-wise. On the search-wise level explanations were sought for the impact of

given variables on precision and recall and on the odds of either being above average; that is, the macro analysis focused on the effectiveness of the searches as a whole. On the item-wise level, explanations were sought for the impact of given variables on the odds that retrieved items were relevant as opposed to not relevant. As the names imply, the search-wise analysis concentrates on traditional measures of precision and recall of the whole search, while the item-wise analysis ignores these measures and concentrates on the relevance of each item retrieved by a search. Regression and logarithmic cross ratio analyses were used for the former and logarithmic cross ratio analysis only for the latter.

Regression Analysis

In regression analysis, exploring the dependence of Y on X, we try to find the best straight line describing Y as a function of X. We may imagine all the values of Y and X plotted in a single graph.

When such a graph is made the values of Y will show some substantial variation. This degree of variation is conveniently summarized by a statistical quantity called the variance. The variance is the average value of the square of the difference between any particular value of Y and the average of all the values of Y. When a line is fit to the data, to explain Y, a certain amount of the value of Y remains unexplained. The average square of the unexplained part is called the residual mean square variation. The difference between the two is the part of the variance that is explained by the model. This may be expressed as a percentage of the original variance, which is called R-squared. Thus, if R-squared is 80%, the model explains 80% of the original observed variation in the values of Y. If R-squared is 10%, 90% of the original variation remains unexplained.

Logarithmic Cross Ratio Analysis

In analysis of what affects (i) retrieval of items judged by users as relevant or partially relevant, and (ii) precision and recall of searches we have used a powerful technique called cross product ratio analysis. The technique is described in chapter 11 of BMDP Manual [53] and a detailed discussion of the meaning of cross product odds ratio is given by Fleiss [54].

To apply the cross ratio analysis, each variable (for which such a distinction is meaningful) is broken into a class of high values and a class of low values. For convenience the mean is generally taken as the cut point or dividing line, thus "high" means above the mean and "low" below the mean. Since in micro analysis the dependent variable of greatest interest is the relevance of retrieved items we take for high value of the relevance, items judged "relevant" or "partially relevant". In macro analysis we take for high value searches having precision or recall above mean. Every case may then be classified into exactly one of four cells in a 2 × 2 table. The number of cases for which the variable is low and the item is not relevant (or below mean) is designated by "A" and so forth. The cross ratio for this table is defined as the ratio of two products: xpr = AB/CD.

For micro or item-wise analysis the values are displayed in the table as follows:

	Independent Variable	
Dependent Variable	Low (Below Mean)	High (Above Mean)
Not Relevant	A	C
Relevant or Part. Rel.	D	B

Number in A indicates the number of items that were retrieved in association with low (below mean) value of the independent variable (e.g., by searches having below mean number of cycles) and at the same time were judged not relevant by users. When it is written in this form its meaning is rather obscure. But, it is easy to see that if A and B are large while C and D are small the cross product ratio will be large. The meaning becomes clearer if we consider the odds that a high value of the independent variable leads to relevant items. For high values of the independent variable the odds that a retrieved item will be relevant or partially relevant are given by B/C. For low values these odds are given by D/A. The ratio of these two odds ratios reflects the increase in odds due to moving from a low value of the variable to a high value of the variable. This ratio (B/C)/(D/A) is precisely equal to the cross product ratio i.e. AB/CD. For this reason the cross ratio is also referred to as the odds ratio.

For macro or search wise analysis the values are displayed in the table as follows:

		Independent Variable	
	Dependent Variable	Low (Below Mean)	High (Above Mean)
Precision (Recall)	Below mean	A	C
	Above mean	D	B

Number in A indicates the number of searches that were below the mean value of the independent variable (e.g., searches by searchers with below mean score on Remote Associates Test) and at the same time had a below mean precision (recall). As in the case of item-wise analysis described above, the ratio of the two odds ratios ((B/C)/(D/A) = AB/CD) reflects the increase in odds due to moving from a low value of the variable to a high value of the variable.

Since the cross product ratio is always positive and may become infinite, it is replaced by its logarithm which has a more symmetrical distribution and which, for samples as large as the ones we are using, is essentially normally distributed. Thus, in our discussion of the impact of independent variables we have consistently used the log odds ratio as a statistical indicator. Since the log odds ratio is distri-

buted essentially normally, the t statistic (that is, the measured value of the log odds ratio divided by its standard deviation) is a measure of the statistical significance of the observed effect. At the same time, the value of the odds ratio itself gives us a simple way of describing the importance or odds of a particular variable.

The log odds ratio has been used because it is resistant to two types of sample selection bias, which may be present in this study. One type of selection bias is in the distribution of relevance. Although the average precision found in this project (about 50%) is similar to that found in other studies, the end users were self-selected, and this may introduce some unknown bias in judgements of relevance. Similarly, the searchers were self-selected and, particularly with regard to cognitive characteristics, may not be typical of searchers in general. The virtue of the log odds ratio, or of the cross product ratio, is that as long as the selection biases of two variables are independent of each other, the log odds ratio is unaffected by the bias. This feature makes the log odds ratio important in so called retrospective clinical studies, where it is not possible to form a random sample. It is appropriate, for the same reasons, in this study.

Example calculation of item-wise analysis. The variable in this example of log odds ratio calculation is the number of cycles used in searches. The cut point for cycles is 3.40—this is the mean number of cycles per search. Calculations involve 8956 cases representing the total number of items retrieved by 360 searches done by all searchers (9 searches per question, for 40 questions). Of these 5287 (59%) were judged by users as relevant or partially relevant and 3669 (41%) as not relevant. There were 3486 (38.9%) cases with cycles above the mean or cut point of 3.40 and 5470 (61.1%) below the mean. In other words, 3486 items were retrieved by searches that had more than 3.40 cycles per search.

Note that the number of cycles is a property of the search as a whole, and is inherited by each of the items retrieved in that search. Thus, we expect that the items retrieved in searches with high values of cycles have a better chance to be relevant, although each particular item may be either relevant or not relevant. In fact, in searches with cycles above the cut point 2145 were relevant or partially relevant and 1341 were not relevant.

The contingency table looks like this:

		Number of Cycles		
		Below Mean	Above Mean	Total
Users Judgement	Not Relevant	2328	1341	3669 41%
	Rel. or Part Rel.	3142	2145	5287 59%
	Total	5470 61.1%	3486 38.9%	8956 100%

We can calculate the odds of REL PREL in each column:

Above mean odds	2145/1341 = 1.5995:1
Below mean odds	3142/2328 = 1.3496:1
Ratio:	1.5995/1.3496 = 1.1851
LN	(1.1851) = 0.17
STD ERROR	(from BMPD) = 0.04
	t value = 3.84

Note: 0.04 is the standard error assuming the given value of logarithm of the odds ratio. The t value is calculated on the null assumption that the cross product ratio is 1.

This is an example of a statistically significant (t larger than 2) result at 95% significance. It says that items retrieved in cases with a high (above mean) number of command cycles are by a factor 1.1851, or 18% more likely to be relevant. In other words, the odds for an item to be relevant are 18% higher when a high number of cycles is in a search. To generalize, more cycles bring higher chances for relevance.

Example of Calculation of search-wise analysis. The variable in this example is the Concrete Experience (CE) score on the Learning Style Inventory (LSI) as taken by the 36 outside searchers. The cut point for CE is 24.70—this is the mean score on the CE for the searchers (the possible score had a range from 12 to 48). Calculations involve precision for 200 searches; the mean precision was 0.54. One hundred searches had above mean values of precision and 100 below mean. Ninety-nine (49.5%) searches came from searchers with below mean CE score and 101 (50.5%) with above mean CE score. Forty-one searches came from below mean CE scores and had below mean precision.

The contingency table looks like this:

		Learning Style Inventory		
		Concrete Experience		
		Below Mean	Above Mean	Total
Precision	Below Mean	42	59	100 50%
	Above Mean	58	42	100 50%
	Total	99 49.5%	101 50.5%	200 100%

We calculate the odds of precision in each column:

Above mean	42/59 = 0.7119:1
Below mean	58/41 = 1.4146:1
Ratio	0.7119/1.4146 = 0.5032
LN	(0.5032) = −0.69
Standard error	= 0.29
t-value	= −2.39

Since the t-value even if negative is above 2, this is a statistically significant result at 95%, but in a negative sense. It indicates that searches from searchers with above mean values of CE are by a factor of 0.4968 or 49.68% (1-0.5032) less likely to have high (above mean) values of precision. In other words, searches from searchers with high CE scores are about two times (1/0.503) more likely to have low (below mean) precision. To generalize: higher CE scores bring higher chances or odds for low precision or lower chances for high precision.

Such calculations and conclusions are applied to all meaningful variables with significant results, as presented in parts II and III in the series of articles on this study of information seeking and retrieving.

Acknowledgement

We gratefully acknowledge the splendid cooperation from our users, searchers, and classification judges. Although remaining anonymous, as is the custom, they made this study possible. Our thanks to Elizabeth Logan and Nancy Woelfl for participating in set-up of various experiments and for technical advice, and to Jun-Min Jeong, J. J. Lee, Moula Cherikh, and Altay Guvenir for programming.

This was the last research project conducted at the now defunct Mathew A. Baxter School of Information and Library Science, Case Western Reserve University (1904–1986). From its inception until the very end the School encouraged research and provided an atmosphere conducive to scholarship. In recognition, the project is dedicated to the alumni and faculty of the School.

References

1. Belkin N. J.; Oddy, H. M.; Brooks, H. M. "ASK for Information Retrieval: Parts I and II," *Journal of Documentation*. 38: 61–71, 145–164; 1982.
2. Harter, S. P. *Online Information Retrieval. Concepts, Principles, and Techniques*. New York: Academic Press; 1986.
3. Belkin, N.: Vickery, A. *Interaction in Information Systems: A Review of Research From Document Retrieval to Knowledge-Based Systems*. London: The British Library; 1985. (Library and Information Research Report 35).
4. Dervin, B; Nilan, M. "Information Needs and Uses." In: Williams, M., Ed. *Annual Review of Information Science and Technology*, Vol. 21. White Plains, NY: Knowledge Industry; 1986.
5. Borgman, C. "Psychological Research in Human-Computer Interaction." In: Williams, M., Ed., *Annual Review of Information Science and Technology*, Vol. 19. White Plains N. Y.: Knowledge Industry; 1984.
6. Schneiderman, B. "Designing Menu Selection Systems," *Journal of the American Society for Information Science*. 37(2): 57–70; 1986.
7. Fenichel, C. "The Process of Searching Online Bibliographic Databases: A Review of Research," *Library Research*. 2: 107–127; 1980.
8. Bellardo, T. "Scientific Research in Online Retrieval: A Critical Review." *Library Research*. 2: 231–237; 1981.
9. Derr, R. L. "A Classification of Questions in Information Retrieval by Conceptual Presupposition." In: *Proceedings of the 45th American Society for Information Science Annual Meeting*, 19: 69–71; 1982.
10. Derr, R. L. "A Conceptual Analysis of Information Need," *Information Processing and Management*, 19(5): 273–278; 1983.
11. Derr, R. L. "Information Seeking Expressions of Users," *Journal of the American Society for Information Science*, 35(2): 124–128; March 1984.
12. Derr, R. L. "Questions: Definitions, Structure, and Classification," *RQ*. 24: 186–190; 1985.
13. Derr, R. L. "The Concept of Information in Ordinary Discourse," *Information Processing & Management*. 21(6): 489–500; 1985.
14. Pao, M. L. "Specificity of Terms in Questions." In: *Proceedings of the 46th American Society for Information Science Annual Meeting*. 20: 26–27; 1983.
15. Saracevic, T. "On a Method for Studying the Structure and Nature of Requests in Information Retrieval," In: *Proceedings of the 46th American Society for Information Science Annual Meeting*. 20: 22–25; 1983.
16. Saracevic, T. "A Research Project on Classification of Questions in Information Retrieval — Preliminary Work," In: *Proceedings of the 43rd American Society for Information Science Annual Meeting*. 17: 146–148; 1980.
17. Saracevic, T. "Measuring the Degree of Agreement Between Searchers," In: *Proceedings of the 47th American Society for Information Science Annual Meeting*. 21: 227–230; 1984.
18. Saracevic, T. "Information Retrieval," In: E. H. Brenner and T. Saracevic, ed. *Indexing and Searching in Perspective*. Philadelphia, Pa: National Federation of Abstracting and Indexing Services; 1985. Ch. 4, pp. 4-1 to 4-29.
19. Saracevic, T.; Kantor, P.; Chamis, A. Y.; Trivison, D.; *Experiments on the Cognitive Aspects of Information Seeking and Retrieving. Final Report for National Science Foundation Grant IST-8505411*. Washington, D.C.: National Technical Information Service; Educational Research Information Center; (ED 281530). 1987.
20. Fidel, R.; Soergel, D. "Factors Affecting Online Bibliographic Retrieval: A Conceptual Framework for Research," *Journal of the American Society for Information Science*. 34(3): 163–180; 1983.
21. Taylor, R. "Question Negotiation and Information Seeking in Libraries," *College and Research Libraries*. 29(3): 178–194; 1968.
22. Belkin, N. "Anomalous States of Knowledge as a Basis for Information Retrieval," *The Canadian Journal of Information Science*. 5: 133–143; 1980.
23. Belkin, N. "Cognitive Models and Information Transfer," *Social Science Information Studies*. 4: 111–129; 1984.
24. Graesser, A. C.; Black, J. B. *The Psychology of Questions*. Hillsdale, N.J.: Lawrence Erlbaum; 1985.
25. Swift, I. I. "Classifying Readers' Questions," *Wilson Bulletin for Libraries*. 8(5): 274–275; 1934.
26. Lehnert, W. G. *The Process of Question Answering*. Hillsdale, N.J.: Lawrence Erlbaum; 1978.
27. Bates, M. J. "Information Search Tactics," *Journal of the American Society for Information Science*. 30(4): 205–214; 1979.
28. Bates, M. J. "Idea Tactics," *Journal of the American Society for Information Science*. 30(5): 280–289; 1979.
29. Markey, K.: Atherton, P. *ONTAP: Online Training and Practice Manual for ERIC Data Base Searchers*. Syracuse University, ERIC Clearinghouse on Information Research, Syracuse, N.Y.; 1978. (ED 106 109).
30. Soergel, D. *Organizing Information: Principles of Database and Retrieval Systems*. New York: Academic Press, 1985.
31. Harter, S.; Peters, A. "Hueristics for Online Information Retrieval: A Typology and Preliminary Listing," *Online Review*, 9: 407–424; 1985.
32. Fenichel, C. "Online Searching: Measures That Discriminate Among Users With Different Types of Experience," *Journal of the American Society for Information Science*. 23(2): 23–32; 1981.
33. Brindle, E. A. *The Relationship Between Characteristics of Searchers and Their Behavior While Using an Online Interactive Retrieval System*. Ph.D. dissertation. Syracuse, N.Y.: Syracuse University; 1981.
34. Bellardo, T. "An Investigation of Online Searcher Traits and Their Relationship to Search Outcome," *Journal of the American Society for Information Science*. 36(4): 241–250; 1985.

35. Woelfl, N. N. *Individual Differences in Online Bibliographic Searching: The Effect of Learning Styles and Cognitive Abilities on Process and Outcome*. Ph.D. dissertation. Cleveland, Ohio: Case Western Reserve University; 1984.
36. Penniman, W. D. "A Stochastic Process Analysis of Online User Behavior," *Proceedings of the American Society for Information Science*. 12: 147–148; 1975.
37. Penniman, W. D. *Modeling and Evaluation of Online User Behavior*. Final Report to the National Library of Medicine. Columbus, Oh: OCLC, 1981.
38. Fidel, R. "Online Searching Styles: A Case-Study-Based Model of Searching Behavior," *Journal of the American Society for Information Science*. 35(4): 211–221; July 1984.
39. Oldroyd, B. K.; Citroen, C. L. "Study of the Strategies Used in Online Searching," *Online Review*. 1: 295–310; 1977.
40. Borgman, C. *User's Mental Model of an Information Retrieval System: Effects of Performance*. Ph.D. dissertation. Palo Alto, CA: Stanford University; 1984.
41. Rouse, W.; Rouse, S. "Human Information Seeking and Design of Information Systems," *Information Processing and Management*. 20(3): 92–138; 1984.
42. Sewell, W.; Teitelbaum, S. "Observations of End-User Online Searching Behavior over Eleven Years." *Journal of the American Society for Information Science*. 37(4): 234–245; 1986.
43. Fischhoff, B.; MacGregor, D. "Calibrating Databases." *Journal of the American Society for Information Science*. 37(4): 222–233; 1986.
44. Newell, A. C.; Simon, H. A. *Human Problem Solving*. Englewood Cliffs, N.J.: Prentice-Hall; 1972.
45. Meyer, R. E. *Thinking and Problem Solving: An Introduction to Human Cognition and Learning*. Glenview, IL: Scott Foresman; 1977.
46. Simon, H. A. "Information-Processing Models of Cognition," *Journal of the American Society for Information Science*. 32(5): 364–377; 1981.
47. Belnap, N. D.; Steel, T. B. *The Logic of Questions and Answers*. New Haven, CT: Yale University Press; 1976.
48. Harrah, D. "A Logic of Questions and Answers." *Philosophy of Science*. 28: 40–46; 1961.
49. Kearsley, G. P. "Questions and Question-Asking in Verbal Discourse: A Crossdisciplinary Review." *Journal of Psychological Research*. 5(4): 355–375; 1976.
50. Mednick, S. A.; Mednick, M. T. *Examiner's Manual: Remote Associates Test, College and Adult Forms 1 and 2*. Boston: Houghton Miffin; 1967.
51. Ruck, F. L.; Ruck, W. L. *Employee Aptitude Survey Technical Report*. Los Angeles: Psychological Services; 1980.
52. Kolb, D. A. *Experiential Learning: Experience as the Source of Learning and Development*. Englewood Cliffs, N.J.: Prentice-Hall; 1984.
53. *BMDP Statistical Software, 1985 Printing*. Berkeley, CA: University of California Press; 1985.
54. Fleiss, J. L. *Statistical Methods for Rates and Proportions*. New York: Wiley; 1981 (see especially chapter 5).